ORTHOGRAPHY AND LINGUISTIC THEORY: THE SYNTACTIC BASIS OF MASORETIC HEBREW PUNCTUATION

MARK ARONOFF

SUNY Stony Brook

The punctuation (accent) system of the Masoretic Hebrew Bible contains a complete unlabeled binary phrase-structure analysis of every verse, based on a single parsing principle. The systems of punctuation, phrase structure, and parsing are each presented here in detail and contrasted with their counterparts in modern linguistics. The entire system is considered as the product of linguistic analysis, rather than as a linguistic system per se; and implications are drawn for the study of written language and writing systems.*

To modern linguistics, discussion of written language has been taboo. Long ago we found a rationale for its banishment in the undoubtedly correct observation that spoken language is 'true' language, while written language is an artifact—the necessarily imperfect product of human intelligence. Nonetheless, linguistics continues to be preoccupied with written language: thus modern syntax and semantics deal with little else, and terms like 'comma intonation' and 'text linguistics' vividly attest the power of written language in governing our thought. Banishment has not worked. Written language continues to exert a strong, though unacknowledged, force on the field. As long as we ignore it, it will not go away. We must face it, if only to put it behind us.

A force so powerful cannot be approached directly, at least at first. If we wish to understand writing, it is best to try initially to understand someone else's version of it, rather than our own—to treat it as we have done religion and sex. I have, therefore, chosen as the topic of this article a writing system remote from our own. I will show, however, that it is a rich system which greatly rewards close study. I believe that, in studying it, we will begin to understand not only writing but, more importantly, the true relation between written and spoken language.[1]

My position is simple: written language is a product of linguistic awareness, the objectification of spoken language. Any orthography must therefore involve a linguistic theory. In most cases, the theory is fairly trivial, involving a simple awareness of recurring units of sound (syllables) or meaning (morphemes). In a few instances, though, orthography has moved from substance to form. Al-

* This work was made possible by a Research Fellowship from the National Endowment for the Humanities, which allowed me to devote a year to the study of orthography. The impetus to look at the syntactic basis of Masoretic punctuation came from reading Rotenberg 1978 and an unfinished paper of the same vintage by Rotenberg and John McCarthy. An early version of this article was circulated some years ago, and other papers on the accents have appeared since—notably Dresher 1981a,b. But Dresher is not concerned with the relationship between language and orthography that is my main theme here; his work is more concerned with the Masoretic text as a linguistic object. I was greatly aided in my revision by Robert Hoberman and John McCarthy.

[1] There have been a few brave attempts to break the taboo, notably by Francis 1958, Vachek 1973, Klima 1972, Mattingly 1972, Haas 1976, and especially Justeson 1976. My own views of the general relation between written and spoken language are close to those of Mattingly; but rather than risk the danger inherent in re-telling, I refer the reader directly to the sources.

phabetic writing, with its recognition of the phoneme, is an example. Less well known, but just as remarkable, is the subject of this article, the punctuation system (or accentuation, as it is traditionally called) of Masoretic Hebrew. As I will show, Masoretic punctuation is based on a remarkably elegant syntactic theory, unsurpassed in descriptive power until very recently, yet almost completely unknown because it is embedded in an orthographic system. This theory, once revealed, shows how orthography is inextricably tied to linguistic analysis.

I have chosen Masoretic punctuation because it is so startling, and because it is little known outside a small circle of Semitists; indeed, even among Hebraists—despite the fact that the basis of the accents has been well understood for a century—an acquaintance with more than the most elementary rudiments of the system is rare. The Masoretes themselves would have scorned such advertisement, as they scorned all worldly fame; but we now inhabit a different age.

1. HEBREW, THE MASORETES, AND MASORETIC HEBREW. In order to discuss the Masoretic accents, I must first lay a groundwork for those unfamiliar with the history of the Hebrew language, especially in its written manifestations. As a Northwest Semitic language, Hebrew is closely related to Phoenician, the language of the first alphabet, and to Aramaic, the language of the Babylonian and Persian empires and of Christ. It is somewhat less closely related to Arabic. Biblical Hebrew (hereafter BH) is the language of the standard text of the Hebrew Bible as codified between about 500 B.C. and 200 A.D. (Talmon 1964).[2] The orthography of BH was probably fixed by about 200 B.C. (Freedman 1962). This orthography, descended from Phoenician via Aramaic, is usually termed 'consonantal' or 'defective'. Indeed, the early Northwest scripts marked only consonants (Cross & Freedman 1952); but the term 'consonantal' is misleading when applied to BH, since the orthography also marks long vowels, by using glide signs to indicate most final and some medial long vowels.[3]

The standard consonantal text, especially that of the Pentateuch, is regarded in the Jewish tradition as divinely inspired. Torah scrolls, the proper production of which is circumscribed by numerous guidelines (extending even to the mental purity of the scribe), may contain only the consonantal text. The text is fixed; even when an error is known to exist in the standard version, it is left standing, though it may be replaced in oral reading by its correction. As a fixed object, it may even be used extralinguistically—e.g. in the numerological tradition

[2] This is not to say that BH was a standardized language. The language of the books varies according to their age and history. For our purpose, though, these differences are not crucial.

[3] The introduction of these vowel letters was fairly gradual, and those books which were established earlier have fewer vowel letters. Scribes must have been aware of this development, since very late Biblical books are sometimes written in an 'archaizing' style, with fewer vowel letters than one might expect. This archaizing was supposed to give them an air of antiquity and authenticity. It should also be noted that vowel length was probably no longer distinctive by the time of the Masoretes, so that the glides no longer marked only long vowels in the Masoretic text. Schramm 1964 therefore disregards the glides in his transcription of pure vowels, as I will do here.

whereby passages are interpreted according to their purely numerical value, which is obtained by treating each of the 22 letters as a positive integer.[4]

If we disregard its sacredness, and look at it simply as a transmitter, the BH text has a number of deficiencies. For one, it contains acknowledged errors, which cannot be noted in the text itself because of its very sacredness. It does not indicate short vowels, and even the marking of long vowels is sometimes deficient or ambiguous. It is not punctuated. Remember also that this text became the center of post-exilic Jewish life: it was studied intensely as a guide to proper conduct, and its recitation was the focal point of religious worship. Inevitably, a tradition grew up concerning the proper way of preserving the integrity of the text and of its oral recitation. This tradition is called MAS(S)ORAH, a word of controversial etymology, and the propagators of the tradition are called MASORETES. Masorah was at first preserved orally, because of the inviolable nature of the written text; and it must have developed at an early date, because one finds references to it in the Talmud. Eventually, a distinction was made between tokens of the Biblical text used in religious worship and those which were used only for study. The former could not be violated, and had to meet all traditional criteria of material and form; but texts destined solely for study were exempted from certain restrictions. In particular, they no longer had to be written on vellum scrolls, usually taking instead the form of bound books (codices); and they could be annotated. Masoretic Hebrew, then, is the language of the annotated study text, which is called the MASORETIC TEXT. There were in fact a number of written Masoretic traditions (Kahle 1913, 1927, 1930); but the Tiberian system, associated with the school of Tiberias in Palestine, won out over the rest at an early date, and is universally recognized as standard.

The Masoretic annotation takes several forms. First are the marginal annotations, which indicate variant readings and orthographic peculiarities, as well as much statistical information about the numbers of occurrences of individual word forms. The term 'Masorah' is sometimes reserved for these marginalia. Within the text itself, we find three distinct types of annotation. The first is the division of the text into verses. The second is phonological, indicating all those segmental phonological properties of the recited text which are not clearly marked in the original consonantal text; this includes vowels and certain properties of consonants (cf. Schramm). The last is the accent notation, which forms my topic. The segmental notation has extended its domain beyond the Bible, and is now regarded as an integral, if only optional, part of (even Modern) Hebrew orthography. The accent system, by contrast, has fallen into general disuse, and is not found outside the Bible.

There is no apparent connection between the Masoretes and any other scholarly group. The Masoretes were preceded by the *tannaim* and the *amoraim*. The former compiled the Mishna, while the latter compiled the Gemara, a commentary on the Mishna. Together, the Mishna and Gemara form the Tal-

[4] The use of these letters to represent integers is ancient (cf. Greek).

mud, which was completed by about the 5th century A.D., and forms the basis of legalistic rabbinical Judaism. One main concern of the Talmud is the interpretation of the Biblical text, and the codification of the practices which developed around it. The language of the text is seldom treated by Talmudists as a topic, since they are interested in what lies behind the language. The Masoretes were followed in time by the Grammarians, who flourished from the 10th to the 14th centuries. The first well-known Grammarian was Saadia Gaon, who wrote in Arabic and was heavily influenced by the scholars who codified Classical Arabic. The last of the great Grammarians was David Kimchi, whose *Mikhlol* (W. Chomsky 1952) was the standard Hebrew grammar until the 19th century. The Grammarians apparently knew little about the Masoretes, since they treated their annotations as a given part of the text—something to be explained. We can therefore assume that the Masoretes had completed their work by the time of the first Grammarians, in fact early enough to have been forgotten.

The Masoretes lie apart from the mainstream of Jewish scholarship. We know little about them, aside from some names (e.g. the ben-Asher family), and the place of their greatest flourishing, Tiberias in Palestine. Unlike the Grammarians, many of whom were known as scholars apart from their grammatical work (e.g. ibn-Ezra), the Masoretes apparently produced nothing but the edition of the Biblical text which bears their name. Nor did they leave behind any explanation of the principles behind their work. It is generally assumed that the names which were given to the Masoretic symbols originated with the Masoretes themselves, since most of these names are Aramaic, and the Masoretes were the last of the great Aramaic-speaking scholars; but even this is not known for certain, since they left behind almost nothing but their text. It has sometimes been claimed—and this claim is endorsed by the most prominent of modern Masoretic scholars (Kahle 1959)—that they were Karaites, members of a fundamentalist sect which rejected the authority of the Talmud and all other rabbinical writings in favor of the actual text of the Bible; certainly the complete isolation of the Masoretes from the entire rabbinical tradition, and their concern solely with textual matters, make sense if we assume that they were Karaites. But whatever their beliefs may have been, their edition of the Biblical text became standard throughout Judaism.

The work of the Masoretes was complete by the end of the 9th century, for Aaron ben-Asher, the last well-known member of his family, wrote his treatise *diqduqe haṭṭʕɔmim* in the 10th century, and the text must have been complete sometime before the treatise was written. The idea of annotating a consonantal text seems to have come from Syriac, where vocalization signs were in use from about the 5th century. The Tiberian system of punctuation was more fully developed than the Babylonian and Palestinian systems, which have survived only in fragmentary texts (Kahle 1913, 1930); it is therefore reasonable to assume that it was the latest (but see Revell 1970). With these endpoints in mind, we can assume that the Masoretic system with which we are dealing was constructed sometime between about 600 and 800 by a group of scholars working

somewhat apart from the rest of the Jewish scholarly community. Their great concern was the regulation of the proper reading of the Biblical text in Hebrew, and they developed a system of symbols designed to aid in this task.

BH is a fairly typical V(erb) S(ubject) O(bject) language, similar in most respects to other North Semitic languages such as Arabic and Aramaic. The order of basic constituents is usually VSO (27 of 31 verses of the first chapter of Genesis begin with a verb); but in accordance with Greenberg's 1966 observation that initial position is emphatic—and with his Universal 6, by which all VSO languages have SVO as an alternative order—emphasized elements may be placed in front of the verb, so that SVO, OVS, and SOV all occur. In verbless sentences, i.e. in sentences where the predicate is not an inflected main verb, the subject generally precedes the predicate, unless the predicate is emphasized. Pronouns differ from nouns: a separate pronoun subject is optional—but when it occurs, it usually precedes the verb, even when not emphatic. Aside from pronoun subjects, the other elements which usually precede the verb are conjunctions, interrogatives, and the sentence negative. Object pronouns are usually cliticized after the verb, unless they are emphatic. Modifiers of the VSO core generally follow it; their general pattern is essentially similar to that described by Williams 1975 and Jackendoff 1978 for analogous adverbial elements in English (VP Complements, S Complements, Adverbial Clauses and Phrases). As in English, the outermost modifiers may precede the VSO core, though they do so infrequently.

BH is also a typical Greenberg VSO language in other respects: it has prepositions rather than postpositions; the genitive follows the noun which it modifies, as do adjectives and demonstratives; adverbs follow adjectives; WH-words are sentence-initial; relative clauses follow their head nouns. Hebrew has a definite article, two genders, and no cases. Verbs and adjectives agree with nouns in number and gender; verbs also agree in person, and adjectives in definiteness.

2. THE ACCENT SYSTEM.[5] Having put the Masoretes in context, I now turn to their work. I will not concern myself with their work on segmental phonology—though it contains some fascinating problems, such as the use of the shwa symbol to mark both the vowel and the absence of a vowel (ambiguously in some cases) and the diverse uses of the *dageš*, an internal dot (Malone 1975; in general, *dageš* marks a consonant symbol as having occlusive rather than fricative value). Nor will I say anything about the marginal Masorah, which has a distinct aim and is not truly orthographic. That leaves the accents.

The accents have three distinct discernible functions. We do not know which was basic, or whether all three were tied together from the beginning. First, and of least interest to us, is the use of the accents to mark stress. Every

[5] I will devote my analysis only to the function and distribution of the DISJUNCTIVE accents, which are used for marking syntactic breaks. The other, CONJUNCTIVE, accents are used in most instances to indicate absence of a syntactic break. The other markers—*maqqef* (hyphen) and *meθeg*, which marks secondary stress (Dresher 1981a)—do not have melodic values; they form a distinct class, not traditionally included among the accents.

orthographic word bears an accent (two words joined by a hyphen are treated as one orthographic word), and the accent is usually placed on the initial consonant of the stressed syllable, though a few accent symbols are word-initial or word-final. Stress in Masoretic Hebrew is usually final, but sometimes falls on the next-to-last syllable. It is possible to predict where it will fall in almost all cases (Prince 1975), but the regularity is not transparent; it is therefore convenient to have the accent marked. Nonetheless, most words have final stress, and penultimate stress comes only in certain restricted environments—mostly before an easily learned list of suffixes—so that, in practice, fluent readers rarely need the accents. It is in fact unlikely that this elaborate system had the primary function of marking stress, and it is misleading to call the marks 'accents'. Indeed, the word 'accent' does not properly translate the original Hebrew word, which is better rendered as 'sense'; but 'accent' has always been the accepted term in the non-Hebrew European scholarly tradition. It makes little sense to change it in this article, so I continue to use it despite its misleading connotations.

