Text segmentation and linguistic levels.
Preparing data for SESB

Version 5, Oct., 2003

Eep Talstra
Vrije Universiteit - Amsterdam
Faculty of Theology - Werkgroep Informatica

1. Linguistic data beyond word level?

One question for many years already has been dominant in the research of the *'Werkgroep Informatica'*[1] in its preparation of linguistically analysed text data for projects such as the Stuttgart Electronic Study Bible (SESB). It is the question how to get beyond word-level in the linguistic analysis of Hebrew and Aramaic texts of the Bible.[2] It may sound as an relatively easy task: once you have made a database of the Bible with an analysis of all its Hebrew and Aramaic words, simply proceed by analysing phrases, subjects, predications, sentences, etc. However, in trying to do so, one rapidly finds out two things. First, when trying to perform computer-assisted syntactic analysis, one immediately detects how great the variety of the linguistic data is. One comes accross very complex sentences, such as in Deuteronomy, very dense constructions as in Psalms or Job, elliptic constructions, recursivity in the realisation of grammatical constructions, and much more. Second, one becomes much more aware of the enormous variation in linguistic and literary interest among biblical scholars. They want to study rhetorical features of text, semantical connections and associations, syntactic patterns, morphological system, textual participants, etc. This means that the suggestion to move on and produce computer-readable, linguistically analyzed texts beyond word level, implies at the same time a challenge to find an appropriate model of linguistic description and a challenge to meet the categories used by translators and exegetes in their work with texts and textual structures.
Linguists are interested in a database that allows them to search for features of Hebrew morphology, patterns of word order in prose or poetry, or the use of verbal forms in main clauses and dependent clauses. Translators need that information too but add to it their interest in patterns of verbal valency[3], or

[1] *'Werkgroep Informatica'* is a research group of the Faculty of Theology, *Vrije Universiteit*, Amsterdam. The group concentrates on research in linguistic analysis of the Hebrew Bible and Methods of Old Testament exegesis. An important instrument is the design and development of data bases of Hebrew and Aramaic text. In cooperation with the Peshitta Institute of the University of Leiden contributes to a similar project on the text of the Syriac Version.

[2] E. Talstra, 'Desk and Discipline. The Impact of Computers on the Study of the Bible'. Opening Adress of the 4th AIBI Conference, in: *Proceedings of the Fourth International Colloquium Bible and Computer: Desk and Discipline, in Amsterdam August 15-18 1994*, Paris/Geneva, 1995, p 25 - 43.

[3] E. Talstra, 'Texts for Recitation', in: *Unless some one guide me ... . Festschrift Karel A. Deurloo* (ACEBT Suppl.2), Maastricht: Shaker, 2001, 67-76.

patterns of nominalisation and renominalisation in prose texts. Exegetes share the interest in all these questions, but they add questions of word frequency in particular texts or books, questions of participants present in a particular text, or questions of idiom belonging to particular books or traditions.

Once aware of the great variety of language-oriented questions, one also begins to understand that most of the questions biblical scholars ask in their daily work, usually are characterized already by a high level of abstraction. For example, could I find those clauses where God (אל) is the subject? Could one collect those cases of direct speech that begin with the conjunction 'and' (ו) or 'since' (כי) ? Who are the participants in particular texts? What are the narrative sections and what are the direct speech sections of a composition? The commonly used Bible search programmes being word-oriented are basically incapable of collecting data of these higher levels of linguistic analysis. Even when it is clear that the SESB project will not yet be capable of answering all those questions, its data are being produced in a research project that already allows for computer guided proposals to the majority of the sample questions mentioned here.

In order to clarify what the features are of the Hebrew data base underlying the SESB, this introduction to the Hebrew data used in SESB in the first place will analyse what kind of questions asked by exegetes and translators invited the research group at the *'Vrije Universiteit'* in Amsterdam to start the project of computer-assisted biblical research. Secondly this introduction reports what type of text grammatical research lead to the production of the linguistically analyzed Hebrew text data base that is used in SESB (§ 2). Thirdly, the closing section of this introduction will give some examples of the type of linguistic searching and textual research that may benefit from this project (§ 3).

1.1. Some initial questions

The first two verses from Genesis chapter 22 may serve here as introductory materials for some basic questions in the area of Bible translation and biblical scholarship.

Genesis 22:1		
	After this	וַיְהִי אַחַר הַדְּבָרִים הָאֵלֶּה
	God has tested Abraham.	וְהָאֱלֹהִים נִסָּה אֶת אַבְרָהָם
	He said to him:	וַיֹּאמֶר אֵלָיו
	'Abraham!'	אַבְרָהָם
	He said:	וַיֹּאמֶר
	'Here I am.'	הִנֵּנִי :
Genesis 22:2		
	He said:	וַיֹּאמֶר
	'You should take your son, your only one, whom you love, Isaac	
		קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק

and go to the land of Moria וְלֶךְ לְךָ אֶל אֶרֶץ הַמֹּרִיָּה
and offer him there as a burnt offering on one of the mountains
וְהַעֲלֵהוּ שָׁם לְעֹלָה לאַחַד הֶהָרִים
that I will tell you.' אֲשֶׁר אֹמַר אֵלֶיךָ׃

Any student of the Hebrew Bible sooner or later experiences the fact that in trying to understand the system of language used in textual compositions one needs linguistic information beyond the level of collecting and sorting of individual words. Of course, the analysis begins with words. A translator may wish to collect more cases of the word הִנְנִי ('it is me') and ask questions about its rendering in a modern bible translation. But mostly one needs to understand more of certain combinations of words, special idioms, grammatical constructions or the arrangement of particular clauses in a text. So the general task is, could one make an inventory of the linguistic materials beyond word level when trying to solve particular questions the exegete and the bible translator ask? Some examples. Scholars may want to know examples of:

1.1.1. Co-occurrence of words

Usually computer databases provide their users with options to search for words and combinations of words in the context of a verse, such as 'son' (בן) and 'love' (אהב) in verse 2. See Genesis 37:3, Deuteronomy 21:15, Hosea 11:1, Ruth 4:15. This way of collecting material is similar to the way one uses traditional concordances.

1.1.2. Combinations of words and grammatical features.

As a next step computer databases also offer possibilities for the searching of combinations of grammatical and lexical data, such as the verb 'bring up' (עלה in hiph'il) + 'burnt offering' (the noun עֹלָה). See: Genesis 8:20, Judges 6:26, 11:31, 2Samuel 24:25, 2Kings 3:27. Frequent users of text data bases will have experienced the difficulties that are related to examples of this type. If one wants to list cases of עלה in hiph'il + the noun עלה, one faces the problem of context. It is not always helpful just to collect all verses that use the words requested, for in many cases they do not appear as part of one predication frame. Therefore we need a data base that 'knows' where clauses begin and end.

1.1.3. Clause segmentation.

For more precise linguistic or exegetical research one would need a data base that has the Hebrew text segmented into 'clauses', i.e. the arrangement of words within the framework of one predication. Only in that case the user could expect more adequate lists of linguistics data. Once a text data base with clause segmentation has been made, more options are present, e.g. the collection of data for lexical research in the area of verbal valency. With what preposition phrases can verbs be constructed? For example: 'to go to ..' (אל + הלך: Genesis 27:9 2Kings 3:13) or 'bring up as a burnt offering to ..' (עלה in hiph'il + pronominal suffix + ל).

1.1.4. Parsing of Constituents.

Once the computer text has been segmented into clauses, a further step should also be made possible. An exegete can be assisted greatly by information of the type: verb + subject or + object. For example, in verse 1 God is the subject of the testing. So, can one search for other cases where 'God' (אלהים) is subject of the verb "to test" נסה? See: Exodus 20:20, Deuteronomy 8:2. And, in order to be able to compare this text to others, can one also collect cases where "God" is not the subject, but the object? See: Deuteronomy 6:16, Psalm 78:56. Such questions require additional analysis of the clauses of a text. One has to combine words into the larger units of phrases and then label them in terms of 'predicate', 'subject', 'object', and so on.

1.1.5. Phrases and Clauses

The reader of Genesis 22 also detects that the world of textgrammar can be a complicated one. Thinking in terms of smaller and larger linguistic units, one would always expect words simply to be part of phrases. For example in verse 1 the words 'Abraham + Object marker' (אֶת + אַבְרָהָם) produce the object phrase 'Abraham' (אֶת אַבְרָהָם). Similarly one would always expect phrases simply to be part of clauses. Thus, the phrases

> 'Object marker+Abraham' (אֶת אַבְרָהָם [NP]) + 'has tested' (נִסָּה [VP]) + 'God' (הָאֱלֹהִים [NP]) + 'And' (וְ [CjP])

together produce the verbal clause

> וְהָאֱלֹהִים נִסָּה אֶת אַבְרָהָם 'And God has tested Abraham'.