The second function of the accents is musical, and most non-specialists assume that this is their central raison d'être. The text of the Bible is chanted during worship; and in this chanting, called CANTILLATION, the accents are interpreted musically. Each of the 27 different accent symbols is assigned a specific melody (not usually a single note), and the practiced cantor can read the symbols much as dancers read choreographic notation.[6] The practice of chanting the text, rather than simply reading it, is certainly ancient, and it is also true that the distribution of the accents results in part from musical factors.[7] Unfortunately, the original musical system, i.e. the particular musical value of each symbol (if ever there was a single system) has been lost. Individual communities now have their own values, and great variation often exists even within a single community. Nor has there been any real success in reconstructing prior systems (Idelsohn 1929), though the values of some of the symbols can be narrowed by considering their names—which are mostly uniform across communities and are presumably original (e.g. 'expulsion', 'resting', 'sustained', 'stopping')—or from the impressionistic descriptions of early treatises. The latter, however, are 'so brief and enigmatical that no one has yet succeeded in deciphering and explaining them' (Wickes 1887:13). Because of our ignorance of the original musical values of the symbols, it is difficult to understand much of the system from a musical point of view: we can tell that certain regularities must have been musically motivated originally, but can go no further. We

[6] This musical function may explain the stress marking: in the simplest tune–text association, the downbeat of a tune is associated with the stressed syllable of a word. The accents are placed on the stressed syllable because that is the center of the tune.

[7] For example, the accent *pašṭɔ* is replaced by *yθilv* when its word has initial stress and is not preceded by a conjunctive accent (about which see below). Wickes (1887:106) says that 'The substitution is entirely on musical grounds.' He suggests that the melody of *pašṭɔ* is such that it demands at least one syllable before the stressed syllable; when no such syllable occurs, another melody, that of *yθilv*, which occurs only under these limited circumstances, is substituted for it. The explanation is reasonable, and other phenomena seem to have a similar musical basis.

cannot give a particular musical explanation without knowing the original melodies which motivated the phenomena. Thus, though the musical significance of these symbols is what people are most aware of when they use the accents, the study of this phenomenon holds less reward than one might expect, and I will not pursue it further here.

This brings us to the final function of the accents, and the one which is of greatest potential interest to the modern linguist: that of punctuation. As shown below, the Masoretic accents comprise the most detailed, complete, syntactically based system of punctuation ever used. I have already pointed out that the Masoretes, for reasons which we do not know, did not hand down an understanding of the various notations which they invented. Thus, even for the segmental notation, we still do not know why they did certain things as they did; but in comparison to the accents, the segments are like crystal. Jews who use the accents have always known that they have 'pausal' values, and that some have greater such value than others; however, it is not unfair to say that they have had little understanding of the system within which the individual accents are embedded. Our understanding of this system is therefore recent, and results largely from Christian scholarship beginning in the 17th century and culminating in Wickes 1881, 1887. There has also been some work by more recent Semitists (Spanier 1927, Breuer 1958, Cohen 1969, Dotan 1970); but these scholars have only amended the analysis which Wickes provided, and have not questioned its basic tenets. Most of what I will have to say about the system of the accents themselves is based on Wickes and on these subsequent works.[8] My work is therefore not entirely original; and my debt to Wickes is so great that I would consider the present paper a success if it merely made people aware of his remarkable achievement.

However, my concerns are not those of Wickes: I am interested in the syntactic analysis upon which the accent system is constructed. Wickes was aware of this syntactic analysis, but he had no special interest in it; nor would we expect a 19th century Semitist to be overly concerned with fine points of syntactic analysis. Wickes' main interest was in the distribution of the accents themselves, and he devoted the bulk of his work to that topic. His discussion of syntax is thus fairly general, and sometimes impressionistic. I have therefore extracted from his syntactic remarks only those which I myself, in studying the syntax of the accents, have found most trustworthy; and I have tried to restrict myself to fairly common constructions, for which the evidence is clear. I have sometimes translated Wickes' observations into modern terms; and I have added observations of my own, sometimes going beyond Wickes, but never contradicting him. Finally, I have uncovered a general syntactic theory, heretofore unknown, which seems to underlie the whole system.

The accents are distributed according to what Wickes calls 'the law of the CONTINUOUS DICHOTOMY' (29). Each verse of the Bible is divided into two parts,

[8] I have also restricted my attention to the prose books of the Bible analysed by Wickes 1887; I will have nothing to say about the slightly different accentuation of the three poetical books treated by Wickes 1881. All further references to Wickes should therefore be understood as to the later work, and all unattributed quotations are from it.

and each of these in two, and so on until no group of more than two words remains undivided. A disjunctive accent is placed at each dividing point, the accents being ranked such that the higher division is marked by a higher-ranking accent. As for the principle by which the division is made: 'It is found, where the main LOGICAL pause of the clause, or the rules for syntactical division require it' (31). In other words, the accents are arranged so as to provide a complete, unlabeled, binary, constituent structure analysis of each verse. What follows is devoted to a study of this analysis.[9]

Two disclaimers: First, I do not claim that the purpose of the accentuation was syntactic analysis, or that the Masoretes were primarily syntacticians. What they wanted to do was bring out as clearly as possible the literal meaning of the text—its SENSE, in the terminology of modern semantics; and they saw that they could best do this by marking the exact relations of the words of the text to one another, down to the finest detail. They even called the signs that they used SENSES (Heb. *ṭəʕamim*). Given the hierarchical structure of language, it is inevitable that the relations which they noted should be syntactic, even though they may not have thought of them as such; thus modern Semitists think of the accents as pauses, with each higher accent designating a longer pause. But whatever one thinks of the accents, it is clear that the divisions of the verse which the accents encoded were not haphazard; they followed certain syntactic principles, and by studying them we can make these principles plain.

Second, I should comment on my syntactic notation—that of phrase structure trees, familiar to generative syntacticians. I use this notation because it is clear, especially when dealing with long and complicated structures of the sort often encountered in the Bible, and because it is well known. No further significance should be attached to it: I do not wish to imply that the Masoretes were proto-generativists. In fact, certain features of their analysis, such as its binary basis and its overt disregard for category labels, are more reminiscent of American structuralism; the most basic methodological principle of their analysis is related both to Harris 1951 and to N. Chomsky 1970.

After explaining the syntactic analysis upon which the accentual system rests, it will be possible to compare this analysis with the corresponding product of a modern theory. It will be seen that the two coincide in some places and differ in others. In one instance, the Masoretic system treats a construction which has been ignored in the modern literature—that of direct quotation—and provides an insightful analysis which we could adopt. In other instances, the Masoretic analysis appears to be wrong; I will show HOW the Masoretes went wrong, guided by an incomplete theory. This only partial coincidence is important for more than simply historical reasons. My own interest in Masoretic accentuation stems from a concern with the nature of orthography. In partic-

[9] There are instances where the division does not follow the syntax. Wickes' view of these (p. 3) is enlightening: 'Such irregularities (if we are so to term them) cannot be ignored. What then are we to say to them? Are we, on account of them, to reject the whole system, as unreliable for the discrimination of the sense? or are we to try and find some explanation of them, so that we may make due allowance, in every case, for disturbances as they occur? Unquestionably, the latter is the true scientific course; nor till we have failed in discovering the necessary explanation, have we any right to condemn what it may turn out we did not understand.'

ular, I believe that orthography is not the product of primary linguistic activity, but rather of metalinguistic awareness, much as is grammar. If Masoretic accentuation, or any writing system, is the product of linguistic analysis and linguistic theory, then it has a more than reasonable chance of being wrong. Furthermore, if the theory on which it is based is well worked out, we can even work back from the errors of analysis to the theory behind the errors. Thus, if the Masoretes made mistakes, they were linguists. If, however, the Masoretic principles were merely implicit in the minds of those who obeyed them—i.e., if the accentuation were 'psychologically real', a window into the mind of language—then we would expect the analysis contained in the accents to be 'real'. But it is not 'real', for we can discover the principles behind it; we can see its virtues and its deficiencies. The accents must therefore have followed from a conscious system of linguistic analysis. The fact that this system comes close to ours in many respects is then noteworthy, though mere agreement should not be taken as evidence for the correctness of either system. Thus, not only does the Masoretic system of syntactically-based punctuation go beyond any other known orthography in analytic subtlety, it also demonstrates unequivocally the close connection between orthography and linguistic analysis which is my theme.

3. PHRASE STRUCTURE. My original hope was to describe the surface structure syntax of BH in modern terms, then to describe the syntactic analysis of the Masoretic punctuation, and finally to compare the two. Unfortunately, when I set out to do this, I realized that the state of modern syntactic theory, especially with regard to VSO languages, often made it difficult to provide an analysis which would meet with general approval. I therefore decided to treat the Masoretic analysis as central, and to introduce modern analysis only when necessary—i.e. when the two differ significantly, or when the modern analysis might serve to clarify the ancient one.

The Masoretic system is purely relational; it marks only constituent breaks, and gives no labels for the constituents themselves. I will therefore not embark on any discussion of nodes or categories at this point. The Masoretic system is also purely binary, and the initial unit is the Biblical verse. The division of the Bible into verses I accept as given, as it was accepted also by the Masoretes; I will not speculate on the criteria according to which this prior division was made.[10] A cursory examination of any random prose passage from the Bible will reveal that these criteria were not primarily syntactic: a single verse may correspond to any multiple of S (and rarely contains less than a single S; e.g. Exodus 34.6–7, Gen. 23.17–18). I will also take as given the hyphen-like symbol called *maqqef*. Two or more words may be joined by a hyphen, and any such group is treated as a single word for the purpose of accentuation.

The verse is first divided at the major syntactic break. This division proceeds similarly until there is no undivided unit of more than two words. A modern syntactician may mark these divisions by bracketing each unit:

[10] Since the Biblical verse is the domain of rules governing the distribution of the accents, we MUST assume that the division into verses is prior to the assignment of accents, whether or not this reflects actual history—and similarly for the *maqqef* or hyphen.

(1) [John [saw [the light]]].

The nesting of the brackets indicates the inclusion of smaller units within larger ones. In the Biblical system, when a given string is divided in two, the end of the first immediate constituent of that string is marked by a DISJUNCTIVE accent on the last word of that string. The accents are ranked, so that a higher division is marked by a higher-ranking accent. Ex. 1 would be given the following sort of representation (where the integers correspond to accents of different levels, the highest rank having the lowest integer):

(2) John_2 saw_3 the light_1.

Since every verse is its own largest unit, the highest-ranking accent always goes at the end. In ex. 2, the next accent comes between the Subject and the VP, these being the immediate constituents into which the string of Level 1 is divided. Note that, though only the first half bears an accent of Level 2, the remaining half is assumed to be a unit at the same level, since the accent serves to mark the point of division. Since *John*, the Subject, is a single word, it receives no further division; but *saw the light* consists of three words, so it must be divided. The division in this unit comes after *saw*; hence *saw* is given the next following accent. Scanning the example, we see no unaccented sequence of more than two words, and we therefore stop the division. At this point, any unaccented word—the only one in our example being *the*—is given a separate type of accent, called CONJUNCTIVE, to indicate that it does not end a syntactic unit. By this method, any verse, no matter how complex, can always be broken down into its constituent parts; and the resulting division can be indicated by consecutively numbered accents, each number corresponding to a level of division. In actuality, the Masoretic system has only four accent levels; how these are used to represent more than four levels of division will be shown below. I will now discuss the syntactic principles according to which the division is made, beginning with S.

3.1. THE SENTENCE. A verse consisting of more than one S is first divided according to the general rule for conjunctions, which will be discussed below in §3.3. I will therefore begin with the typical V-initial sentence of the form VSO X Y, where X and Y are complement phrases. These are treated as uniformly left-branching structures, firmly rooted in V. Outer complements are peeled off one at a time, the greatest division being at the outermost complement; and the V-headed structure extends all the way down to VSO, which is analysed as VS O. Some examples are given in Figures 1a–d.[11]

[11] Many of these examples are from Wickes or Cohen, some are from Spanier, and some I found myself. In every case where I have used someone else's example, I have adapted his accentuation, when necessary, to that of Elliger & Rudolph 1977, the standard modern scholarly edition of the Hebrew Old Testament. Under the cited passage is given a word-by-word translation into English, with English words not present in the Hebrew being placed in parentheses. For the most part, I have adopted Schramm's phonological transcription. His transcription does not mark shwa or vowel length, except for shwa-colored ultra-short vowels, which are marked with a breve sign. Schramm distinguishes the two types of *dageš*, using a raised dot (C·) to mark strong *dageš* only. I use a double consonant instead of the raised dot, in accord with the usual phonetic interpretation of strong *dageš*.

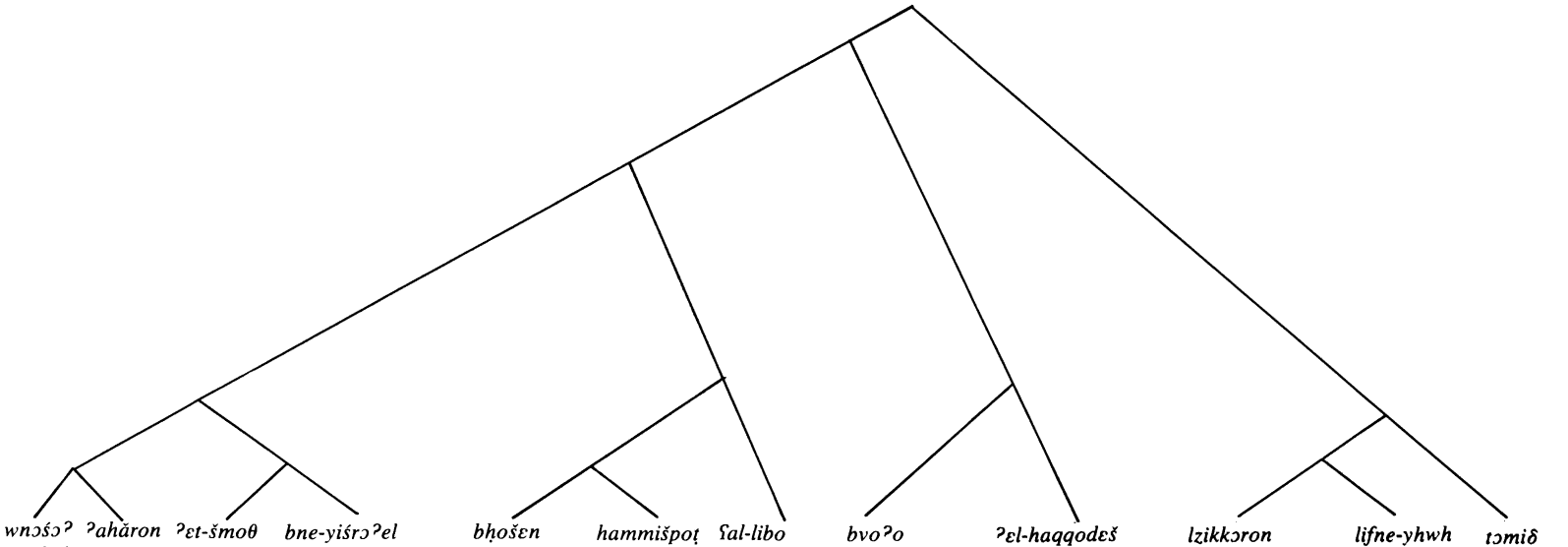


FIGURE 1a. 'Aaron carried the names of the sons of Israel on the plate of justice on his heart on coming to the holy place in perpetual remembrance to God.' (Ex. 28.29)