However, on many occasions the interaction of phrases and clauses in texts is much more complicated. Phrases are part of clauses, but clauses equally well can be part of phrases, as is demonstrated in the second clause of Genesis 22:2.
The relative clause 'whom you love' (אֲשֶׁר אָהַבְתָּ) is part of the object phrase: 'your son, your only one, the one you love, Isaac' in the command given to Abraham: 'Take your son ...'. קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק
As a result of this, the relative clause 'whom you love' (אֲשֶׁר אָהַבְתָּ) should be analysed in a text data base at two levels. That is, it should be labeled as part of a phrase (the object to the verb 'take'), but also as a clause by itself, since it has a predication of its own: 'love'.
Thus, a translator who wants to study the function of אשר-clauses in narrative

prose should get from a computer search two cases from verse 2, i.e., the attributive clause ’the one I tell you’ (אֲשֶׁר אֹמַר אֵלֶיךָ), and also the small clause ’the one you love’ (אֲשֶׁר אָהַבְתָּ) that is part of the larger object ’your son, your only one ...’
Similarly, a search for Objects should produce from verse 2 also two cases, i.e. the full complex phrase:
’your son ...’ אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק
but also the pronominal suffix to the verbal form: ’you offer *him*’ וְהַעֲלֵהוּ.

1.1.6. Present a text by listing the clauses and clause types

As a next step translators and exegetes may benefit from a computer data base presenting to them a text not verse by verse, but clause by clause. Additionally one could ask to add a label indicative of the clause type based on the presence or the absence of verbal forms and the position of the subject. A listing of Genesis 22:1-2 might look like this:

וַיְהִי אַחַר הַדְּבָרִים הָאֵלֶּה	Wayyiqtol	1.a
וְהָאֱלֹהִים נִסָּה אֶת אַבְרָהָם	W-X-Qatal	1.b
וַיֹּאמֶר אֵלָיו	Wayyiqtol	1.c
אַבְרָהָם	Vocative	1.d
וַיֹּאמֶר	Wayyiqtol	1.e
הִנֵּנִי :	Nom.Clause	1.f
וַיֹּאמֶר	Wayyiqtol	2.a
קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק	Imperat.	2.b
וְלֶךְ לְךָ אֶל אֶרֶץ הַמֹּרִיָּה	Imperat.	2.c
וְהַעֲלֵהוּ שָׁם לְעֹלָה לאַחַד הֶהָרִים	Imperat.	2.d
אֲשֶׁר אֹמַר אֵלֶיךָ :	Rel.Qatal	2.e

Of course, the presentation of verse 2.b. creates a difficulty. If a clause is embedded in another clause, as is the case with ’whom you love’ (אֲשֶׁר אָהַבְתָּ), how to present them, as one clause or as two? This type of linguistic problems will be dealt with in the second part of the introduction.
A similar point is the presence of different text types, i.e. the narrative text and embedded in it the sections of direct speech. For the study of syntax, style and textual structure it is important to make such distinctions.

1.1.7. Text types

The text of Genesis 22 clearly demonstrates the interaction of narrative sections and direct speech sections. After the heading in verse 1: God has tested Abraham, the request in verse 2 with respect to Isaac is formulated by three imperatives addressed to Abraham: ’take’, ’go’, ’lift up/offer’. If one continues the reading of the chapter one will repeatedly see the use of ’to take’ and ’to go’ in the narrator’s text (verse 3 and 6). The verb ’to offer’, however, is not used until in verse 13, where the full command is executed. There these three verbs are used, the last one now being applied to the ram in stead of Isaac:

> 'Abraham went, he took the ram
> and offered it as burnt offering in stead of his son'.
> וַיֵּלֶךְ אַבְרָהָם וַיִּקַּח אֶת הָאַיִל וַיַּעֲלֵהוּ לְעֹלָה תַּחַת בְּנוֹ:

To contribute to the study of textual features a computer database should be able to present a text according to these basic text types such as 'narrative' and 'direct speech'. The Hebrew text data base used in SESB marks narrative text by 'N' and the direct speech by 'Q'. Applied to Genesis 22:1-2 a listing might look like this:

```
                 וַיְהִי אַחַר הַדְּבָרִים הָאֵלֶּה      Wayyiqtol    N     1.a
                 וְהָאֱלֹהִים נִסָּה אֶת אַבְרָהָם      W-X-Qatal    N     1.b
                                 וַיֹּאמֶר אֵלָיו      Wayyiqtol    N     1.c
----------------------------------+
                               | אַבְרָהָם         Vocative     NQ    1.d
----------------------------------+
                                       וַיֹּאמֶר      Wayyiqtol    N     1.e
----------------------------------+
                                  | הִנֵּנִי:         Nom.Cl.      NQ    1.f
----------------------------------+
                                       וַיֹּאמֶר      Wayyiqtol    N     2.a
----------------------------------+
| קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק     Imperat.     NQ    2.b
              | וְלֶךְ לְךָ אֶל אֶרֶץ הַמֹּרִיָּה         Imperat.     NQ    2.c
     | וְהַעֲלֵהוּ שָׁם לְעֹלָה לְאַחַד הֶהָרִים         Imperat.     NQ    2.d
                     | אֲשֶׁר אֹמַר אֵלֶיךָ:         Rel.Qatal    NQ    2.e
```

1.1.8. Clause connections

There is at least one other important linguistic feature that needs to be considered here. This regards the matter of text syntax. How does the text use clauses and clause types to build its structure? For example, what is the effect of the special clause type in vers 1 and what is the effect of on another special clause type in vers 14? How to translate them?
The setting of the story is made by the statement in vers 1:

> 'God has tested Abraham' וְהָאֱלֹהִים נִסָּה אֶת אַבְרָהָם

This clause has not been constructed by the frequent type Wayyiqtol ('and did X'), but by the less frequent type W-X-Qatal ('And X has done').
A first concluding statement is found in verse 14, where after Abraham's naming of the offering place again a special clause type is used in the narrator's text:

> 'that's why one says today:' אֲשֶׁר יֵאָמֵר הַיּוֹם בְּהַר יְהוָה יֵרָאֶה:

This clause is constructed by Relative + Yiqtol, a clause type that is common in direct speech section, but rather uncommon in narrative texts.
Can one collect more clauses of these particular types and find a proper way to translate them in accordance with their text level functions?
To do so it should be possible to do searches of the type:

> find in Narrative texts, clauses of the type W-X-Qatal,
> connected to a Wayyiqtol clause,

or

> find in Narrative texts, clauses of the type Relative-Yiqtol,
> connected to a Wayyiqtol clause.

1.2. The task

The examples mentioned in this paragraph may have clarified two things.
First, most of the text linguistic questions translators and exegetes have, arise at a linguistic level that goes far beyond the level of words and the co-occurrence of words. So, one needs tools that could perform more effective searching or listing, without just leaving most of the selection from the raw data to the user.
Secondly, the construction of a text data base of linguistically analysed material is not a matter of only climbing from lower level data (words) to higher level data (phrase, clauses, texts). The real challenge is the presence of iterative processes in natural language to the effect that higher level units can be embedded in lower level units: phrases can be part of higher level phrases, clauses can be part of phrases, clauses can be embedded in higher level clauses, single words can become an entire direct speech section within the narrative.
These complexities may explain why most of the computer-assisted analytical tools in literary research so far have remained restricted to word level analysis.
At the same time biblical scholars at a number of places in the world have continued experimenting, since most of the questions translators and exegetes have generate new questions for those scholars who work on the preparation of text data bases: how to produce a clause segmentation? How to analyse the syntactic connection of clauses? How to identify subjects and objects?
Sofar one thing at least has become clear: the production of a linguistically analyed text data base is not just preparing a tool. It is biblical research by itself in a very intensive way. Therefore, the next section of this introduction (§ 2) will report on the research that lays underneath the Hebrew linguistic data that are use in SESB.

2. Data and Theory: in which order?

In this section I will explain first in what manner data categories have been designed to store and analyse elements of text and language of the Hebrew Bible. The challenge was made clear by the inventory of research question in the previous section. Before we ever will be able to use a programme that can collect on request various types of linguistic and textual features, how to prepare a database that allows for searching complex linguistic data? Do we understand enough of the linguistic system of an ancient language such as biblical Hebrew to write computerprogrammes that will come up with intelligent proposals for syntactic analysis? Do we need to develop or to adopt a complete theory of linguistic description before we could begin using a computer to help us analyse texts?
Experience of the last 25 years has taught us: the best way to find out is 'just' to try it[4]. Which does not imply, of course, the naive view that one could do linguistic analysis without having any theory about categories of description. Rather, it is a matter of the appropriate order of doing the analysis. Computers are not just easy tools to execute a grammatical theory expressed in a particular set of instructions. Computers also, even better to our experience, can be used to help find a set of grammatical rules. Programmes can help detect linguistic regularity within the complexity of linguistic data, by a procedure of trial and error and checking for consistency. Therefore, when I started computer-assisted analysis of Hebrew texts, more than 25 years ago, I decided not to try to begin with the design of a set of grammatical rules, to be applied by a computer programme in performing the morphological and syntactic parsing. But from that very start and continually so in the group of the colleagues that joined me in the project, we have tried to use the Biblical texts as an area of testing proposals of syntactic parsing. Proposals for parsing a particular phrase or clause usually had an *ad hoc* basis first, since they were made on the basis of the researcher's philological knowledge and linguistic intuition. In the course of the project parsing proposals could also be made by programmes using the data derived from texts already parsed in earlier sessions. This procedure has helped us to define what types of categories and relations one should design for a proper linguistic description of an ancient language and an ancient textual corpus.

2.1. Language and Text. What is system, what is strategy? Linguistic Layers

Computer assisted analysis of biblical or ancient texts in general requires a dialogue of classical philological, exegetical methods and computer-assisted

[4] E. Talstra, 'An hierarchically structured database of Biblical Hebrew Texts. The relationship of grammar and encoding', in: *Actes du premier Colloque Internationale "Bible et Informatique: Le Texte"*, Louvain-la-Neuve, 2-4 sept. 1985. Genève, 1986, p 335-349.

design of data types to store linguistic and textual information. Can one rephrase categories used in philological interpretation into categories of segmentation, combination and function? One question, therefore, that needed to be answered in advance, before starting the actual analysis of biblical texts, was the problem of linguistic layers, or levels of analysis. Once working beyond word level, in what categories could one describe the organisation of language and text? Therefore a dialogue was needed between various disciplines: traditional philology, Hebrew linguistics and data base oriented system analysis.

2.1.1. Philology, Grammar and Computer: segmentation of texts

Traditional grammars of Hebrew philology stay close to the categories of the exegete. They speak of phonology, words and sentences.[5] Thus T. Muraoka's new edition of Joüon's Hebrew Grammar has not rearranged the materials according to analytical linguistic categories, but has left the organisation of Joüon's book as it was in the original French edition. The use of these classical categroeis make clear where the area of debate between philology and computer-assisted analysis is located. Since this introduction is on word-level and beyond, I skip matter of phonology. Morphology is dealt with here, though only to the extent it is needed to understand grammatical word functions. Syntax is the main area of discussion.

In classical Hebrew syntax usually not much is said about the way linguistic units such as phrases and clauses are built and connected. The descriptions given in most cases concentrate on syntactic functions, e.g. accusative or genitive (of phrases [6]); conditional or consecutive (of clauses [7]). General linguistics, however, has urged Hebraists to become more precise, and make further distinctions: lexemes, phrases, clauses and sentences.[8]

The actual situation implies that the concrete subdivision of a text into its constituent parts in traditional grammars usually is not dealt with as part of the domain of grammatical systematizations, but is seen as an ad-hoc decision to be left to the insights of the reader of the text.

The computer-assisted construction of a text-data base, however, cannot leave the demarcation of the syntactic textual elements to be the ad-hoc product made by the linguistic knowledge of the individual reader. One needs to be able to define exactly with what linguistic markers new phrases and clauses in a text begin, otherwise a computer program will not be able to find and analyze them. For this procedure one would need a consistent descriptive grammar of phrases

[5] Cf. P. Joüon, T. Muraoka, *A Grammar of Biblical Hebrew (Subdidia Biblica 14/I, 14/II)*, Rome: Pontificio Istituto Biblico, 1991.

[6] P. Joüon, T. Muraoka, *A Grammar of Biblical Hebrew* (Subsidia Biblica 14/1-2), Rome, 1991, § 128ff.

[7] Cf. Joüon - Muraoka, *op.cit.* § 157ff.

[8] B.K. Waltke, M. O'Connor, *An Introduction to Biblical Hebrew Syntax*, Eisenbrauns: Winona Lake, Indiana, 1990. p. 63.

and clauses. The paradox, however, is, that the entire enterprise to create a grammatically analyzed text-data base is meant to *produce* a consistent grammar of phrases, clauses and textual structure, rather then *base* itself *upon* such a grammar, simply because it does not yet exist and can only be made with the help of preliminary analyzed data. So we are moving in a circle.

The method proposed in the project preparing data for SESB, therefore, is based on the assumption that the construction of a text data base is *not* the fabrication of a tool for linguistic research, but it is *grammatical research* in itself. On the one hand the text data base is not made by applying traditional philological knowledge to all elements of the text. This would lead to a data base that only mirrors classical grammatical solutions for individual linguistic phenomena. Thus it would not be very consistent and certainly not represent independent linguistic research. On the other hand, the text data base is not made either by applying categories and results of any modern linguistic theory to all elements of a text. This would leave the analysis to be performed in the head of the researcher first and only afterwards its results being inserted into the data base [9].

It is only when one tries to let the machine perform the grammatical parsing of linguistic elements beyond word level, that one will be able to achieve any methodological advantage. These computer operations force the user to make clear what the origin is of the linguistic information the machine uses in the parsing process.

In the method presented here, the *morphological paradigm* for parsing Hebrew words is taken from traditional grammar and is then reformatted into a much more formal pattern; *lexical data* (lemmatization, part of speech) are taken from existing dictionaries, but extended with information on lexical and morphological lexeme types; *morphosyntactical* data are gathered from observations on distribution and combinations during the process of data analysis. The parsing process beyond word level, therefore, is iterative by definition.

2.2. Reading and analysing. A dialogue on 'understanding' and 'processing of information'

The examples below, taken from Genesis 23, may clarify the difference between the skills of the experienced reader of a text on the one hand and the need for using precisely defined descriptive categories, taken from existing grammars and newly made observations in the texts, on the other hand. To put it in theoretical terminology first: The Hebrew data base used in SESB makes a sharp distinction between the *distributional definition* of data, i.e. larger units are built form smaller units, and the *functional labelling* of data, i.e. the grammatical function of smaller units is defined based on its role in a larger unit. Some practical examples may

[9] Cf. W. Richter, *Biblia Hebraica transscripta (BH^t) Genesis* (ATSAT 33.1), *Exodus, Leviticus* (ATSAT 33.2), *Numeri, Deuteronomium* (ATSAT 33.3), St Ottilien, 1991.

help to clarify on what kind of experiences this distinction is based. First a simple verbal clause:

Abraham bowed down for the people of the land.

```
Gen 23,12 וַיִּשְׁתַּחוּ אַבְרָהָם לִפְנֵי עַם הָאָרֶץ:
```

The majority of the readers of this verse will have little doubts about the definition of what are to be regarded as the phrases and clauses in this verse. It has only one verbal clause, consisting of four phrases:

Conjunction - Verb (Predicate) - NP (Subject) - PP (Complement).

```
Gen 23,12 [<Cj>וַ-] [<Pr>יִּשְׁתַּחוּ] [<Su>אַבְרָהָם] [<Co>לִפְנֵי עַם הָאָרֶץ:]
```

As human readers usually perform grammatical analysis and textual interpretation in one run at the same time, they consciously or unconsciously work with routines of morphosyntactical parsing in combination with a functional, content-based concept of clauses and clause constituents. Recognition of patterns of the concatenation of words, the segmentation into clause constituents and the process of interpretation, all are performed in the same process of reading. All constituents that can be identified as Subject, Object or any type of Complement of a verbal or nominal predication are taken together as the basic elements that compose one clause. However, when using a computer to identify patterns of concatenation and segmentation and also to assign to them labels of grammatical functions, one has to separate the various routines in order to make clear how exactly the flow of information is established in the process of reading.