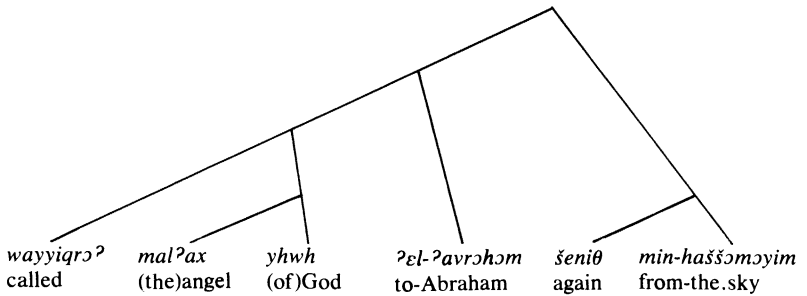


FIGURE 1b. 'The angel of God called to Abraham again from the sky.' (Gen. 22.15)

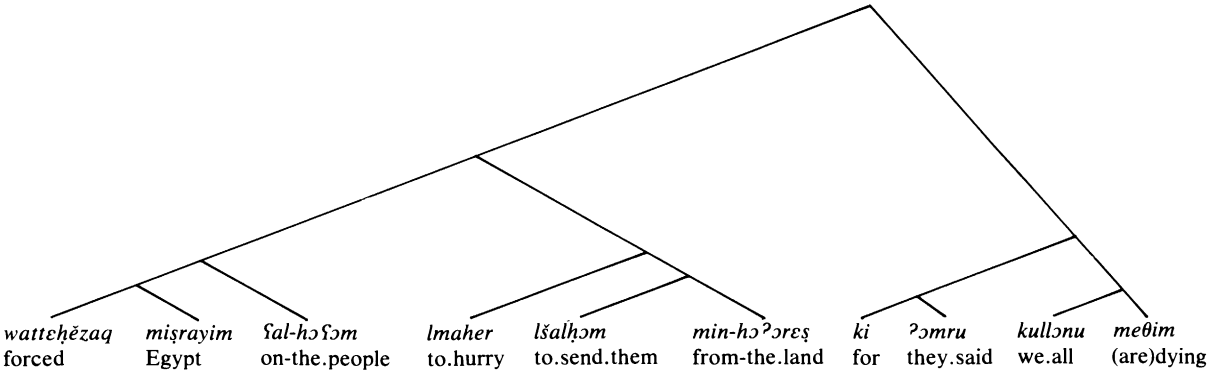


FIGURE 1c. 'Egypt pressured the people to hurry, in order to send them out of the land, for they said, we are all dying.' (Ex. 12.33)

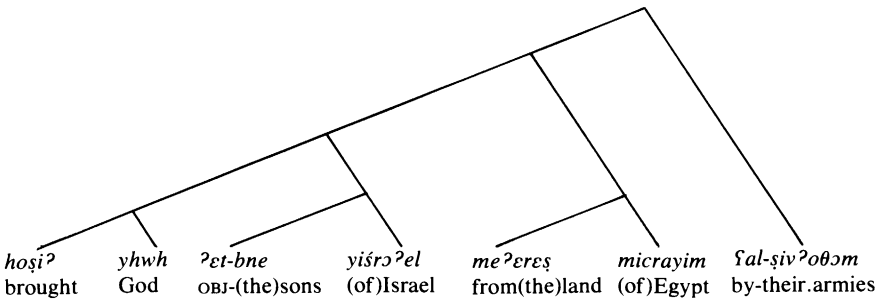


FIGURE 1d. 'God brought the sons of Israel out of the land of Egypt in armies.' (Ex. 12.51)

Wickes notes a few exceptions to this general division; however, 'Against them are to be set the thousands of instances in which THE RULE FOR THE DIVISION OF THE VERBAL CLAUSE IS CARRIED OUT' (51). Such a layered analysis is similar to those proposed by modern syntacticians (cf. Jackendoff, Williams) for such structures—though analysts now tend to posit a specific number of layers, and to assign complements to these according to how closely bound they are to the verb. Furthermore, working with SVO languages, neither Jackendoff nor Williams confronts the subject in exactly the same way.

In direct contrast with the treatment of V-initial sentences is that of sentences in which a normally postverbal element is found preverbally. Greenberg points out that VSO languages always allow certain elements to appear preverbally for reasons of emphasis. This emphatic (or topicalized) value of preverbal elements is recognized in the Masoretic accentuation. Sentences of this pattern are analysed as having a major syntactic break between the topicalized element and the rest of the sentence. In Figure 2a, the topicalized phrase is a subject,

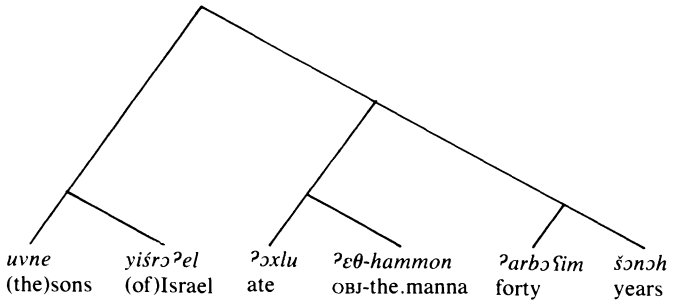


FIGURE 2a. 'The sons of Israel ate the manna for forty years.' (EX. 16.35)

and in Fig. 2b an object. In Fig. 2c, a PP is fronted, and in Fig. 2d an adverb. When two distinct phrases have been fronted, we find the major break after the first, and the next greatest break after the second, as in Fig. 2e.

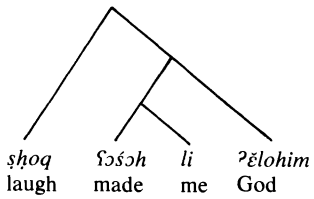


FIGURE 2b. 'God made me laugh.' (Gen. 21.6)

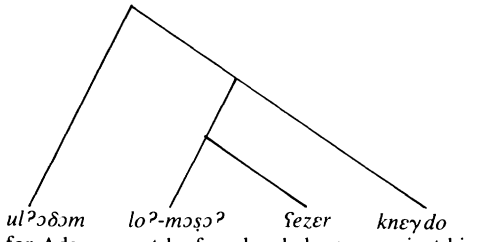


FIGURE 2c. 'For Adam he did not find a worthy helper.' (Gen. 2.20)

Not all preverbal elements, though, are regarded as topicalized; those elements which normally precede the verb are not. The most remarkable of these is the pronoun subject. Subject pronouns are optional with tensed verbs, and they are fairly infrequent. However, when they do occur, they generally pre-

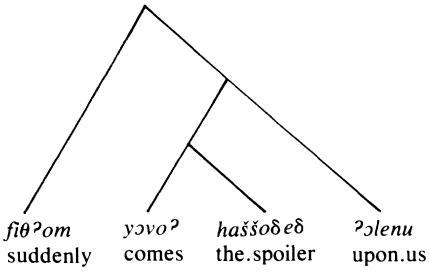


FIGURE 2d. 'Suddenly, the spoiler comes upon us.' (Jeremiah 6.26)

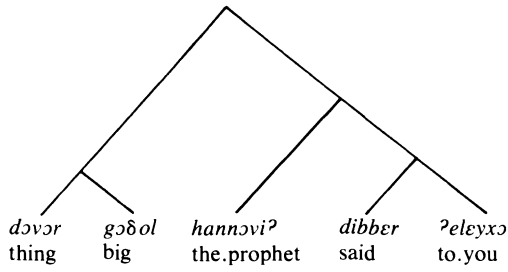


FIGURE 2e. 'The prophet told you a great thing.' (2Kings 5.13)

cede the verb, whether or not they are emphatic.[12] When the pronouns are not emphatic, the Masoretic accents do not indicate a major break between the pronoun and the verb. In Wickes' words, 'THE PERSONAL AND OTHER PRONOUNS are not always considered important enough to stand by themselves' (45). For examples, see Figures 3a–c.

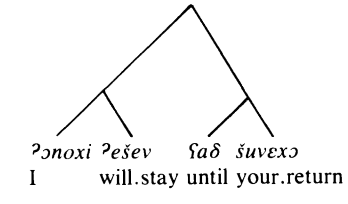


FIGURE 3a. 'I will stay until your return.' (Judges 6.18)

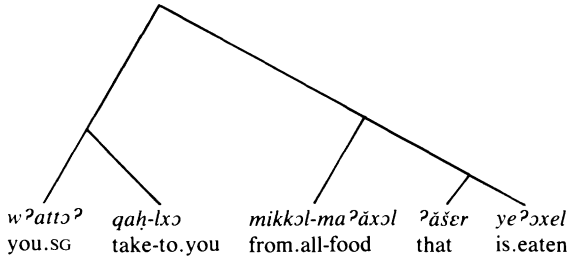


FIGURE 3b. 'Take to you from all edible food.' (Gen. 6.21)

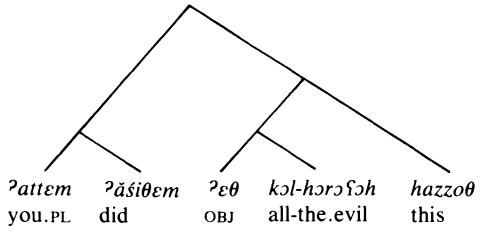


FIGURE 3c. 'You did all this evil.' (1Samuel 12.20)

Not all sentences contain a main verb. Those whose predicate is not a finite main verb are termed NOMINAL in traditional Hebrew grammar.[13] In such sen-

[12] Modern Hebrew uses subject pronouns much more frequently, and is also largely SVO. The two facts are probably not unrelated.

[13] Traditional Arabic grammar groups those sentences which I call nominal together with any sentence which begins with a noun—including those discussed above in which the sentence contains a finite main verb, and the initial noun is merely topicalized. I assume that this grouping is incorrect for Hebrew, though it may be more defensible in Arabic, where fronted nouns leave a pronoun copy. See Anshen & Schreiber 1968 for a discussion of the Arabic cases.

tences, the subject usually comes first, regardless of whether it is emphatic. Here the major break may come after the subject (i.e., the subject may be treated as topicalized); but alternatively, a minor break (or none) may come after the subject, as in Figures 4a–c.

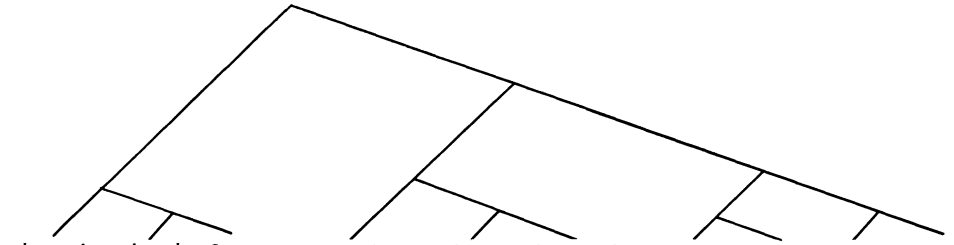


FIGURE 4a. 'The snake was more subtle than all the beasts of the field that God made.' (Gen. 3.1)

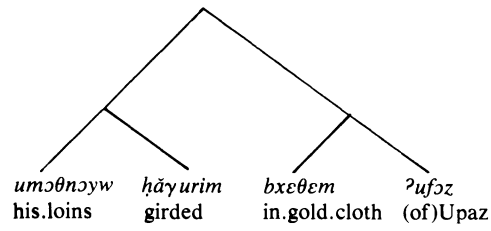


FIGURE 4b. 'His loins were girded in gold cloth of Upaz.' (Daniel 10.5)

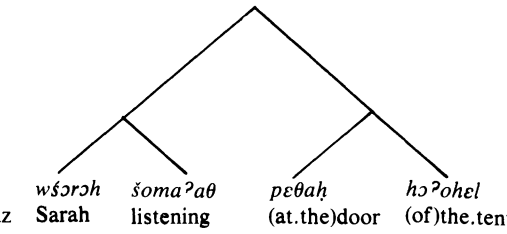
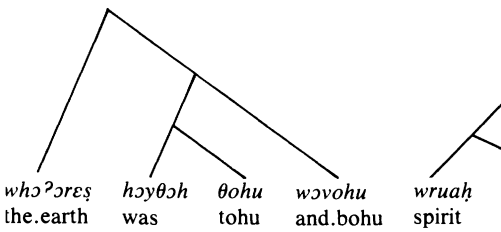


FIGURE 4c. 'Sarah was listening at the door of the tent.' (Gen. 18.10)

The major break can come after the subject, especially if the subject phrase is long or is emphasized. Nonetheless, a major break is much less common here than it is after a topicalized subject in a verbal sentence; see Figures 5a–c.



IGURE 5a. 'The earth was tohu and bohu.' (Gen. 1.2)

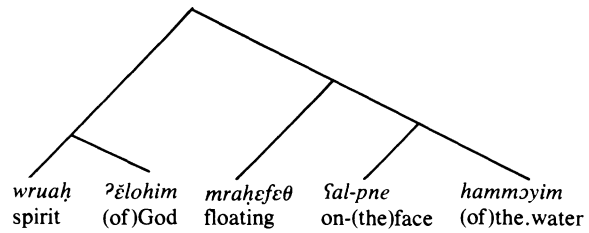


FIGURE 5b. 'The spirit of God was floating on the face of the water.' (Gen. 1.2)

3.2. NOUN PHRASES. Apart from relative clauses (which I will discuss below), modifiers of nouns are not as common as one might wish. Nonetheless, there is enough material for us to establish the general analysis of NP with some certainty. Except for cardinal numerals, which immediately precede the head noun, all NP modifiers follow the head. The order of constituents within the NP is basically as follows:

(3) NUM N NP_{gen} AP_0 DEM

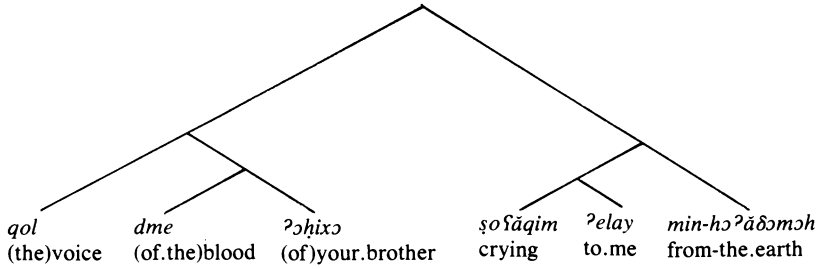


FIGURE 5c. 'The voice of the blood of your brother is crying to me from the earth.' (Gen. 4.10)

Except for the numeral, whose analysis is problematic, the analysis of NP is similar to that of S; i.e., we find a left-branching tree rooted in the noun, as in Figures 6a–c.

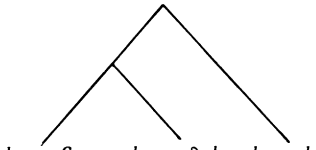


FIGURE 6a. 'this big storm' (Jonah 1.12)

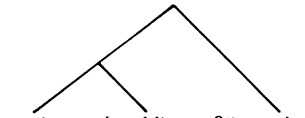


FIGURE 6b. 'grizzled bay horses' (Zechariah 6.3)

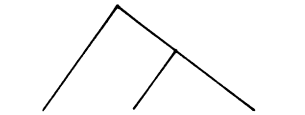


FIGURE 6c. 'the four winds of the sky' (Zech. 6.5)

This left-branching structure is contradicted in one aspect of nominal syntax, the so-called CONSTRUCT PHRASE. 'Construct' is the term used for the phrase which consists of the head noun and its following genitive. What is peculiar about Hebrew is that the genitive noun itself is not distinguished either by affixation, or by any internal phonological change from a noun standing alone; nor is it preceded by any preposition like Eng. *of*. Instead, the governing head noun is often phonologically distinct from its independent form:

(4) a. *brɔxɔ* 'blessing'
birkat hammɔzon 'the blessing of the food'
b. *dɔvɔr* 'word'
dvar mošɛh 'the word of Moses'

This modified form of the head noun is called the CONSTRUCT FORM. Most phonological differences between the independent and the construct forms of nouns can be derived by regular rules of the phonology if it is assumed that the entire construct phrase forms a single phonological word for purposes of stress—the nouns being separated only by a single word-boundary, rather than the usual two.[14] The first consequence of this dependence is a subordination of the stress of the construct noun; from this, the other phonological differences between the construct and the independent form follow naturally. I will not discuss the

[14] The Masoretes sometimes joined the members of the construct phrase with a hyphen, but at other times the construct form was given its own accent. The distribution of the two treatments is not predictable.

phonology in detail here, but readers are encouraged to consult Prince for a complete presentation of the analysis outlined here.

Recursion of the construct phrase is possible, and is in fact quite common, the construct having a wide range of application in Hebrew. Embedded constructs can be found quite easily:

(5) *pri γoδɛl lvav mɛlɛx ʔaššur*
fruit size heart king Assyria
'the fruit of the size of the heart of the King of Assyria' (Isaiah 10.12)

From a phonological point of view, these longer sequences are exactly analogous to simple two-word construct phrases: they form single phonological words. Syntactically, however, they are more complex. Let us assume that the simple construct would be treated in a modern analysis as a mirror image (in simple linear terms) of the Eng. genitive: thus, parallel to Figure 7a, we would have the equivalent BH form in Figure 7b.

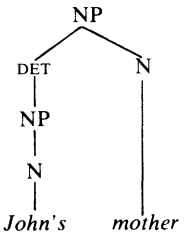


FIGURE 7a.

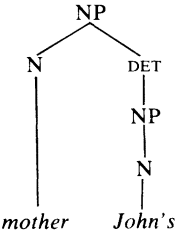


FIGURE 7b.

Parallel to the analysis of the embedded Eng. genitive given in Figure 8a, we would expect the BH equivalent of Figure 8b.

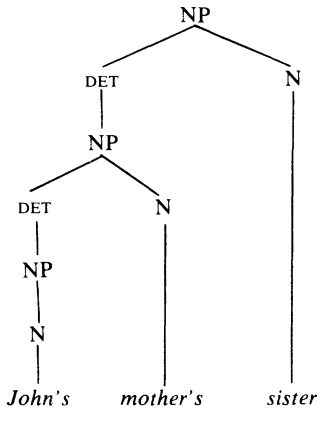


FIGURE 8a.

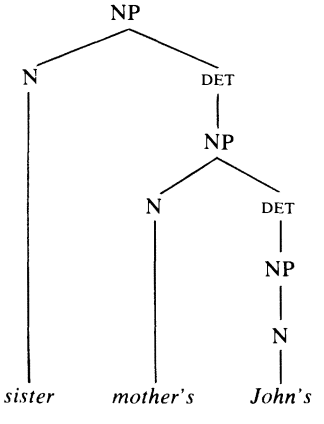


FIGURE 8b.

Let us now ignore all non-branching nodes in the last examples. This turns the Eng. genitive into a left-branching tree, and the BH construct phrase into a right-branching tree, as in Figures 9a–b.

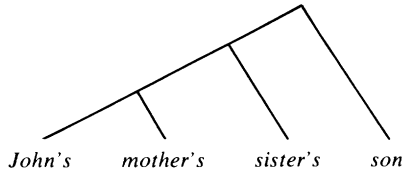


FIGURE 9a.

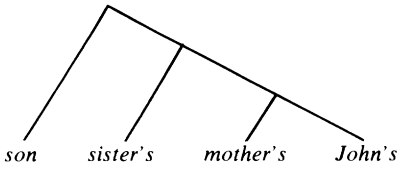


FIGURE 9b.

Recall that the Masoretic accents do not provide labeled trees. Within such a system there are no non-branching nodes. The Masoretic analysis of embedded constructs is therefore exactly that of Fig. 9: a uniform right-branching tree, as in Figures 10a–b.

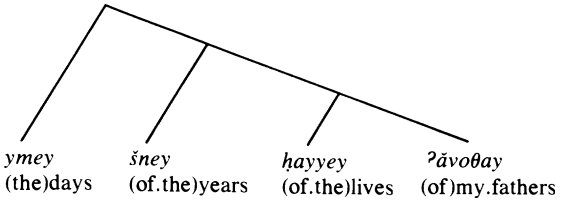


FIGURE 10a. 'the days of the years of the lives of my fathers' (Gen. 47.9)

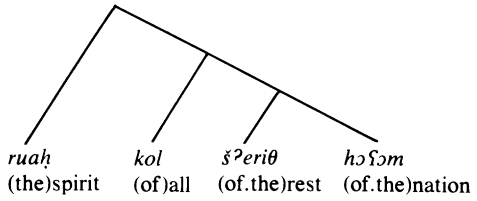


FIGURE 10b. 'the spirit of all the rest of the nation' (Haggai 1.14)

This uniform branching for complex construct phrases will be violated only when the head of a construct phrase is itself a construct phrase, as in Figure 11a. This is in opposition to a possible right-branching analysis, as in Figure 11b (which is not the Masoretic analysis).

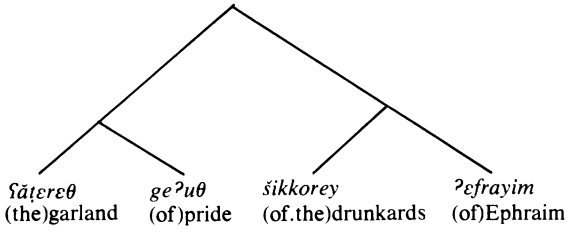


FIGURE 11a. 'the garland of pride of the drunkards of Ephraim' (Is. 28.1)

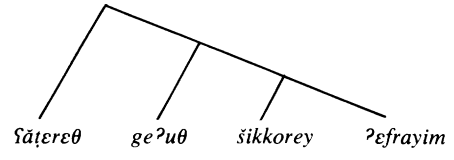


FIGURE 11b.

Note that the two structures are phonologically identical, and that they are fairly similar semantically. The distinction is indeed subtle, and few examples like this are found. The mere existence of these examples, however, is good evidence for the position that the essential purpose of the Masoretic accents was syntactic, rather than musical—since such differences in accent make little sense unless we assume them to have been motivated by a desire to bring out the proper relationships among the words, i.e. the syntax.

As noted above, when a noun is modified by both a genitive and an adjective, the adjective follows the genitive. This is true no matter how complex the genitive. Any adjective must follow the entire construct phrase, regardless of which noun it modifies. In most cases, the adjective will modify either the head noun or the last noun, and it agrees with whichever one it modifies:

(6) a. *bsefɛr hattorɔh hazzoʔθ* (Deuteronomy 28.61)
in.book the.Torah this
'in the book of this Torah'
b. *bsefɛr hattorɔh hazzɛh* (Deut. 29.20)
'in this book of the Torah'

In 6a, the demonstrative agrees with 'Torah', which is feminine; in 6b, it agrees with 'book', which is masculine. The accentuation of these phrases gives the analyses shown in Figures 12a–b.

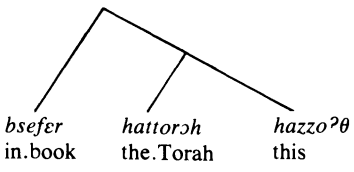


FIGURE 12a.

FIGURE 12b.

3.3. CONJUNCTION. The representation of conjunction in a purely binary system is always problematic—not because it is difficult, but rather because the binary hierarchy which results when more than two elements are concatenated is intuitively unsatisfying. For example, given the schema X → X CONJ X, and given two conjuncts, the structure produced is shown in Figure 13a. This is fine; but with three conjuncts, the binary schema produces the two structures of Figures 13b–c.

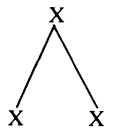


FIGURE 13a.

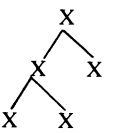


FIGURE 13b.

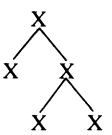


FIGURE 13c.

Intuitively, many feel that three or more conjuncts should not automatically engender more embedding than two; nor should any single conjunct be more deeply embedded than the others, as will always be the case when the total number of conjuncts is odd. But despite these objections, the binary schema

makes up in elegance for what it lacks in intuitive appeal; and though the extra structure is cumbersome, I know of no argument which proves it wrong.[15] Nor apparently did the Masoretes, since their general schema is that above, with one condition:

(7) MASORETIC CONJUNCTION RULE: X → X CONJ X
Condition: expand from left to right at each level of analysis.

The schema and condition produce the structures of Figures 14a–d for up to five conjuncts (instances of more than five conjuncts seldom occur).

FIGURE 14a.

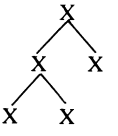


FIGURE 14b.

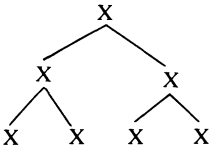


FIGURE 14c.

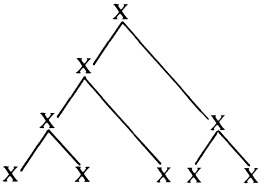


FIGURE 14d.

Examples of various numbers of conjuncts are given in Figures 15a–d (overleaf). The schema applies to all conjuncts (sentences, verbs, nouns etc.).

3.4. CLITICS. The representation of clitics in Hebrew orthography, both Biblical and Masoretic, is complex. Certain elements—such as the conjunction *w*, the question particle *ha*, the definite article *haC*, the complementizer *š*, and the C(V) prepositions *b, l, k, mi(C)*—are prefixed to the following word, and are treated as part of that word with respect to the phonology. Object and possessive pronouns are suffixes. All these affixal elements are best treated phonologically as '+' boundary affixes (cf. Prince). The orthography, in keeping with the phonological analysis, does not separate these particular affixes from their bases, but puts them together as single words. This practice is ancient, and is standard in the consonantal text which predates the Masoretes.

Certain other elements appear to be weakly cliticized to the following word. These are: prepositions longer than CV (*ʕaδ, lifney, ʔaḥar, biγlal* etc.); conjunctions (*ki, ʔim, ʔo* etc.); the complementizer *ʔăšɛr*; and the sentence negatives *loʔ, ʔen*. These elements are never written as a single unit with the next word. However, they are often joined by means of a hyphen, and the resultant combination is treated as a single word by the accents. (As noted above, the accents presuppose the hyphen, and I will not discuss its various uses further—

[15] However, see McCawley (1976:301–2).

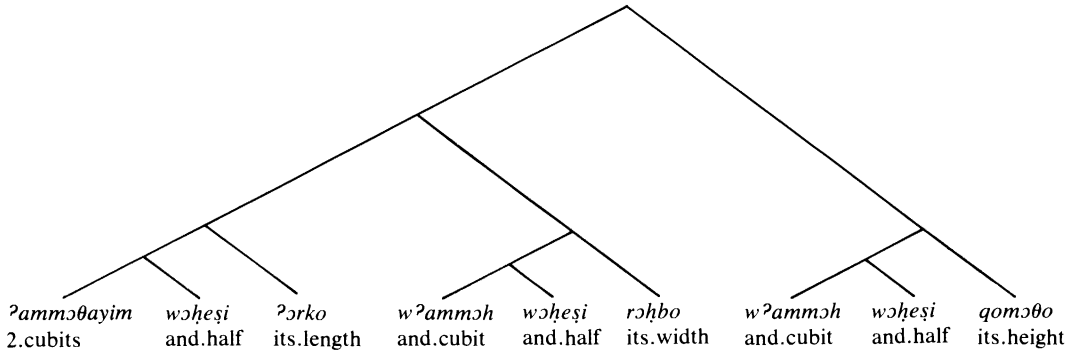


FIGURE 15a. Three conjuncts: 'two cubits and a half in length, and a cubit and a half in width, and a cubit and a half in height' (Ex. 25.10)

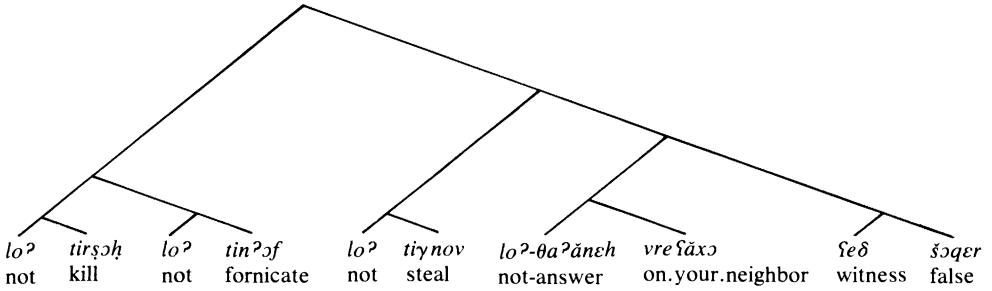


FIGURE 15b. Four conjuncts: 'Do not kill, do not fornicate, do not steal, do not bear false witness on your neighbor.' (Ex. 20.13)

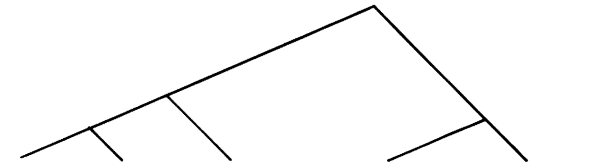


FIGURE 15c. Five conjuncts: 'Er and Onan and Shela and Perez and Zorah' (Gen. 46.12)

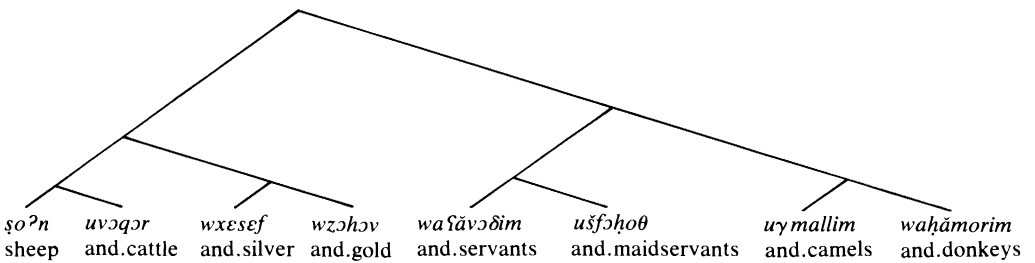


FIGURE 15d. Eight conjuncts: 'sheep and cattle and silver and gold and servants and maidservants and camels and donkeys' (Gen. 24.35)

though they are of some interest.) When not joined by a hyphen, the weakly cliticized elements often bear a conjunctive accent. There is no apparent principle determining exactly when we will find a hyphen, when a conjunctive accent, and when neither. In fact, it is in the punctuation of these elements that individual codices are in least agreement. Nonetheless, though the details of treatment are not entirely clear-cut, the classification of these elements as clitics, and of a looser sort than the affixal ones, is apparent.