If applied to more complex situations, any attempt to imitate the integrated process of subdividing a text into functional units causes even more difficulties to computer-assisted textual analysis. For example, the text of Genesis 23,17-19.

```
Gen 23,17 [וַ-] [יָּקָם] [שְׂדֵה עֶפְרוֹן]
Gen 23,17 [אֲשֶׁר] [בַּ--מַּכְפֵּלָה ]
Gen 23,17 [אֲשֶׁר] [לִ-פְנֵי מַמְרֵא]
Gen 23,17 [הַ-שָּׂדֶה וְ-הַ-מְּעָרָה]
Gen 23,17 [אֲשֶׁר ] [בּוֹ]
Gen 23,17 [וְ-] [כָל הָ-עֵץ]
Gen 23,17 [אֲשֶׁר] [בַּ--שָּׂדֶה]
Gen 23,17 [אֲשֶׁר] [בְּ-כָל גְּבֻלוֹ] [סָבִיב:]
Gen 23,18 [לְ-אַבְרָהָם] [לְ-מִקְנָה] [לְ-עֵינֵי בְנֵי חֵת] [בְּ-כֹל בָּאֵי שַׁעַר עִירוֹ:]
Gen 23,19 [וְ-] [אַחֲרֵי כֵן] [קָבַר] [אַבְרָהָם] [אֶת שָׂרָה (אִשְׁתּוֹ) ]
          [אֶל מְעָרַת שְׂדֵה הַ-מַּכְפֵּלָה] [עַל פְּנֵי מַמְרֵא]
Gen 23,19 [הִוא] [חֶבְרוֹן]
Gen 23,19 [בְּ-אֶרֶץ כְּנָעַן:]
```

What are the complications?
First, one has the format of the text as a document. The traditional Massoretic division in verses and half verses does not match with a syntactical analysis of the text. The transition of verse 17 to verse 18 is made before the clause has ended. So one has to continue the reading from verse 17 into verse 18: 'And passed the field of Efron ... to Abraham ...' (וַיָּקָם שְׂדֵה עֶפְרוֹן --> 18 .. לְאַבְרָהָם). Compare the parallel construction of verse 20.
Second, the definition of the clauses in verses 17-18 itself raises questions. A number of nominal constituents is expanded by אֲשֶׁר expressions, for example in verse 17: אֲשֶׁר בְּכָל גְּבֻלוֹ סָבִיב:. So the question is: which sequences of phrases should be called 'clauses' here? The entire section of verse 17 and 18? The אֲשֶׁר expressions? Or both? If a computer program were to be asked to list or to describe the clauses in Genesis 23,17-18, what should it reproduce? List the entire section once? List the entire section and in addition also the אֲשֶׁר expressions separately?

From this situation is may be clear that is was necessary to search for an analytical system that

- is able to do justice to the Massoretic text of the Bible as a *document*: keep intact the division of books and chapters; accept the subdivision of verses and half verses, etc.
- is capable of dealing with the hierarchical grammatical organization of this document as a *linguistically structured text*: phrases, clauses, embedded clauses and clause connections should be marked as separate units;
- can be used as a *research tool* for further linguistic analysis rather than as a data base of fixed linguistic data. It should work with *distributional linguistic units*, the basic 'building blocks' to be discerned at each linguistic level of a text, in a procedure that does not have to rely on a complete grammar already being produced completely in advance. It should be capable of working with the elements of texts in a way that uses a minimum of predefined grammatical paradigms and remains open to further linguistic research and labeling.

The example of Genesis 23,17ff may explain why a computer-assisted linguistic analysis has to define more categories of linguistic description and define them more precisely than classical Hebrew grammars usually do. It appeared necessary to make a distinction between the *'building blocks'* (the distributional units, we called 'atoms') on the one hand, and the *functional units* to be defined by 'linguistic concepts' on the other. It was also clear one needs an explicit linguistic definition of the concept 'clause' if one wants to be able to decide upon the number of clauses to be found in complex texts such as in Genesis 23, 17-18. However, more preliminary linguistic procedures were found to be sufficient to decide upon the 'building blocks' that constitute in this text the level of phrase combinations, i.e. clauses. For example, in verse 17 one can make the following clause level segmentations:

וַיָּקָם שְׂדֵה עֶפְרוֹן Gen 23,17a
אֲשֶׁר בַּמַּכְפֵּלָה Gen 23,17b
אֲשֶׁר לִפְנֵי מַמְרֵא Gen 23,17c
הַשָּׂדֶה וְהַמְּעָרָה Gen 23,17d
אֲשֶׁר בּוֹ Gen 23,17e

based on a list of user-accepted patterns:

```
[ConjP:ו + VP + NP] and
[ConjP:אשר + PP]
```

These patterns can be matched with line a,b,c and e. The exception is in line 17d. But, the repeated application of the second pattern (`[ConjP:אשר + PP]`) to the text implies that the analyst has to accept the sequence `[NPdet + ConjP:ו + NPdet]` in line 17d, because it is left behind as a separate string between two אשר+PP patterns. In fact such preliminary divisions of the text into rows of phrases, based on pattern matching, constitute a *hypothesis* about the textual structure, to be verified or falsified at the next level, i.e. the level of clause combinations (to be mentioned 'clause hierarchy' later). This regards the status of line 17d. Only at the level of clause hierarchy, where these sequences of phrases are combined into higher blocks, the pattern of phrases in 17d `[NPdet + ConjP:ו + NPdet]` can be accepted as a continuation of the pattern of phrases in 17a, and eventually be redefined as an apposition to the phrase שְׂדֵה עֶפְרוֹן.

Similar to the procedure to define clause level 'building blocks' a procedure has been developed that makes it possible to construct the lower level 'building blocks', i.e. the elements used tot compose phrases. For example, in verse 17 one can define the following phrase level segmentations:

הַ שָּׂדֶה Gen 23,17d
וְ Gen 23,17d
הַ מְּעָרָה Gen 23,17d

אַבְרָהָם Gen 23,19a
אֶת־שָׂרָה Gen 23,19a
אִשְׁתּוֹ Gen 23,19a

based on a list of user-accepted patterns:

```
[ Definite Article + Noun ]
[ Conjunction ]
[ Proper Name ]
[ Preposition + Proper Name ]
[ Noun + pronominal suffix ]
```

These patterns are applied to establish the primary sequences of lexemes in the text. As a next step paradigms of combinations of such basic patterns allow for the construction of complete phrases, such as

[Def.art.+ Noun + Conj.:ו + Def.art + Noun] = NPdet
[הַ-שָּׂדֶה וְ-הַ-מְּעָרָה] Gen 23,17d

[Obj.marker-את + Proper Name + Noun + pron.sfx] = PP-את + NPapp.

[אֶת שָׂרָה (אִשְׁתּוֹ)] Gen 23,19a

In most cases, however, a *distributional unit* (a 'building block') will be identical to a *functional unit* (a phrase, for instance), e.g.

[Proper Name] [אַבְרָהָם] Gen 23,19a

The preliminary divisions made by recognition of distributional patterns create the elementary linguistic 'building stones' at a particular linguistic level. From now on these will be referred to as *'atoms'*: phrase-atoms, clause-atoms. They are called 'atoms' because they are the composing parts of the functional units at a particular linguistic level, i.e. phrases and clauses (and at a higher level: sentences). At each linguistic level 'atom' is a label that can be assigned to all grammatically acceptable units of that level that can be found by pattern recognition directly. The combination 'Def.art. + Noun' (ה-שׂרה), for instance, is used frequently as a pattern of its own in the texts. It can also be used in combination with other phrase-atoms to build one phrase (המלך דוד), but it can not be subdivided further into smaller parts without being divided into elements of a lower linguistic level, i.e. lexemes. The largest possible units built from them in a certain context by applying the paradigms of linguistic theory are labeled with the more traditional linguistic terminology of the functional type: phrase, clause, or sentence.

2.3. The production of the SESB Hebrew data

After the dialogue of philology and data analysis on the differences between 'understanding' and 'processing' some introduction is presented here into the actual steps taken in preparing the Hebrew data used in SESB.

2.3.1. An assumption to start from: a text is a regularly built linguistic piramide

At first sight linguistic analysis beyond word level seems to be possible more or less straight forward. Since phrases are composed of words, clauses are composed of phrases and sentences are composed of clauses, it seems as if we could easily use these categories to divide and analyse a text, working our way up from smaller to larger units. Of course, on will find out quickly that texts usually are not organized in such a piramide-like construction. But, adopting a trial and error stategy, we nevertheless start from that assumption. Some texts indeed fit the ideal of a piramide and they can be used to present most of the categories needed.

For example, Genesis 22, verse 1, taking the first thirteen words.

Genesis 22:1

Levels:

```
                 וַ יְהִי אַחַר הַ דְּבָרִים הָ אֵלֶּה וְ הָ אֱלֹהִים נִסָּה אֶת אַבְרָהָם

lexeme      |====| || |=| |====| | | |=| | |====| | |=| |=| |

phrase      |====.==| |=| |====.=| | |==.=.======.=.==| |=| |

clause      |========.===.=======.=| |=================.==.=|

sentence    |======================.======================|
```

Proceding from lexeme level to sentence level, what are the tasks to perform in order to built a linguistic text data base?[10]

The*first* task is a morpological analysis of words, in order to find their respective lexemes (e.g. דְּבָרִים is a form of the lexeme דבר) and to establish for each word what grammatical functions are determined by its morphological features (e.g. the ending ים determines the function 'plural' as a feature of דָּבָר). See the overview presented below. For that reason the very first task completed by F. Postma and A.J.C. Verheij a number of years ago, has been to produce a morphologically analysed Hebrew and Aramaic text of the Old Testament. [11]

The *second* task, closely connected to morphological analysis is the lexical analysis: searching a lexicon, to establish the part of speech for each lexeme. The result of the first and second task is: *lexeme* level data.

> A question pertaining at this level of research: in a considerable number of cases it is important to allow for a change of part of speech, due to morpho-syntactic conditions. For example, the lexeme אַחַר (back side) in the lexicon has a primary part of speech: noun. In constructions such as here in Genesis 22,1, it appears that in this particular phrase it functions as a preposition ('after'). To be able to store these data each word in principle has a primary, lexical part of speech (to be derived from the lexicon) and a secondary, morphosyntactic part of speech, resulting from syntactic patterns used in a text. In about 90 % of the words these two part of speech labels will be identical.

[10] C. Hardmeier, E. Talstra, Sprachgestalt und Sinngehalt. Wege zu neuen Instrumenten der computergestützten Textwahrnehmung, *ZAW (Zeitschrift für die alttest. Wissenschaft)* 101 (1989) 408 - 428.

[11] A.J.C. Verheij, *Grammatica Digitalis I. The Morphological Code in the "Werkgroep Informatica" Computer Text of the Hebrew Bible* (Applicatio 11), Amsterdam: Vrije Universiteit Publishers, 1994.

The *third* task is to regroup lexemes into grammatically acceptable strings: i.e. phrases. For example, the arrangement 'article ה' + Name, constitutes a phrase of two elements: הָאֱלֹהִים. The arrangement 'preposition', 'def.article','noun [plur]', 'def.article', 'dem.pronoun [plur]', constitutes a phrase of five elements: אַחַר הַדְּבָרִים הָאֵלֶּה. The result of this operation is a division of the words listed into six phrases. A programme to perform his task has been developed. It searches a list of user-made patterns and lists of morhological and lexical conditions to may influence the construction of a phrase (e.g. whether a pronominal suffix would be acceptable or not). Patterns accepted by the data producer are stored in the list, to be re-used in the analysis of the next texts.
For example:

noun (=אַחַר)+DefArt+noun(=plur.)+DefArt+PronDem(+plur.) = Prep. Phr.

An additional task, not presented separately here, is the internal parsing of phrases. For example, genetive relations (=relations of "regens" and "rectum") in noun phrases, attributive relations, demonstrative relations. Compare the phrase אַחַר הַדְּבָרִים הָאֵלֶּה.
It can be subdvided further into: [((הָאֵלֶּה) <dem> (הַדְּבָרִים)) אַחַר].
The result of the third task is: *phrase* level data.

> A particular question to be answered at this level of research: how to deal with the sequence of ו + Yiqtol, to be read as one verbal form: Wayyiqtol? In fact he combination of the words וַיְהִי produce one phrase. Some would even claim: one word. For practical reasons in such cases the two words are kept separate, since we need the conjunction "ו" at the next level: It does double duty, for it is part fo the verbal form, but at the same time it also the the opening signal of the clause. In syntactic analysis one cannot miss that point.

The *fourth* task is to regroup phrases into clauses. This is done in a way similar to the production of phrases. Certain patterns of phrase sequences constitute particular clauses. Once a producer of clause data has accepted a particular arrengement of phrases as a clause, it is stored in a list, to be used in further production. An important check is found in the way conjuntions behave. For example, if the phrases left and right to the conjunction "ו" are of a different grammatical type, one can assume that this "ו" functions as a conjunction that starts a new clause. Genesis 22: 1 הָאֱלֹהִים (NP) וְ אַחַר הַדְּבָרִים הָאֵלֶּה (PP).
An additional task at this level is the parsing of phrases as clause constituents. For example, the first clause of verse 1 has three phrases: conjunction (phrase) - VP (verbal phrase) and PP (prepositional phrase). The conjunction functions as an introduction to the clause, so it does not need further labeling: <Cj>. The VP

is the verbal predication: <Pr>; the PP is a time referecne: <Ti>.[12]
The result of the fourth task is: *clause* level data.

The results of the data production in task 1 - 4

Genesis 22:1

	(4)	(3)			(2)	(1)	
Way0	Cj	CjP	–	ו	conj.	וַ	וַ
	Pred	VP	ipfc.3ms	היה	verb	יְ/הִי(ה)/ø	יְהִי
	Time	PP	sing.	אחר	noun	אַחַר/ø	אַחַר
			–	ה	def.Art.	הַ	הַ
			plur.ms	דבר	noun	דְּבָר/ים	דְּבָרִים
			–	ה	def.Art.	הָ	הָ
			plur.	אלה	pr.dem.	אֵלֶּה	אֵלֶּה
WXQt	Cj	CjP	–	ו	conj.	וְ	וְ
	Subj	NP	–	ה	def.Art.	הָ	הָ
			plur.	אלהים	noun	אֱלֹה/ים	אֱלֹהִים
	Pred	VP	pf.3ms	נסה	verb	נִסָּה	נִסָּה
	Obj	NP	–	את	prep.	אֶת	אֶת
			–	אברהם	Name	אַבְרָהָם	אַבְרָהָם

The *fifth* task is[13] to combine clauses into sentences. This implies that one has to establish a number of different positions clauses may take in a text. It is possible that clauses only represent a part of a phrase in a higher level clause; they also may take the position of a phrase (a constituent) in a higher level clause or they may be connected to higher levels clauses in parallel or in dependent constructions. At the moment research to construct programmes that come up with linguistically meaningful analytical proposals is going on and is as yet far from completed. For practical reasons the SESB data use a very restricted definition of sentences: only where relative clauses or infinitive clauses are used to expand parts of a main clause, these clause combinations are called 'sentences'. The

[12] Janet W. Dyk and Eep Talstra, 'Paradigmatic and Syntagmatic Features in Identifying Subject and Predicate in Nominal Clauses', in: Cynthia L. Miller (ed.), *The verbless Clause in Biblical Hebrew. Linguistic Approaches. [LSAWS: Linguistic Studies in Ancient West Semitic, Volume 1]*, Winona Lake: Eisenbrauns, 1999, p. 133- 185.

[13] E. Talstra, 'A Hierarchy of Clauses in Biblical Hebrew Narrative', in: E.J. Van Wolde (ed.), *Narrative Syntax and the Hebrew Bible. Papers of the Tilburg Conference 1996 (Biblical Interpretation Series 29)*, Leiden: Brill, 1997, p. 85-118.

analysis of other combinations of clauses is postponed to a later phrase of data production. In the data the connections between clauses are only marked by a code that indicates the relationship. If principle then code consists of a figure for the conjunction used (e.g. "3" = וַ), a figure for the verbal form used (e.g. "2" = Qatal [affirmative conjugation]) and a figure for the verbal form in the "mother clause", i.e. gouverning clause (e.g. "7" = wayyiqtol [preformative conjugation]. Thus "327" would mean: a clause with conjunction וַ and verbal form wayyiqtol, refers back to a gouverning clause with verbal form Qatal. In this way continuing syntactic research is made possible, independent from any grammatical labelling I might wish to give to particular clause connections. One can search them by using the codes.
The result of the fifth task is: *sentence* level and *text* level data.