3.5. DIRECT DISCOURSE. Up to this point, my discussion of individual constructions has touched upon those which modern theoreticians have treated in some detail, and the Masoretic analysis has by and large been similar to the modern one. However, one construction whose Masoretic analysis is quite clear, but which has received little attention from modern syntax, is direct discourse. I do not know why so little attention has been paid to this construction, since it is so common—though I suspect that it may have been felt to be literary, artificial, and thus of little interest.[16] Be that as it may, most people's ideas about Eng. direct speech seem to have been formed by punctuation, i.e. the use of quotation marks to set off passages of direct speech.

Syntactically, direct speech is interesting only when it occurs inside a sentence together with what I will call the INTRODUCTORY PHRASE (e.g. I *said* or *he exclaimed*). The Eng. punctuation system treats the direct discourse as subordinate to the introductory phrase, no matter where the introductory phrase stands with respect to the quoted discourse:

(8) 'I don't want you living with roaches,' said Muhammed Ali.
Muhammed Ali said, 'I don't want you living with roaches.'
'I,' said Muhammed Ali, 'don't want you living with roaches.'

A modern analysis which followed the punctuation would consider the discourse to be subordinate to the introductory phrase, and would treat the direct discourse as analogous to indirect discourse—except for the lack of a complementizer, and differences in sequence of tense and pronouns. The introductory phrase, however, differs from ordinary higher sentences introducing indirect discourse in that the latter cannot be shifted about:

(9) Muhammed Ali said that he doesn't want you living with roaches.
?That he doesn't want you living with roaches Muhammed Ali said.
*That he, Muhammed Ali said, doesn't want you living with roaches.

This is not to say that introductory phrases can be shifted anywhere in the sentence:

(10) *'I don't want,' said Muhammed Ali, 'you living with roaches.'
*'I don't want you living with,' said Muhammed Ali, 'roaches.'

The acceptability pattern above is paralleled by that of parenthetical expressions:

[16] Banfield 1973 is a notable exception.

(11) Muhammed Ali doesn't want you living with roaches, you know.
You know, Muhammed Ali doesn't want you living with roaches.
Muhammed Ali, you know, doesn't want you living with roaches.
*Muhammed Ali doesn't want, you know, you living with roaches.
*Muhammed Ali doesn't want you, you know, living with roaches.

Parentheticals in turn pattern like Sentence Adverbs:

(12) Muhammed Ali doesn't want you living with roaches, honestly.
Honestly, Muhammed Ali doesn't want you living with roaches.
Muhammed Ali, honestly, doesn't want you living with roaches.
*Muhammed Ali doesn't want, honestly, you living with roaches.
*Muhammed Ali doesn't want you, honestly, living with roaches.

Following this line of reasoning to its natural conclusion, we find that introductory phrases are best treated as a sort of sentence adverbial; i.e., they are subordinate to the direct discourse rather than superordinate to it. The punctuation is therefore misleading.[17]

The reader can most likely discern my next statement, which is that the treatment of direct discourse in the Masoretic system is, by and large, that which I have here advocated on general linguistic grounds. Introductory phrases are subordinated to the discourse in the manner of adverbial expressions. To quote Wickes (35–6) in full on this matter:

> 'Particularly noteworthy is the way in which the words that introduce a speech—or anything similar, as a command, decree, oath, covenant, &c.—are treated. They constantly occupy a SUBORDINATE POSITION, as far as the accents are concerned. The clause containing THE SPEECH ITSELF, THE COMMAND, &c., is counted the more important, and receives the main accentuation. In short, the division is made (as above) as if the introductory words were absent, e.g.
>
> "And God said, Let there be a firmament in the midst of the waters, // and let it divide the waters from the waters" (Gen. 1.6).
>
> "And Jehovah said to him, Therefore whosoever slayeth Cain, // vengeance shall be taken sevenfold" (4.15).
>
> Such cases occur in every page.'

This ends my discussion of individual analyses contained in the accentual system. I turn now to more general matters.

[17] The unorthodox treatment which James Joyce gave to the punctuation of his work, probably modeled on that of the Romance languages, eschews the use of quotation marks (Joyce thought that they were ugly); this is much more in tune with an analysis of introductory phrases as parenthetical adverbial expressions. The following passage from *A portrait of the artist as a young man* exemplifies Joyce's punctuation:

> Uncle Charles smoked such black twist that at last his nephew suggested to him to enjoy his morning smoke in a little outhouse at the end of the garden.
>
> —Very good, Simon. All serene, Simon, said the old man tranquilly. Anywhere you like. The outhouse will do me nicely: it will be more salubrious.
>
> —Damn me, said Mr. Daedalus frankly, if I know how you can smoke such villainous awful tobacco. It's like gunpowder, by God.
>
> —It's very nice, Simon, replied the old man. Very cool and mollifying.

The literary-minded reader might also look at *Tristram Shandy* or at some of the stories of Grace Paley for punctuation which eschews quotation marks in a similar fashion.

4. THE BASIS OF THE ACCENTS. Wickes proposes that the accents have their origin in the metrical structure of Hebrew poetry, each verse of which consists of two equal halves—with equality probably being defined in terms of the number of syllables—and where this division is strongly echoed syntactically and semantically as well, primarily by means of the well-known principle of parallelism.[18] As Wickes points out, the poetry is usually very simple from a syntactic point of view. The structure of the poetry could thus provide a foundation for a basic principle of dichotomy which was then extended to the prose passages, which were both more complex syntactically and less obviously binary in structure. The first advantage of Wickes' proposal is that it provides a reasonable explanation for why the system is strictly binary—even where a binary analysis is not optimal, as with conjunction, or when there are more than two consecutive sentences in a given Biblical verse. In addition, the basis in parallelism serves to explain several classes of exceptions to the syntactic division of the verse. These include cases where a major break is disregarded, and the main dichotomy is instead placed between two subordinate phrases, which are parallel:[19]

(13) 'Like as many were astonished at thee,—his visage was so marred more than man, // and his form more than the sons of men.' (Is. 52.14)

It is as if parallelism were viewed as the basic determiner of dichotomy, even when the parallel structure was itself syntactically subordinate.

There are also many cases where a major break will be ignored if it comes too close to the beginning or end of a verse. The main dichotomy will be placed closer to the middle of the verse, ignoring syntax for the sake of rhythmic balance:

(14) 'And he took a calf // fine and good.' (Gen. 18.7)

Syntactically, the main dichotomy should come after *took*, but this would impair the rhythm of the phrase. Note, though, that the syntax is not always overridden for the sake of rhythm. The syntax usually wins out. We can therefore conclude that the principle of dichotomy probably had its origin in the structure of Hebrew verse, and that vestiges of this origin emerge in cases where syntactic structure was overridden in favor of more purely poetic features.[20]

The historical basis of Masoretic accent may be found in poetry. But it is clear that, by the time the analysis had crystallized into the system under discussion here, it had moved far beyond its origins, and had developed into

[18] O'Connor 1980 argues that even the regularity in line length results directly from syntactic patterning.

[19] All examples where // is used to indicate the break are taken directly from Wickes.

[20] This putative poetic origin is also of general theoretical interest, since it serves to establish yet another connection between types of linguistic awareness. According to linguistic stylistic critics of the Prague School, the most essential property of the poetic use of language is FOREGROUNDING. Among those matters most characteristically foregrounded in Hebrew poetry is syntax. Translating Wickes' claim into a Praguean framework, we might say that the foregrounding of the syntax, as found in the poetry, led to increasing objectification of the syntactic structure of the language, culminating in the syntactically based punctuation.

a consistent theory of syntactic constituent structure. I have dealt thus far with individual syntactic constructions—with single phrase structure rules, so to speak. It is entirely possible that such an enumeration exhausts the field. To take a better-known example, within the theory of phrase structure presented in N. Chomsky 1965, the base rules share only the fact that they are context-free rewrite rules. By contrast, within the bar theory of Chomsky 1970 and subsequent work, an attempt is made to develop a series of phrase structure rules on the basis of a single principle; the same is true of Harris' system of phrase structure, as Chomsky himself notes.

We may therefore ask at this point whether the Masoretic analysis operated in terms of such a single principle. Here I go beyond my sources—yet I do not move far, since my answer develops out of an observation of Wickes (45):

> 'When the clause ... consists of two parts, the first syntactically complete in itself, the second a supplemental appendage (a *Zusatz*, to use a German term, which exactly expresses the construction) consisting generally of a preposition with its government or an adverbial expression, the main dichotomy may be placed at the end of the first part.'

Wickes' observation may be extended into a general principle, which appears to be basic to the Masoretic theory of phrase structure:

(15) THE MASORETIC PARSING PRINCIPLE: Given a constituent X_i of category X, divide it into two continuous subconstituents such that one of them is the maximal continuous constituent of the same category X within X_i.

For example, given a typical V-initial sentence, the principle finds the longest constituent sentence, and makes the major break at the end of it. Given a NP, the principle takes the longest possible NP within it. This has two possible outcomes: if the NP to be analysed has a final Adjective, then the Adjective will be removed; if the NP is a construct, then the first N, the head, will be removed. Thus we can derive the most common divisions, that of V-initial S, and those of NP, from this principle.

Note that the principle makes no claims about the relationship of the constituents to one another. This is why we find such seemingly anomalous parsings as the following, where the smaller constituent modifies only part of the larger constituent:

(16) a. 'Are we not counted of him strangers? For he hath sold us, // and hath also quite devoured our money.' (Gen. 31.15)
b. 'For God doth know that, in the day ye eat thereof, then your eyes shall be opened, // and ye shall be as gods, knowing good and evil.' (3.5)
c. 'The serpent was crafty // beyond all the beasts of the field.' (3.1)
d. 'Him their father loved // above all his brothers.' (37.4)
e. 'And to Abraham he acted kindly // on her account.' (12.16)

In 16a–b, the break is made before a clause which modifies only the second part of the first half; in 16c–d, it is made in the middle of the VP—when it should be made after the initial topicalized NP.

Even more striking is the treatment of appositional phrases exemplified

below, which is not uncommon (each higher integer signifies a successively weaker break, as in §3 above):

(17) 'And I give to you/$_4$ and to your descendants after you/$_3$ the land of your sojourn/$_2$ all the land of Canaan/$_1$ for an everlasting possession.' (Gen. 17.8)

Similarly, final relative clauses are sometimes separated from their head by a major break instead of a minor one:

(18) 'And you came and inherited the land // which the Lord promised to your fathers.' (Deut. 8.1)

This is not always the case, and relatives may be treated as constituents with their heads:

(19) 'And Abraham paid to Ephron/$_1$ the money/$_2$ which he had said/$_3$ in the hearing of the Hittites.' (Gen. 23.16)

Breuer says that the latter analysis is more common. But a relative clause is not separated from its head by the accent unless it is final. Why then does this occur in final position?

All the above anomalies can be accounted for if we assume that they stem from application of the Parsing Principle. In fact, in order to avoid cases like these, the principle would have to be augmented by some fairly complex condition, which I will not attempt to state, whereby the two constituents must stand in construction with one another, where 'in construction with' must be defined.[21]

So far as I can see, the Masoretic principle as stated fails to account for the Masoretic analysis of only two major constructions: topicalization and coördination. I have said that the major break in a sentence containing a topicalized constituent comes after that constituent. But topicalized sentences are generally of the form X S Y, so that the principle would be just as likely to make the break before Y as after X, unless Y is null. A sentence like the following should therefore have two equally probable analyses:

(20) a. 'The Lord / sent an east wind on the land.'
b. 'The Lord sent an east wind / on the land.' (Ex. 10.13)

However, though analyses of the (b) type occur, they are much less frequent than the (a) type. We must therefore assume that the Masoretes took V-initial sentences to be more typical (which they are) and preferred the analysis which maximized this sort of sentence. Whether this decision was a deliberate one cannot be determined. As for coördination, it differs from most other structures in that a coördination of category X, by definition, contains two instances of X. The choice of the Masoretes was apparently to make both maximal. This can be handled by a simple parenthesis:

(21) THE REVISED PARSING PRINCIPLE: Given a constituent of the category X, divide it in two in such a manner as to maximize its continuous subconstituent(s) of the category X.

[21] Modern theories of syntax have had remarkably little success in defining the notion 'in construction with' in terms of phrase structure. Indeed, recent invocations of non-configurationality in syntax (Hale 1983) might be taken as an admission that the notion defies any structural definition.

The Parsing Principle accounts for the great bulk of Masoretic analyses, including those which differ from what a modern syntactician might expect. Since the Masoretes left no record of the system behind their accentuation, it is impossible to say with absolute certainty that this principle lay at the heart of their analytic method. Nonetheless, the laws of scientific inquiry compel us to ascribe such a principle to them. Nor could the principle have been implicit or unconscious. For one thing, the domain of its application is heterogeneous and not always linguistically 'natural'; the Biblical verse, for example, does not correspond to any single linguistic category. The Parsing Principle, therefore, is a linguistic theory, and the accents are distributed according to the binary phrase-structure analyses which result from this theory. In this principle and its application we see the relation between linguistic theory and orthographic practice; and if the relation is striking, it is only because of the depth and elegance of the theory. Every orthography is based on a linguistic theory of some sort—though most such theories are so trivial that they are easily ignored. The non-trivialness of the Parsing Principle is the center of my argument.

One of the most striking features of the Masoretic system is the consistency with which the Parsing Principle is applied, even in cases where the resulting analysis is intuitively unsatisfactory. I have already mentioned a few of these cases, and I will review them more fully here, as well as discussing some other more complex cases.

In 16–18 above, we saw cases of the general pattern AB/C where a modern analysis would be A/BC. Since the Parsing Principle demands the maximal constituent of the category X, the tendency is to maximize X at the expense of other factors, especially the connection between B and C. Thus, in 17, we expect the analysis of 22, but instead we find 23:

(22) 'And I give to you and to your descendants after you / the land of your sojourn, all the land of Canaan.'

(23) 'And I give to you and to your descendants after you the land of your sojourn / all the land of Canaan.'

X has been maximized. In the same example, at the next level of analysis, we expect 24, but we find 25:

(24) 'And I give / to you and to your descendants after you.'

(25) 'And I give to you / and to your descendants after you.'

Again, X has been maximized. The same is true of the relative in 18; and the fact that only final relative clauses are so analysed follows from the maximization of X. Sentences with final relative clauses are of the form Q N S. The Masoretic analysis is Q N / S rather than Q / N S, maximizing the initial sentence.

I noted above that, in coördinate constructions with an odd number of constituents, the first half is given the extra constituent. Three conjuncts will be divided into two plus one, and so on. However, there are exceptions, where the second half is longer, as in Figures 16a–b (overleaf).

A comparison of these examples with those in Fig. 15, especially Fig. 15c, reveals the crucial factor leading to the unusual analysis: normally, every conjunct but the first is preceded by the coördinate conjunction *w*. In the examples

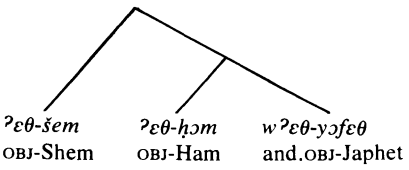


FIGURE 16a. 'Shem, Ham, and Japhet' (Gen. 5.32)

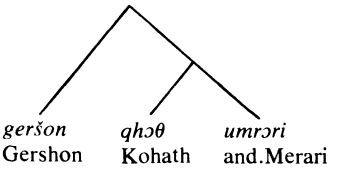


FIGURE 16b. 'Gershon, Kohath, and Merari' (Gen. 46.11)

here, the conjunction has been omitted before all but the last conjunct, as it normally is in English. Because of this omission, the first and second conjuncts taken together do not form a conjunction—the last constituent does not bear a *w*. The second and third together do form a conjunction. By the Parsing Principle, therefore, only the analysis of Figs. 16a–b is possible, since it provides a conjunction within a conjunction, maximizing X.

The Parsing Principle can also explain the existence of two distinct analyses of conjoined subjects. Normally, the conjoined subject is treated as a single constituent, as in Figures 17a–b.

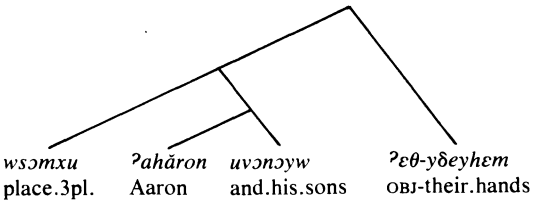


FIGURE 17a. 'Aaron and his sons placed their hands.' (Ex. 29.15)

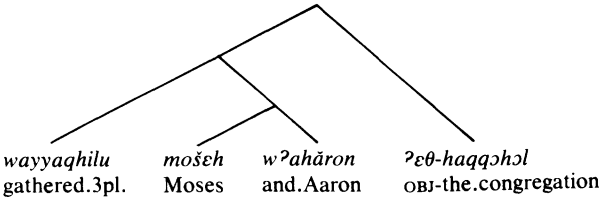


FIGURE 17b. 'Moses and Aaron gathered the congregation.' (Numbers 20.10)

However, when the verb agrees in number and gender only with the first of two coördinate conjoined subjects, we find a different analysis; the verb and first conjunct are separated from the second conjunct, as in Figures 18a–b.

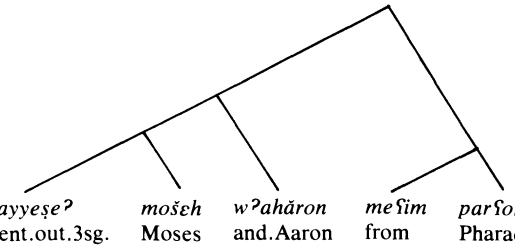


FIGURE 18a. 'Moses and Aaron went out from Pharaoh.' (Ex. 8.8)

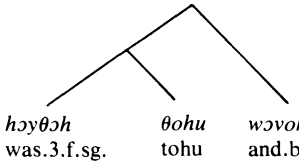


FIGURE 18b. 'was tohu and bohu' (Gen. 1.2)

The working of the Parsing Principle is also clear in this case: given a sentence of the form V_{sg} N N, its maximal sentential constituent is V_{sg} N. When the verb is plural, as in Fig. 17, then such an analysis is impossible, and the 'normal' analysis results by default.

5. THE ACCENTS THEMSELVES. I now turn to a discussion of the accentual notation itself. I do this partly to satisfy the curiosity of those readers who may be interested in the actual accents, and partly because of an interesting property of the system which cannot be understood without some grasp of the notation. This property makes it possible for people using the notation to know, at any point in their reading, the relationship between that point and the rest of the verse in question. Needless to say, this facet of the notation could be of great benefit to the reader. However, the extent to which modern readers take advantage of it is questionable, mainly because of most readers' limited understanding of the accent system. Apart from its possible utility, though, it is a nice example of how the result of a global system of analysis—which, as its basic principle shows, must operate over an entire verse—can be converted into a simple Markovian notation, once the analysis is completed.

First, a note on directionality: Hebrew is written from right to left. Since certain principles of accentuation are directional, this difference between the original script and the transcribed script may potentially cause some confusion in an explication of the accents. I have decided, therefore, to use the terms 'beginning' and 'end', 'before' and 'after', 'first' and 'second', rather than 'left' and 'right' in my discussion. The only potentially difficult term of these four is END, which must be read only in the sense of 'terminus ad quem' and not as 'a quo'.

There are eighteen disjunctive accents; five are simple contextual variants of others, and two are also limited in context. This leaves eleven accents, which can be divided into four ranks. According to Cohen's notation, which I will adopt here, the ranks are D_0, D_1, D_2, and D_3: D_0 is the strongest (marking the greatest break), and D_3 the weakest. D_0 is also the simplest to describe. It has only two members, and each member can occur only once in a verse. The first, *silluq*, is placed at the end of the verse, and the second, *ʔaθnɔḥ*, is placed at the major break in the verse. Let us call the constituent before any given accent its DOMAIN (alternatively, we could say that an accent is placed at the end of its domain). We can now say that the domain of *ʔaθnɔḥ* is the first half of the verse. The domain of *silluq* is either the entire verse or the second half of the verse. I know of no empirical difference between the two descriptions, but I will use the second. Thus a verse is divided at its major constituent break; *ʔaθnɔḥ* is placed at the end of the first half, which is its domain; and *silluq* is placed at the end of the second half, which is its domain.

There are two important D_1 accents; and, as expected, they are used in the division of D_0 domains. But unlike D_0 accents, D_1 accents may be repeated: any given D_0 domain will be divided by a D_1 accent at its major constituent break. The following are schematic examples of the placement of D_1, with capital letters representing words:

(26) a. $X\ Y\ Z\ W\ P\ T_0 \rightarrow X\ Y\ \ Z_1\ W\ P\ T_0$
b. $X\ Y\ Z\ W\ P\ T_0 \rightarrow X\ Y_1\ Z\ \ W\ P\ T_0$

In 26a, [XYZ] and [WPT] are the immediate constituents, and 1 marks the constituent break; analogously for 26b. In 26a we would say that [XYZ] is the domain of 1. Now, since each constituent of 26a has more than two members, each must be analysed further. The first constituent will be divided by a Level 2 accent:

(27) $X\ Y_2\ Z_1$

The second constituent, however, which itself contains no Level 1 accent, will be divided by a Level 1 accent:

(28) $W\ P_1\ T_0$

This Level 1 accent, despite its being of the same level as the preceding one, marks a lower break. No confusion arises, however, because of the following convention (which operates at Levels 1 and 2):

(29) SEQUENCE OF ACCENTS: Of two accents A_i and A_j of the same level in sequence, the first is stronger.

We can explain this convention as follows: A Level 0 constituent is scanned, and a Level 1 accent is placed at its major break. The first constituent of the original constituent is now regarded as being in the domain of a Level 1 accent, since an accent is placed at the end of its domain. We now disregard this first constituent, and move on to the second. This part is still in the domain of a Level 0 accent and must be divided by a Level 1 accent, giving the following division of the second constituent of 26a:[22]

(30) $W\ P\ T_0 \rightarrow W\ P_1\ T_0$

In 26b, let us assume that the second constituent has the following structure, after the assignment of a second Level 1 accent:

(31) $Z_1\ W\ P\ T_0$

The string W P T must be analysed further; and since it is not in the domain of a Level 1 accent, it will be divided by a Level 1 accent, say after W:

(32) $W_1\ P\ T_0$

No further division is possible, and 26b now has the accent structure:

(33) $X\ Y_1\ Z_1\ W_1\ P\ T_0$

The accents provide the structure of Figure 19.

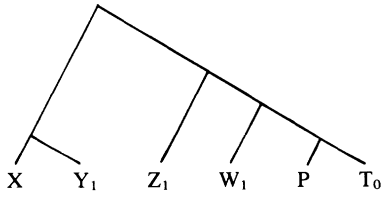


FIGURE 19.

[22] The first constituent of 26a cannot be divided by a Level 1 accent, because it ends in one. We will see below that this constituent is divided by a Level 2 accent.

The Sequence of Accents convention is easily interpretable, if we see that the relative strength of identical accents is determined directionally, and remember that an accent marks the end of its domain. Furthermore, once we overcome initial confusions, we can see the utility of the convention, since it allows us to indicate any number of successively smaller constituent breaks.

Why are there two Level 1 accents? Given the picture thus far, we would expect only one, with that accent being repeated as necessary within a given Level 0 domain. The answer is that one particular type of Level 1 accent is a FINAL ACCENT, or FORETONE. That accent, *ṭifḥɔ*, occurs as the last Level 1 accent preceding a Level 0 accent. Let us identify a final accent by subscript *f*; then Level 1 will contain two accents, A_1 and A_{1f}. The rule of final accent placement applies even if there is only one Level 1 accent, so that the non-final Level 1 accent will appear only when there is a sequence of accents. The origin of the final accent is not clear. Wickes calls it a 'foretone', suggesting a musical rationale, but gives no real argument for it. But whatever its origin, the final/non-final distinction allows the reader to know what is coming next. The accent after A_{1f} will always be D_0 (marking a stronger break), while that after A_1 will always mark a weaker break.

I will now give some examples of how the Level 0 and Level 1 accents are distributed. The examples and analysis are from Cohen, whose exposition is very clear. I use subscripts to represent the accents. The first examples, in Figures 20a–b (overleaf), show the simple D_0 division.

Next is an initial D_1 division of the same verse, as in Figures 21a–b (p. 61).

Finally, an example of the successive application of D_1 accents, taken step by step, appears in Figures 22a–d (p. 62).

Level 1 accentuation still leaves certain elements of a verse unanalysed. For example, in Fig. 20a, in the second D_0 domain, the first D_1 division leaves the entire first subordinate clause unanalysed, because the first major D_1 break comes after this clause; see Figure 23 (p. 62).

This clause cannot be divided by a D_1 accent, since it precedes the first D_1 accent in the D_0 domain. It will therefore be divided by a D_2 accent, as in Figure 24 (p. 63).

The rule of Sequence of Accents also applies to D_2, so that successively smaller breaks within the domain of D_1 are marked by successive D_2 accents; see Figure 25 (p. 63).

Just as D_2 divides a D_1 phrase at the next level down, D_3 divides a D_2 phrase, as in Figures 26a–b (p. 64).

It would appear at first glance that the sequence of accents applies to D_3 since each successive D_3 accent is lower in disjunctive value in Figure 27 (p. 65).

However, there is no level lower than D_3; thus a D_3 phrase, if divided (though often it is simply not analysed any further), will be divided by a D_3 accent. Such a division runs counter to the Sequence of Accents rule, which specifically prohibits a weaker accent of a given level from preceding a stronger one. The result is that, in principle, no way exists to establish a hierarchy among consecutive D_3 accents. In practice, this is not as serious a breakdown of the system as one might at first think, simply because relatively few verses are so complex

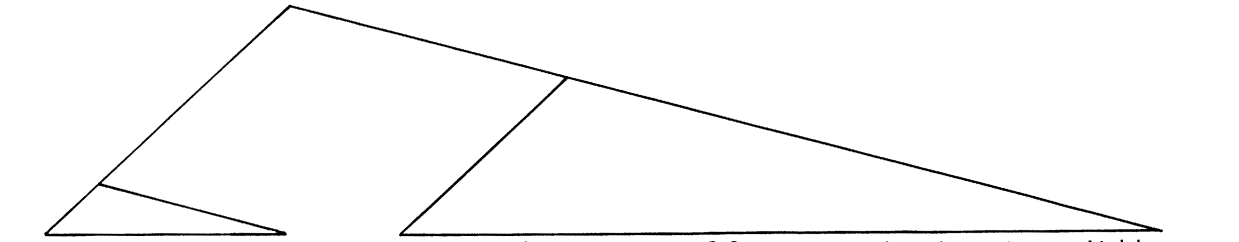


FIGURE 20a. 'Please say you are my sister, so that it goes well for me on your account and so that my soul lives because of you.' (Gen. 12.13)

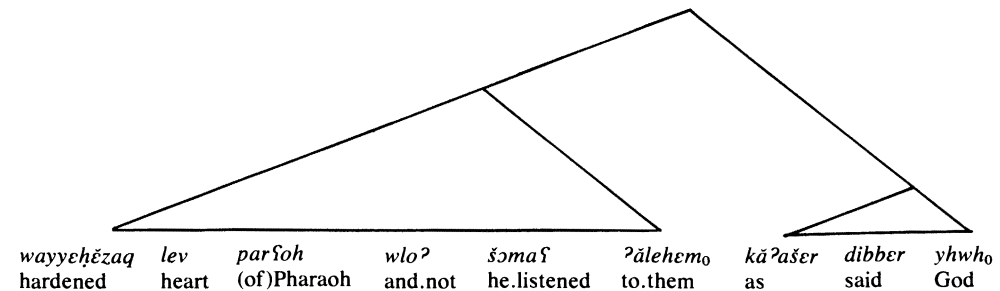


FIGURE 20b. 'The heart of Pharaoh hardened and he didn't listen to them, as God said.' (Ex. 7.13)

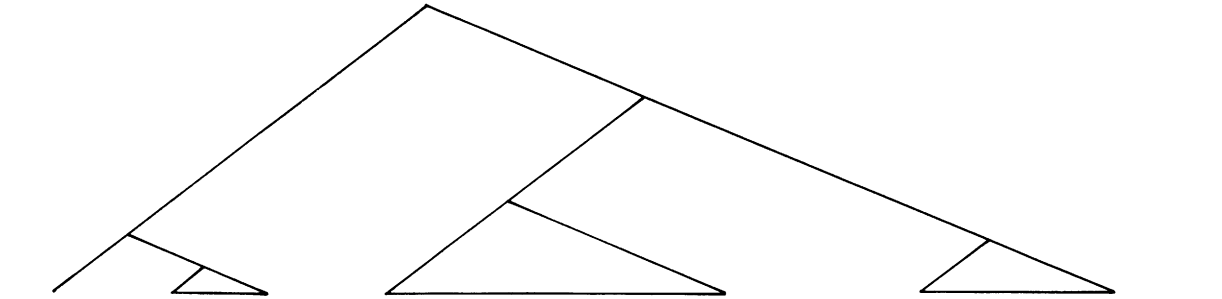


FIGURE 21a.