The results of the data production in task 5

Genesis 22:1 text of clauses type line clause connection codes

```
          [<Ti>אַחַר הַדְּבָרִים הָאֵלֶּה] [<Pr>יְהִי] [<Cj>וַ-] Way0  <+1 327><+2 200>
[<Ob>אֶת אַבְרָהָם] [<Pr>נִסָּה] [<Su>הָאֱלֹהִים] [<Cj>וְ-] WXQt  <-1 327>
                      [<Co>אֵלָיו] [<Pr>יֹּאמֶר] [<Cj>וַ-] Way0  <-2 200><+2 200><+1 999>
                                      [<Vo>אַבְרָהָם] Voct  <-1 999>
                               [<Pr>יֹּאמֶר] [<Cj>וַ-] Way0  <-2 200><+1 999>
                                       [<Is>הִנֵּנִי׃] NmCl  <-1 999>
```

327: conjunction "W"; Qatal; Wayyiqtol
200: coordination of identical clause types
999: clause connection start a direct speech section

At this level of the computer assisted analysis it is very important to experiment with various approaches of language and text. Thus this work is much more a matter of experimental research than a matter of data production. The research regards especially attempts to find more insight in the interaction of a formal, distributional approach and functional, pragmatic models of linguistic analysis.[14]

2.3.2. The complication: a text is a linguistic piramide with recursion

So far the analysis, even when becoming increasingly more complicated, could be executed as a linear process, climbing, so to speak, the piramide, ascending from smaller elements (morphemes and lexemes) up to the complex structures

[14] E. Talstra - C.H.J. van der Merwe, 'Analysis, retrieval and the demand for more Data. Integrating the results of a formal textlinguistic and cognitive based pragmatic approach to the analysis of Deut 4:1-40', in: J. Cook (ed.), *Proceedings of the 5th AIBI conference on Bible and Computing in Stellenbosch, July 2000*, Leiden: Brill, 2000

(clauses and sentences. However, only only has to continue the reading of Genesis 22 to find a completely different kind of complexity. The example is in verse 2.[15] Proceding as if we encounter also in this verse a regular piramide of linguistic units, the result of the analysis is very similar to the analysis of verse 1.

Genesis 22:2

```
Levels:      קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ אֲשֶׁר אָהַבְתָּ אֶת יִצְחָק . . .

lexeme       |==| || |==| |=| |===| || |=| || || ||

phrase       |=====| |==| |=| |(=app=)======| || ||

clause       |=====| |======| |===================|

sentence     |======.========.====================|
```

Genesis 22:2

```
                                  וַיֹּאמֶר   Way0  a
              קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ   Impv  b
                             אֲשֶׁר אָהַבְתָּ   xQtl  c
                              אֶת יִצְחָק   ....  d
              וְלֶךְ לְךָ אֶל אֶרֶץ הַמֹּרִיָּה   Impv  e
   וְהַעֲלֵהוּ שָׁם לְעֹלָה לאַחַד הֶהָרִים   WQtl  f
                         אֲשֶׁר אֹמַר אֵלֶיךָ׃   xQtl  g
```

The complication is in line d. This line cannot be analysed as a clause, since it is in fact a part of the clause in line b, that is interrupted by an attribute clause in line c.

Additional analysis:

```
|=====|  |======|  |===================|

      ___________________________
     |                           |
     |                ________   |
     |               |        |  |
|(app)|  |======|  |(=app=)===========|

|=====|  |======|  |===================|
```

[15] Eep Talstra - Constantijn Sikkel, 'Genese und Kategorienentwicklung der WIVU-Datenbank, oder: ein Versuch, dem Computer Hebräisch beizubringen', in: Christof Hardmeier, Wolf-Dieter Syring, Jochen D. Range, Eep Talstra (eds.), *Ad Fontes! Quellen erfassen - lesen - deuten. Was ist Computerphilologie?* [Applicatio 15], Amsterdam: VU University Press, 2000, p. 33-68.

Genesis 22:2

וַיֹּאמֶר	Way0	<+1 999>
קַח נָא אֶת בִּנְךָ אֶת יְחִידְךָ	Impv	< -1 999> <+1 12> <+2 223> <+3 201>
אֲשֶׁר אָהַבְתָּ	xQtl	< -1 12>
אֶת יִצְחָק	Impv	< -2 223>
וְלֶךְ לְךָ אֶל אֶרֶץ הַמֹּרִיָּה	Impv	<+1 200> < -3 201>
וְהַעֲלֵהוּ שָׁם לְעֹלָה לאַחַד הֶהָרִים	WQtl	<+1 11> < -1 200>
אֲשֶׁר אֹמַר אֵלֶיךָ׃	xYqt	< -1 11>

The solution adopted is these cases is this: make a distinction between 'distributional data' and 'functional data'.

Distributional data consist of patterns composed from lower level units. They present a piramide of linguistic data and do no allow any gapping or embedding of data.

Functional data consist of patterns recombined from lower level data. They represent functional units, such as 'clause constituents', 'clauses'.

Below, the analysed data of Genesis 22,2,3 are presented. The columns to the right present the 'distributional units': lexemes, phrase atoms and phrase atom relations, clause atoms and clause atom relations, sentence atoms. The columns to the left present the 'functional units: sentences (at the actual stage of the research they are treated as identical to sentence atoms), clauses, phrases. One could add here again the lexemes, since that actually function in both areas: distributional and functional.

I assume that the majority of the user of SESB will orient themselves on the (classical) functional units. They may want to search for: clauses using the verb X and an object. As a result they will receive from the text below: 'verse 2, sentence 2, clause 1, phrase 1 + 3 and verse 3, clause 1, phrase 2 + 3.

Syntactic Analysis: *Verse 2 - line a/b/c/d/e*

```
phrase    clause    sentence
                    Sent.At. C.A.rel. Cl.Atom P.A.rel. phr.atom lexeme
+=========+=========+=======+========+=======+-------+---------+------   a
| 1.      | 1.      | 1.    |        | 1.    |       | 1.      |    וַ
| Conj.   |         |       | 200 ↑2 |       |       | CjP     | 1.
+---------+ VC      |       |        |       | - - - +---------+------
| 2.      |         |       |200 ↓7  |       |   ↑   | 2.      | יֹּאמֶר
| Predic. | WayQ    |       | 999 ↓1 |       | Way-Y | VP      | 2.
+=========+=========+=======+========+=======+-------+---------+------   b
| 1.      | 1.      | 2.    |        |  2.   |       | 1.      |    קַח
| Predic. |         |       | 999 ↑1 |       |       | VP      | 3.
+---------+ VC      |       |        |       +-------+---------+------
| 2.      |         |       | 201 ↓3 |       |       | 2.      |    נָא
| Modif.  | ImpC    |       |        |       |       | ModP    | 4.
+---------+         |       | 223 ↓2 |       +-------+---------+------
| 3.      |         |       |        |       |       | 3.      |    אֶת
| Object  |         |       |  12 ↓1 |       |       | PP      | 5.
|         |         |       |        |       |       |         +------
|         |         |       |        |       |       |         |  בִּנְךָ
|         |         |       |        |       |       |         | 6.
|         |         |       |        |       + - - - +---------+------
|         |         |       |        |       | ↑ -1p.| 4.      |    אֶת
|         |         |       |        |       | appos.| PP      | 7.
|         |         |       |        |       |       |         +------
|         |         |       |        |       |       |         | יְחִידְךָ
|         |         |       |        |       |       |         | 8.
|  +======+=======+ |       |========+=======+-------+---------+------   c
|  |1.    |  2.   | |       |  12 ↑1 | 3.    |       | 5.      |   אֲשֶׁר
|  |Relat.|  VC   | |       |        |       |       | RelP    | 9.
|  +------+  xQtl | |       | VC     |       +-------+---------+------
|  |2.    |       | |       |        |       |       | 6.      |  אָהַבְתָּ
|  |Predic|  attr.| |       |        |       |       | VP      | 10.
|  +======+=======+ |       |========+=======+-------+---------+------   d
| 3.      | 1.      |       | 223 ↑2 | 4.    | ↑ -3p.| 7.      |    אֶת
|         |         |       |        |       | appos.| PP      | 11.
|         |         |       |        |       |       |         +------
|         |         |       |        |       |       |         |  יִצְחָק
|         |         |       |        |       |       |         | 12.
+=========+=========+=======+========+=======+-------+---------+------   e
| 1.      | 1.      | 3.    |        | 5.    |       | 1.      |     וְ
| Conj.   |         |       | 201 ↑3 |       |       | CjP     | 13.
+---------+ VC      |       |        |       +-------+---------+------
| 2.      |         |       |        |       |       | 2.      |    לֶךְ
| Predic. | ImpC    |       |        |       |       | VP      | 14.
+---------+         |       |        |       +-------+---------+------
| 3.      |         |       |        |       |       | 3.      |    לְךָ
| sup.Comp|         |       |        |       |       | PP      | 15.
+---------+         |       |        |       +-------+---------+------
| 4.      |         |       |        |       |       | 4.      |    אֶל
| Compl.  |         |       |        |       |       | PP      | 16.
|         |         |       |        |       |       |         |------
|         |         |       |        |       |       |         |   אֶרֶץ
|         |         |       |        |       |       |         | 17.
|         |         |       |        |       | - - - +---------+------
|         |         |       |        |       | ↑ -1p | 5.      |     הַ
|         |         |       |        |       | specif| NP      | 18.
|         |         |       |        |       |       |         |------
|         |         |       |        |       |       |         |  מֹּרִיָּה
|         |         |       |        |       |       |         | 19.
+=========+=========+=======+========+=======+-------+---------+------
```