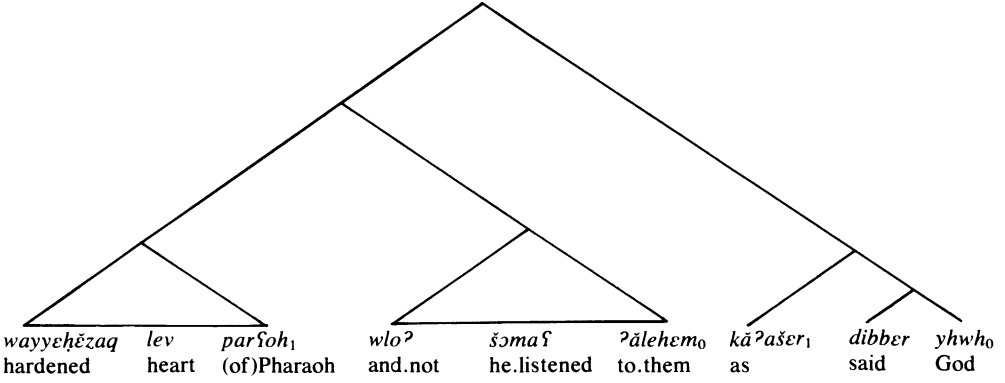


FIGURE 21b.

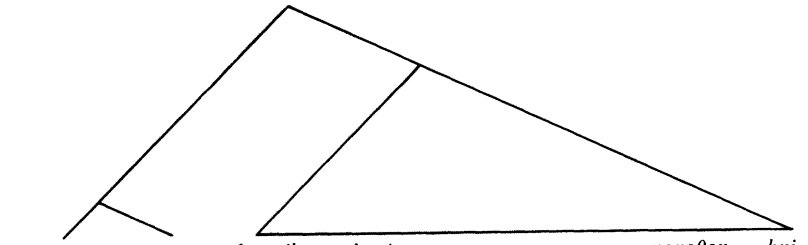


FIGURE 22a. 'And Saul said, Cast lots between me and Jonathan my son.' (1Samuel 14.42)

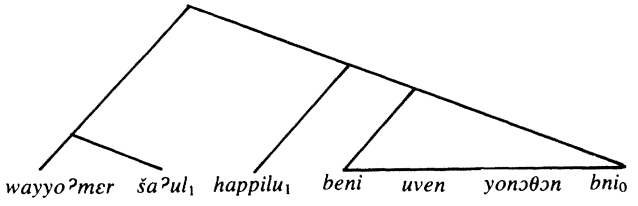


FIGURE 22b.

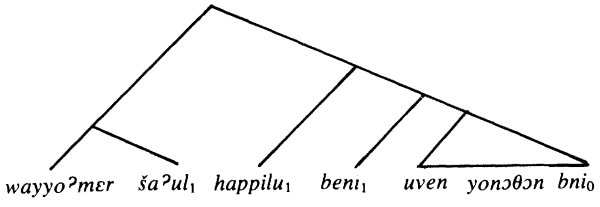


FIGURE 22c.

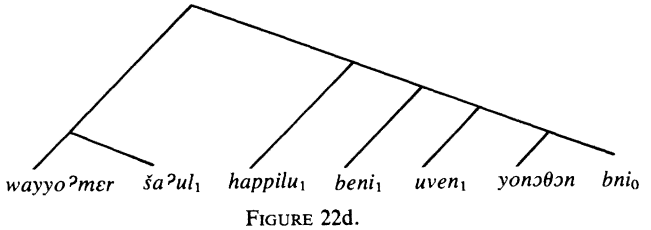


FIGURE 22d.

FIGURE 25. 'He came before God to the entrance of the tabernacle of the congregation.' (Leviticus 15.14)

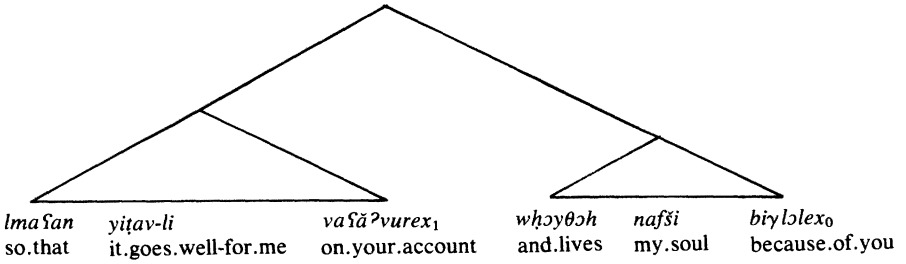


FIGURE 23.

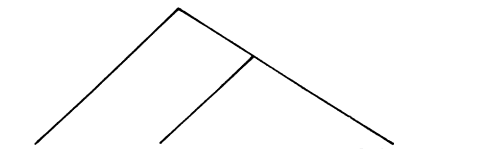


FIGURE 24.

as to require analysis of a D_3 phrase. It is also true that the internal organization of the set of D_3 accents is not nearly as clear as that of the other levels, again probably because of the relative infrequency of D_3 accents. In any case, the inadequacy of D_3 accents, though a fault in the system, is not a major one; and it does serve to highlight the virtues of the rest of the system. The failure of the sequence of accents in particular shows how well this sequence works at the D_1 and D_2 levels.

I have not mentioned the final accents of levels D_2 and D_3. There is one D_3 final; but at D_2 there are three, each occurring before a particular D_1 accent. Thus one, which we may label $A_{2f(1)}$, appears only before the non-final D_1 accent which I have labeled A_1. The second occurs only before A_{1f}, and may be labeled $A_{2f(1f)}$. The third occurs before a contextual variant of A_1. Just as with the final accent at the D_1 level, the origin of these finals is not clear; it may be musical. From the point of view of parsing, however, they serve an anticipatory function, allowing the reader to know not only that the next break is a greater one than the last, but also to know whether this next break will be followed by a yet greater break (which it will be if it is A_{1f}) or by a lesser one (if the next accent is A_1).

I will delve no deeper into the accents themselves. There is much more contextual variation than I have described, not all of which is regularly predictable. What I have described here, however, is very regular and forms the core of the system. Its main features are the four levels of dichotomous accents; the concept of domain, with a given accent occurring at the end of its domain; the sequence of accents at levels D_1 and D_2, with each succeeding accent being weaker than the last; and the distinction between final and non-final accents at a given level. The main advantage of the system, for the reader who knows it, is that it converts a hierarchical structure into a linear sequence of symbols

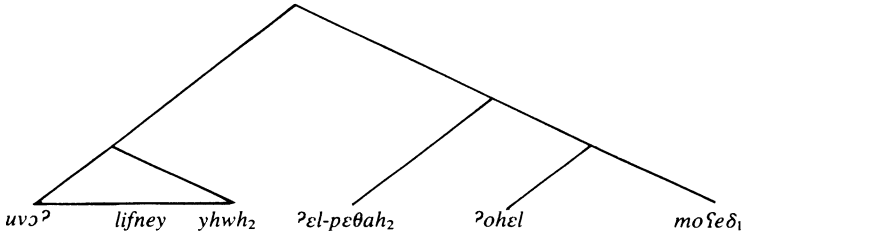


FIGURE 25. 'He came before God to the entrance of the tabernacle of the congregation.' (Leviticus 15.14)

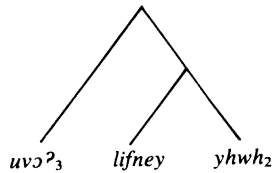


FIGURE 26a.

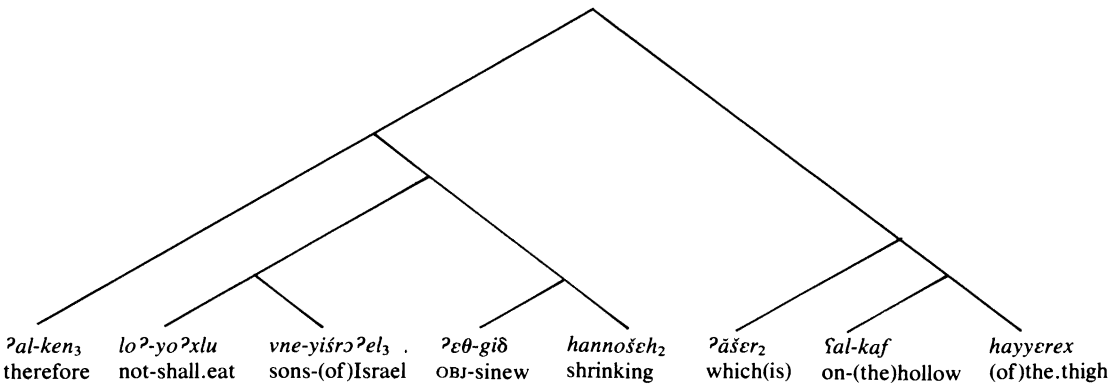


FIGURE 26b. 'Therefore, the sons of Israel shall not eat the shrinking sinew that is on the hollow of the thigh.' (Gen. 32.32)

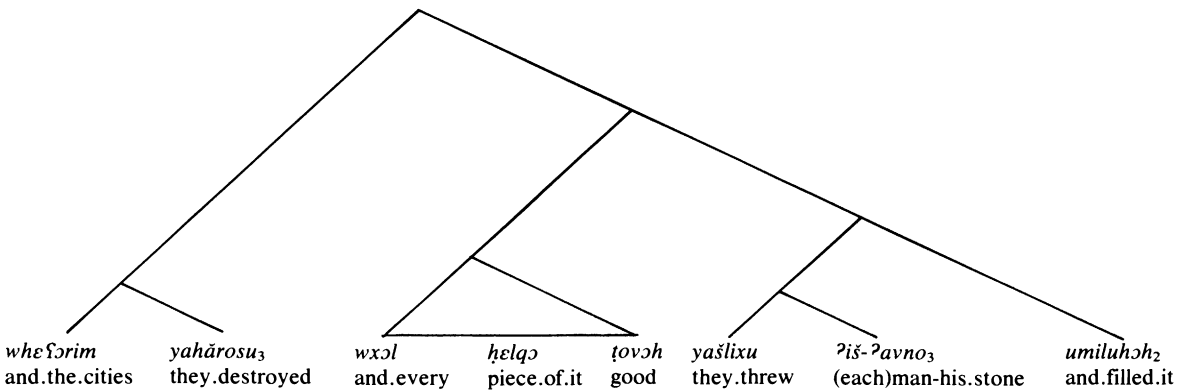


FIGURE 27. 'And they destroyed the cities and each man threw a stone on every good piece of land and filled it.' (2Kings 3.25)

whose direction conforms to that of the text. Each symbol is placed at the point in the text where the break which it represents occurs; from any given symbol, the reader may divine with a good degree of certainty the identity of the next. The advantage of this last feature, apparent to those familiar with forward-moving parsers, is that it may allow the reader to choose among competing analyses of the sequence which he has already passed, in those cases when the analysis of the preceding sequence is dependent on the nature of the following one (see Wanner & Maratsos 1978 for a simple discussion of forward-moving directional parsing systems).

6. MUSIC AND RECITATION. For what purpose did the Masoretes design such a complex accentuation system? Spanier (110) suggests three possibilities. First, the accentuation may have had a musical basis; second, it may have been designed 'to fix the relation of the words and phrases to one another, purely theoretically'; or finally, it may have been that 'the goal and sense of the accents was proper recitation'. Spanier opts for the third, but it seems to me that such a decision is hasty.

It is clear that the accentuation system was not PURELY musical—since, as Spanier observes, the possibilities of combination of the accents are narrowly fixed by syntactic principles of language. As a purely musical system, the accents would have been uninteresting, being completely determined. This does not mean, however, that the accents were not musically significant; as Rotenberg (159) points out, 'in many religious traditions, the melodic side of the scriptural reading is what is called "logogenic", i.e. more or less servile to and reflective of the syntactic and rhetorical structure of the text'.

Thus, though it is true that the distribution of the accents was not governed primarily by musical criteria, there is no reason to conclude that they had no musical value. Furthermore, as Wickes argues in some detail, there are a number of alternations which are most plausibly explained by assuming that each accent had a fixed musical value.

Nonetheless, though the accents may have had musical values, the relations among the accents did not have a musical basis. Compare in this regard the system of Gregorian chant.[23] In that system, which has a similar musical use to the accents in the recitation of the Psalms, the melodies are NOT rooted in the text, even though their application to the text is governed to a certain extent by linguistic criteria.

As regards the 'purely theoretical' value of the accents, Spanier argues that the division of the verse recorded by the accents does not always seem to correspond to logic. As I have shown above, it is indeed true that the Masoretic analysis does not always correspond closely to a modern one. In most cases, however, I have shown that the reason for the discrepancy lies precisely in the reconstructed Masoretic theory of syntax, embodied in the Parsing Principle, which ignored the semantic relations of constituents to one another. These intuitively unsatisfying analyses, however, reveal not that the Masoretes

[23] See Chen 1980 for an enlightening discussion of Gregorian chant from a linguistic point of view.

were illogical, but rather that they were excessively logical—permitting analyses which were theoretically valid, but did not correspond to common-sense expectations.[24] This excessive logic indicates that the basic motivation for the accents, and the principles according to which they are distributed, must have been a desire to 'fix the relation of the words to one another'.

This is not to deny that the accents were used as a guide to the recitation of the text. It is quite clear that they had phonological implications which would have been reasonable only if they were treated as a guide to recitation. Furthermore, we know that a fairly rigorous tradition of recitation existed long before the written accent system came into being, and that the accents supplanted and augmented this tradition. But though the accents governed recitation, even in fairly detailed aspects, it must be the case that they were based on syntax, and that their relation to recitation was similar to their relation to music. To put it another way, suppose that there was a traditional way of reciting the text, and that this tradition was fairly rigorously defined, down to the most minute details of segmental phonology and of intonation. One could devise a system which would record all this; in fact, the Sanskrit system comes close, and certain aspects of the Masoretic segmental notation are certainly designed with such a purpose in mind. This system would be related to syntax insofar as intonation and sandhi are a function of syntactic structure (Selkirk 1980, Rotenberg 1978); but the relationship would be fortuitous. My point is that the Masoretic accents do not represent such a direct codification of a traditional recitation.

What evidence exists for this assertion? First, there is the top-down organization of the system, especially the fact that the primitive unit of this organization is the Biblical verse—a linguistically arbitrary unit. As Wickes emphasizes, there is no correspondence between individual accents and individual syntactic constructions; the accents are purely relational. The D_0 accents mark the division of the verse; and no matter whether a verse consists of three sentences or of a single NP, the main division of that verse will be marked by a particular accent—similarly for D_1, D_2, and D_3. On the assumption that each accent had a fixed intonational value, this particular distribution could serve only to obscure the natural intonation of the text, since there is no simple one-

[24] A referee for *Lg.* notes that the Masoretes, not being native speakers of Hebrew—a language then several centuries dead (Kutscher 1982)—did not have the same intuitions about syntactic structure that a native might; thus we should not expect them to be able to intuit properties of a language not their own. From this point of view, says the referee, the results of this article are unremarkable, and are unlikely to generalize to cases of native speakers who devise their own orthographies. There are several problems with this objection. First, the Masoretes were not decoding a long-dead language, but rather devising a means for parsing a text which had been in continuous use for centuries. Second, the Masoretes were native speakers of a closely related language, Aramaic, and were also familiar with Mishnaic Hebrew as a language of learned discourse. Finally, there is no reason to believe that any of us have privileged access to theoretically interesting syntactic or even morphophonemic properties of any language, including our own. If we did, linguists would have little reason to be in business. There is a much larger hermeneutic question at issue about the relation between theory and data in human science; but it is precisely a corner of that question that I am investigating here.

to-one correspondence between accents and the syntactic units with which intonation is correlated. This lack of correspondence is only aggravated in modern recitation, where there is a tendency for the rarer accents to be more prominent—with the result that, the greater the syntactic break, the less distinguished it is musically. The Masoretic system is not merely inefficient from a prosodic point of view, it is downright confusing; however, from a purely syntactic point of view, as a way of indicating the relationships among the parts, it is perfectly reasonable.

The actual treatment of sandhi is also instructive. In Masoretic Hebrew, the simple stops *p t k b d g* become fricatives postvocalically ('aspiration'; for details, see Gesenius 1910; for analysis, see Prince). This is indicated in the Masoretic system by means of *dageš*: stops have it, fricatives do not. When a word ends in a vowel, and the initial segment of the following word is a simple stop, this stop is realized as a stop after a disjunctive accent, but as a fricative after a conjunctive accent. One should take this as evidence of the close relationship between the accents and the recitation of the text. Note, however, that the actual syntactic relationship between the two words, the first ending in a vowel and the second beginning in a consonant, is irrelevant for the operation of aspiration. In fact, pairs exist in which the syntactic relationship is identical, but in which, because of the top-down distribution of the accents, a disjunctive accent appears in one case and a conjunctive one in the other. In every case, the distribution of the segments follows the accentuation rather than the actual syntax. The following near-minimal pair is cited by Rotenberg:

(34) a. *wayyišʔălu₂ bney yiśrɔʔel₁ bayhwh*
asked sons Israel in.God
'the sons of Israel asked God' (Judges 1.1)

b. *wayyillɔḥămu vney-yhuδɔh₂ birušɔlaym*
fought sons-Judah in.Jerusalem
'the sons of Judah fought against Jerusalem' (Judges 1.8)

The two examples have the same syntactic pattern. They are accented differently because there is a hyphen between the two words of the construct phrase in 34b, but none in 34a. I noted above that the distribution of hyphens is prior to the accent system, and that it is not consistent. However, the presence of the hyphen in 34b means that there will be no disjunctive accent between the verb and the subject, since they form a two-word phrase. In 34a, since there is no hyphen, the subject is treated as two words, and a disjunctive accent appears between the subject and the verb. The result is that the *b* of 34a is a stop, while that of 34b is a fricative. Sandhi is therefore dependent on accent, not on syntactic or phonological structure. We must conclude that this particular detail of pronunciation (spirantization across words) was governed by the accents and not vice versa. The accents can therefore not be a record of actual recitation; rather, recitation must have come to be based on the accents, at least by the time of the earliest surviving codices (ca. 1000).[25]

[25] There are additional sandhi rules, including initial gemination and a variety of stress retraction rules that are sensitive to the conjunctive/disjunctive accent distinction (McCarthy 1979). The

One might claim that the syntactic break in 34a–b, between verb and subject, is minor enough so that aspiration may have been optional. In the following example, the break is between a topicalized Prepositional Phrase and the main verb. It is the strongest such break in terms of both modern analysis and the Masoretic analysis. But the sentence consists of only two words; thus there is no disjunction. Aspiration therefore applies across this major break:

(35) *ʕal-gḥonxɔ θelex*
on-your.belly you.will.go (Gen. 3.14)

There is a late rule whereby certain occurrences of the D_3 accent *gereš* are changed into conjunctives. The rule is called a 'transformation' by Wickes because the accents preceding it pattern exactly as they would if the original *gereš* were actually present; the simplest description assumes that first *gereš* is assigned, then the accents preceding it are assigned, and finally *gereš* is transformed into a conjunctive (100, 117). In these cases, where an original disjunctive is transformed, the aspiration follows the accent which is found on the surface, i.e. the conjunctive. We therefore find aspiration in the following example:

(36) *hevi^{ʔ} γam-hu^{ʔ}*
brought also-him

Why *gereš* should be transformed in these cases is not clear, but the reason is most likely musical. The fact that aspiration follows the transformation, however, is yet another piece of evidence that accentuation governs recitation rather than vice versa.

The other major sandhi phenomenon which can be related to the accents is PAUSAL LENGTHENING. A variety of phonological changes—most, but not all, involving lengthening of one sort or another—take place in pausal forms. For example, a short stressed *a* will be raised to *ɔ*, a reflex of long /a/: *kɔtav* 'he wrote' becomes *kɔtɔv*. Lengthening toward the end of the word is expected on general phonetic grounds (Devine & Stephens 1980), so there is no doubt that the phenomenon has a real basis. Pausal forms occur only before disjunctive accents. According to Gesenius (96), pausal forms occur 'in the last word of a sentence (verse) or clause'. They also are usually marked by a D_0 accent, but may be marked by a lesser accent. From our point of view, the important question is whether the pausal forms are triggered by the accents or by the syntactic position. Unfortunately, this question is not easily answered—partly because the pausal forms are, by and large, optional, and partly because the question is at root a statistical one which cannot be given a reasonably accurate answer without immense labor. I will therefore leave the question as highly relevant but unanswered for now.

---

complexity of these phenomena and their interaction with the rest of the phonology indicate that they are not late accretions to the language. It is therefore reasonable to assume, with McCarthy and with Rotenberg, that these rules were originally conditioned by the syntax of the language. What I am trying to demonstrate is that the Masoretes, in their codification, replaced the original linguistic conditions with more theoretically explicit, albeit linguistically inaccurate, ones. I am not questioning the status of the rules.

I have tried to demonstrate that the accents are primary and govern the recitation of the text, rather than vice versa. The practical difference between the two possibilities is not great. In terms of my own thesis, however, the difference is important. If the accents were simply a record of a traditional recitation, and not at base a purely theoretical notation, they could be dismissed as yet another phonetic record, and the entire system would be in the end only a curio.

7. CONCLUSION. I have shown that the accentual system of Tiberian Masoretic Hebrew is based on a theory of syntactic analysis. From a general point of view, the goal of this demonstration is to exemplify the close connection which exists between orthography and linguistic analysis. I do not wish to identify the two. Both depend on the objectification of language, yet they differ in function. Orthography is a communication device, which depicts language in order to transmit it, while linguistic analysis has a more abstract purpose, that of science or understanding. But as a recording device, orthography differs from the merely mechanical: it has a human component at its core. That humans cannot depict without understanding is a commonplace, and my attempt has been to take this commonplace beyond the visual in as rigorous a fashion as possible.

REFERENCES

ANSHEN, FRANK, and PETER SCHREIBER. 1968. A focus transformation of Modern Standard Arabic. Lg. 44.792–7.

BANFIELD, ANN. 1973. Narrative style and the grammar of direct and indirect speech. Foundations of Language 10.1–39.

BREUER, MORDECAI. 1958. Pisuq hateʕamim šebamiqra. Jerusalem: Hahistadrut Hacionit.

CHEN, MATTHEW. 1980. Gregorian chant and Latin prosody. University of California, San Diego, MS.

CHOMSKY, NOAM. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

——. 1970. Remarks on nominalization. Readings in English transformational grammar, ed. by Roderick Jacobs & Peter Rosenbaum, 184–221. Waltham, MA: Ginn.

CHOMSKY, WILLIAM. 1952. David Kimchi's Hebrew grammar: Mikhlol. New York: Bloch.

COHEN, MILES B. 1969. The system of accentuation in the Hebrew Bible. Minneapolis: Milco.

CROSS, FRANK M., and DAVID N. FREEDMAN. 1952. Early Hebrew orthography. New Haven: American Oriental Society.

DEVINE, ANDREW M., and LAURENCE D. STEPHENS. 1980. On the phonological definition of boundaries. Juncture, ed. by M. Aronoff & Mary Louise Kean, 57–78. Saratoga, CA: Anma Libri.

DOTAN, ARON. 1970. Two treatises on the accentuation of the Old Testament: Prolegomenon. New York: Ktav.

DRESHER, B. ELAN. 1981a. Metrical structure and secondary stress in Tiberian Hebrew. Brown University Working Papers in Linguistics 4.24–37.

——. 1981b. Accentuation and metrical structure in Tiberian Hebrew. MIT Working Papers in Linguistics 3.180–208.

ELLIGER, KARL, and WILHELM RUDOLPH. 1977 (eds.) Biblia Hebraica Stuttgartensia. Stuttgart: Deutsche Bibelstiftung.

FRANCIS, W. NELSON. 1958. The structure of American English. New York: Ronald Press.

FREEDMAN, DAVID N. 1962. The Masoretic text and the Qumran Scrolls: A study in orthography. Textus 2.87–102. [Reprinted in Qumran and the history of the Biblical text, ed. by Frank M. Cross & S. Talmon, 196–211. Cambridge, MA: Harvard University Press, 1975.]

GESENIUS, WILHELM. 1910. Hebrew grammar, as edited and enlarged by the late E. Kautzsch. Second English edition revised in accordance with the 28th German edition. Oxford: University Press.

GREENBERG, JOSEPH. 1966. Some universals of grammar with particular reference to the order of meaningful elements. Universals of language, 2nd edn., ed. by J. Greenberg, 73–113. Cambridge, MA: MIT Press.

HAAS, W. 1976. Writings: The basic options. Writing without letters, ed. by W. Haas, 131–208. Manchester: Manchester University Press.

HALE, KENNETH L. 1983. Warlpiri and the grammar of non-configurational languages. Natural Language and Linguistic Theory 1.5–47.

HARRIS, ZELLIG S. 1951. Methods in structural linguistics. Chicago: University of Chicago Press.

IDELSOHN, ABRAHAM J. 1929. Jewish music in its historical development. New York: Holt.

JACKENDOFF, RAY S. 1978. $\bar{X}$ syntax: A study of phrase structure. (LI monograph 2.) Cambridge, MA: MIT Press.

JUSTESON, JOHN S. 1976. Universals of language and universals of writing. Linguistic studies offered to Joseph Greenberg, ed. by Alphonse Juilland, 1.54–94. Saratoga, CA: Anma Libri.

KAHLE, PAUL. 1913. Masoreten des Ostens. Leipzig.

——. 1927. Masoreten des Westens I. Stuttgart.

——. 1930. Masoreten des Westens II. Stuttgart.

——. 1959. The Cairo Geniza. 2nd edn. New York: Praeger.

KLIMA, EDWARD S. 1972. How alphabets might reflect language. Language by eye and by ear, ed. by J. F. Kavanagh & Ignatius Mattingly, 57–80. Cambridge, MA: MIT Press.

KUTSCHER, EDUARD Y. 1982. A history of the Hebrew language. Jerusalem: Magnes.

MCCARTHY, JOHN. 1979. Formal problems in Semitic phonology and morphology. MIT dissertation.

MCCAWLEY, JAMES. 1976. Grammar and meaning. New York: Academic Press.

MALONE, JOSEPH. 1975. Systematic vs. autonomous phonemics and the Hebrew grapheme dagesh. Afroasiatic Linguistics 2:7.113–29.

MATTINGLY, IGNATIUS G. 1972. Reading, the linguistic process, and linguistic awareness. Language by eye and by ear, ed. by J. F. Kavanagh & I. G. Mattingly, 133–48. Cambridge, MA: MIT Press.

O'CONNOR, MICHAEL. 1980. Hebrew verse structure. Winona Lake, IN: Eisenbrauns.

PRINCE, ALAN S. 1975. The phonology and morphology of Tiberian Hebrew. MIT dissertation.

REVELL, E. J. 1970. Studies in the Palestinian vocalization of Hebrew. Essays on the ancient Semitic world, ed. by John W. Wevers & Donald B. Redford, 51–100. Toronto: University of Toronto Press.

ROTENBERG, JOEL. 1978. The syntax of phonology. MIT dissertation.

SCHRAMM, GENE M. 1964. The graphemes of Tiberian Hebrew. (University of California studies, Near Eastern studies, 2.) Berkeley & Los Angeles: University of California Press.

SELKIRK, ELISABETH O. 1980. Prosodic domains in phonology: Sanskrit revisited. Juncture, ed. by M. Aronoff & Mary Louise Kean, 107–30. Saratoga, CA: Anma Libri.

SPANIER, ARTHUR. 1927. Die Massoretischen Akzente. Berlin: Akademie-Verlag.

TALMON, SHEMARYAHU. 1964. Aspects of the textual transmission of the Bible in the light of Qumran manuscripts. Textus 4.95–132. [Reprinted in Qumran and the history of the Biblical text, ed. by Frank M. Cross & S. Talmon, 226–63. Cambridge, MA: Harvard University Press, 1975.]

VACHEK, JOZEF. 1973. Written language. The Hague: Mouton.
WANNER, ERIC, and MICHAEL MARATSOS. 1978. An ATN approach to comprehension. Linguistic theory and psychological reality, ed. by Morris Halle et al., 119–61. Cambridge, MA: MIT Press.
WICKES, WILLIAM. 1881. A treatise on the accentuation of the three so-called poetical books of the Old Testament: Psalms, Proverbs, and Job. Oxford: Clarendon Press.
——. 1887. A treatise on the accentuation of the twenty-one so-called prose books of the Old Testament. Oxford: Clarendon Press.
WILLIAMS, EDWIN S. 1975. Small clauses in English. Syntax and semantics 4, ed. by John Kimball, 249–73. New York: Academic Press.

[Received 5 October 1983;
revision received 3 March 1984;
accepted 18 June 1984.]