Verse 3 - line c

```
phrase    clause    sentence
                    Sent.At. C.A.rel. Cl.Atom   P.A.rel. phr.atom  lexeme

+=========+=========+=======+========+=======+-------+---------+-------
| 1.      | 1.      | 1.    |        | 1.    |       | 1.      |     וַ
| Conj.   |         |       |  200 ↑1|       |       | CjP     |
+---------+ VC      |       |        |       | - - - +---------+-------
| 2.      |         |       |200 ↓1  |       |  ↑    | 2.      | יִּקַּח
| Predic. | WayQ    |       |        |       | Way-Y | VP      |
+---------+         |       |        |       |-------+---------+-------
| 3.      |         |       |        |       |       | 3.      |    אֶת
| Object  |         |       |        |       |       | PP      |
|         |         |       |        |       |       |         |-------
|         |         |       |        |       |       |         |   שְׁנֵי
|         |         |       |        |       |       |         |
|         |         |       |        |       |       |         |-------
|         |         |       |        |       |       |         | נְעָרָיו
|         |         |       |        |       |       |         |
| +-------+         |       |        |       |-------+---------+-------
| |4.     |         |       |        |       |       | 4.      |   אִתּוֹ
| |Compl  |         |       |        |       |       | PP      |
| +-------+         |       |        |       |-------+---------+-------
|         |         |       |        |       |↑ -2p.| 5.      |     וְ
| 3.      |         |       |        |       |Link   | CjP     |
|         |         |       |        |       |-------+---------+-------
|         |         |       |        |       |↑ -3p.| 6.      |    אֵת
|         |         |       |        |       |Paral.| PP      |
|         |         |       |        |       |       |         |-------
|         |         |       |        |       |       |         | יִצְחָק
|         |         |       |        |       |       |         |
|         |         |       |        |       |- - - -+---------+-------
|         |         |       |        |       |↑ -1p.| 7.      |   בְּנוֹ
|         |         |       |        |       |appos.| NP      |
+=========+=========+=======+========+=======+-------+---------+-------
```

2.4. Elements from the method of grammatical parsing

The method used here implies an *iterative or recursive process* of the analysis.

- The computer proposes the construction of linguistic elements on the basis of *patterns of distribution*. It will work from simple constructions to complex constructions.
- Once constructions have been made, their *functions* and *relations* can be calculated.
- If calculations of this type meet problems, a revision of the elementary constructions proposed may be necessary.

This method of computer-assisted grammatical parsing is a compromise between a distributional and a functional approach. In collecting linguistic data one concentrates on phenomena and patterns of the surface text. In organizing and interpreting these data more functional or even traditional linguistic concepts are adopted, such as 'apposition', 'attribute', 'genitive'. These concept are used as labels for patterns of linguistic data established at a lower linguistic level.
One has, in fact, to avoid two extremes. On the one hand the application of complete philological knowledge to all cases in a text by a scholar who afterwards enters this interpretative knowledge into a data base. In such an approach one would miss to much of the formal and distributional information. On the other hand one should not try to define a complete formal grammar of the linguistic material first and expect the machine to perform the parsing of the texts by applying this grammar. This might work with a set of more or less isolated clauses. It will become much more difficult when applied to a real text with a considerable amount of embedded or elliptic clauses. And it certainly will not work with an ancient *text* corpus as the Hebrew Bible where it will be extremely difficult to extract from texts that have been reworked and re-edited by generations the generative power of the language used. It therefore seems to be much more effective to reverse the procedure and try to construct a grammar of Biblical Hebrew at the very end of computer-assisted procedures rather than as their first step.
Therefore, the method of production used has the following features.

2.4.1. *Ascendent procedure*

The computer builds up linguistic information in a cumulative way, working from the smallest grammatical units, (i.e. morphemes) up to the larger ones, (i.e. words, phrases, clauses, sentences). [16]

[16] W. Richter, *Grundlagen einer althebräischen Grammatik. A. Grundfragen einer sprachwissenschaftlichen Grammatik. B. Die Beschreibungsebenen: I. Das Wort (Morphologie),*(ATSAT 8) St Ottilien, 1978

2.4.2. *Pattern matching*

The programs do not propose grammatical parsings by applying analytical rules, but by a process of pattern matching. This means that parsing programs have at their disposal lists of patterns accepted by the users as grammatically correct parsings. These patterns, therefore, have the status of a *preliminary hypothesis* about the analysis of the linguistic material. For instance: once the sequence 'definite article + noun' has been found in the text, it is accepted as a pattern that builds a phrase-atom. From now on it can be used to propose other phrase-atoms in the texts: [ה מלך], [ה ארץ], etc. The sequence 'personal pronoun + proper name', however, will not be accepted as a phrase, which means that with each occurrence in the text the program will propose two phrases, e.g.: [חברון] [היא], [דוד] [אני], etc. The list of accepted patterns is expanded with new patterns in an interactive process of textual analysis. (See below § 2.5 about the principles of the parsing process.)

2.4.3. *Distributional patterns*

Distributional observations are the basis for the parsing process, not functional ones. This does not mean that the parsing process could be performed without a linguistic interpretation of the data. It means that one avoids the definitions of an observed element in terms of a function at a *higher* linguistic level. For instance, phrase definition *rules* of the following type are avoided:

NP	= head + attribute
head	= definite noun
attribute	= adjectival group
defin. noun	= Noun + def. art.
etc.	

Similarly clause definition *rules* are avoided [17], for example:

Clause	= Margin + Verb + Subject + Object + Margin

The parsing process is used to describe linguistic elements in terms of the distribution of *lower* level parts. For example, distributional patterns of NP's used for further pattern matching are for instance:

def.art. - noun - def.art. - adjective	= NPdet
def.art. - noun	= NPdet
def.art. - noun - proper name	= NPdet + NPdet(app)

[17] K.E. Lowery, *Towards a Discourse Grammar of Biblical Hebrew (Ph.D. University of California)*, Los Angeles, 1985, (U.M.I., 1988), p. ...

noun - def.art - noun - def.art. - adjective	= NPdet
noun - proper-noun - def.art. - adjective	= NPdet
noun - proper-noun	= NPdet
noun [18]	= NP

Similar patterns are used for the marking of clause divisions.

2.4.4. *Secondary functional analysis*

Some elements at a lower linguistic level can only get a label indicative of their grammatical function if this can be based on information of a *higher* level.
This regards for example: the *functional or phrase-dependent part of speech* (due to the phrase construction it belongs to a lexeme may change its part of speech, e.g. a verb in infinitive absolute may function as an adverbial phrase), *state* (when this is based on phrase construction rather than on nominal endings or vowel patterns), *phrase type* or *clause type*.

2.5. Elements from the parsing process

As mentioned in the previous paragraph, the process of computer-assisted grammatical analysis proposed here is an iterative one. Sets of tentatively defined phrase-atoms and clause-atoms are used to let the machine propose the segmentation of phrase-atoms and clause-atoms of a new text. Traditional grammatical knowledge is used, though not for the interpretation of individual textual data, but to identify sets of elements (a taxonomy) from which phrases and clauses are built. Functional features or relations between phrases or clauses are not yet identified at this stage. Only the patterns of distribution are involved here. Greater consistency in applying the sets of phrase-atoms and clause-atoms to the texts means better grammatical quality of these sets. In the end, these data sets may be expected to present a correct grammatical description of the textual corpus.
The data sets (lists of patterns) are created as follows:

- The parsing program reads a text, and shows to the user the patterns (phrase atoms or clause atoms) it has been able to recognize in it. Of course, when in the beginning of a project the list of known patterns is empty or still is very short, no patterns at all or only very few patterns will be recognized.
- The user checks whether the recognized patterns actually fit this text. If not, the user can delimit the correct patterns.
- If a correction by the user results in a pattern that was not yet included in the list of known patterns, the list is expanded with this new pattern.

[18] I do not yet enter here the additional morphological or lexical features that may be part of these patterns.

- From then on, the new pattern counts as a 'known pattern' and can be used by the program when parsing new text.

When the entire textual corpus has been parsed in this way, the resulting lists of patterns (either phrase atoms or clause atoms) has the status of a consistent, though preliminary hypothesis about the analysis of the linguistic material in this particular textual corpus. The lists, or data sets will rather represent a first index or existing morphosyntactical patterns than an adequate grammatical system [19]. But they will provide the researcher with the material relevant for the development of a syntactical theory of a more functional or generative kind.
The iterative process of the analysis not only proceeds from more tentatively made decisions to a more consistently organized set of analysis data, it also moves from lower levels of analysis to higher ones.

2.5.1. *At Phrase level*

External division

The production of phrase-atoms is based on word-level information. The program tries to match previously accepted patterns of word connections from a list (Phrase Set) with patterns of words found in the text to be analyzed. In addition to the Phrase Set the program uses a file of Morphological Conditions and a file of Lexical Conditions, where patterns of restrictions on lexeme combinations are stored. By being based on pattern recognition only, the procedure is a paradigmatic and a distributional one: in principle each pattern gets only one parsing. The use of the Phrase Set opens up possibilities for syntagmatic or context sensitive analysis at the previous level (i.e. word level), for example, the definition of the functional, phrase-determined part of speech and the determination of nominal state in cases where vowel patterns of nominal morphemes are ambiguous.

Internal structure

Once the phrase-atoms in a text have been identified as special arrangements of lexemes matching the morphological and lexical features defined by the Phrase set, the internal hierarchical structure of the phrase-atoms can be parsed in a similar procedure. For example, the word combination: 'Noun' → 'adjective', once accepted as a phrase-atom, gets an additional parsing of its internal relation: 'Noun' (+attribute) → adjective (-attribute). The signs '+' and '-' indicate the direction of the relation. The word combination 'Noun' → 'Proper name' gets an additional parsing: 'Noun' (+regens) → 'Noun' (-rectum).

[19] C. Hardmeier, Fs. W. Richter

2.5.2. *At Clause level*

External division

The production of clause atoms is based on phrase-level and word-level information. The program tries to match previously accepted patterns of phrase-atom orders (stored in a list: Clause Set) with patterns of phrase-atoms found in the text to be analyzed.
The program uses a file Morphological Conditions, a file Lexical Conditions where patterns of restrictions on phrase orders are stored.

Like the previous level (phrase segmentation) this procedure is a paradigmatic and distributional one, due to the technique of pattern recognition. But it too opens up possibilities for further syntagmatic analysis at the lower level, e.g. the definition of compound phrases.

Internal structure

Once the clause-atoms have been established the clause-internal relations of clause constituents can be parsed: Predicate, Subject, Object, Complement, etc. A file with Verb Valency Patterns is used to check clause_atoms for special restraints in the combination of verbs and complements or adjuncts. A simliar file with patterns of Verbless Clauses is used for the parsing of Nominal Clauses.

VbCl: [<Co> לְ-פְּנֵי עַם הָ-אָרֶץ׃] [<Su> אַבְרָהָם] [<Pr> יִּשְׁתַּחוּ] [<Cj> וַ-] Gen 23,12

NmCl: [<PC> חֶבְרוֹן] [<Su> הִוא] Gen 23,19

2.5.3. *At Text level*

The production of a clause hierarchy is based on the information of all previous levels, e.g. the order of phrases, word features such as person, number and gender of the verb and of pronominal suffixes. This procedure is only partially a paradigmatic one. This the case, for instance, where predefined patterns of clause connections can be matched with clauses in the text on the basis of certain lexical combinations:

כן ← כאשׁר, or אם ← ה .

For the greater part, the construction of clause hierarchies is a syntagmatic procedure. It is not only the *type* of a clause (e.g. **W** - X - Qatal) that is decisive here. It also has to calculate the linguistic information that creates *relations* between clauses: "consecutio verborum", the presence or absenc of an explicit NP for the subject, pronominal reference, distance between clauses in the text, etc.[20].

This method of parsing by using a recursive and cumulative procedure of grammatical parsing to some extent can be regarded as an imitation of the process of reading ancient texts. Activities that are part of this process are:
- the observation of *surface* text data,
- *preliminary grammatically labeling* based on distributional information,
- *correction* or expansion of the linguistic information by feed back from the direct context or, at a later stage, by back-tracking based on information constituted at a higher linguistic level.
- consistency check, by applying newly accepted linguistic information in the continuation of the parsing process.

[20] Cf E. Talstra, "Text Grammar and Computer. The Balance of Interpretation and Calculation", in: *Actes du Troisième Colloque International Bible et Informatique: "Interprétation, Herméneutique, Expertise", Tübingen 28-31 aout 1991 (Paris / Genève 1992)* p.135-149.

2.6. Details of a text syntactic analysis

Presentation of Genesis 22

Symbols used:

Textual Hierarchy

```
                                    Ln   Ttype Cl type PNG   Txt.ref.

          ---------------]          14    N     WayX   3sgM    22,03
          ------------]  .          15    N     Way0   3sgM    22,03
          ------------]  .          16    N     Way0   3sgM    22,03
```

The categories indicated

Ln Line Number

Ttype Text type
N: narrative text (starting from wayyiqtol)
Q: discursive text (direct speech, starting from 'q')
D: discursive text (starting from yiqtol in narrative text)
ClType Syntactic Clause Label
Some examples:
NmCl: nominal clause, with <PC>
WayX: Wayyiqtol + NP <S>
Way0: Wayyiqtol - NP <S>
WQtl: W-Qatal
WXQt: W-X(NP=subject)-Qatal
WxQt: W-x(not subject)-Qatal
0Qtl: asyndetic Qatal
PNG verbal predicate of the clause: Person, Number, Gender

The main parsing labels used

<Pr>	Predicate	<PC>	predicative Complement (adj., nom., ptc.)
<PO>	Predicate + Object (vb.fin. + sfx.)		
<Su>	SubjectSpecifier	<Ob>	ObjectComplement
<Co>	Complement	<Aj>	Adjunct
<Ti>	TimeReference	<Lo>	LocativeReference

```
                                              Genesis 22      Textual Hierarchy   Ln  Ttype  ClLab Vpng Vs

                                   [<Ti> אחר הדברים האלה] [<Pr> יהי] [<Cj>ו]   1  N      Way0  3sgM 01
                   [<Ob> את אברהם] [<Pr> נסה] [<Su> האלהים] [<Cj>ו]      |     2  N      WXQt  3sgM 01
                                    [<Co> אליו] [<Pr> יאמר] [<Cj>ו]   |      |     3  N      Way0  3sgM 01
==============================================================================+|   |      |        =======
                                                      [<Vo> אברהם] ||   |      |     4  NQ     Voct  ---- 01
==============================================================================+|   |      |        =======
                                                [<Pr> יאמר] [<Cj>ו]   |      |     5  N      Way0  3sgM 01
==============================================================================+|   |      |        =======
                                                        [<Is> הנני] ||   |      |     6  NQ     NmCl  ---- 01
==============================================================================+|   |      |        =======
                                                [<Pr> יאמר] [<Cj>ו]   |      |     7  N      Way0  3sgM 02
==============================================================================+    |      |        =======
                   [<Ob><ap> את בנך / את יחידך] [<Ij> נא] [<Pr> קח] |     |      |     8  NQ     imp.  2sgM 02
                                   [<Pr> אהבת] [<Re> אשר]   |   |   |     |      |     9  NQ     XQtl  2sgM 02
                                            [<Qb><ap> את יצחק]   |   |     |      |    10  NQ     Defc  ---- 02
                  [<Co> אל ארץ המריה] [<sc> לך] [<Pr> לך] [<Cj>ו]      |     |      |    11  NQ     imp.  2sgM 02
     [<Lo>על אחד ההרים] [<Co> לעלה] [<Lo> שם] [<PO> העלהו] [<Cj>ו]      |     |      |    12  NQ     imp.  2sgM 02
                           [<Co> אליך] [<Pr> אמר] [<Re> אשר]         |     |      |    13  NQ     Xyqt  1sg- 02
==============================================================================+    |      |        =======
                       [<Ti> בבקר] [<Su> אברהם] [<Pr> ישכם] [<Cj>ו]          |    14  N      WayX  3sgM 03
                              [<Ob> את חמרו] [<Pr> יחבש] [<Cj>ו]   |      |    15  N      Way0  3sgM 03
 [<PA><ap> בנו / את יצחק /ו] [<Co> אתו] [<Ob> את שני נעריו] [<Pr> יקח] [<Cj>ו]   |      |    16  N      Way0  3sgM 03
                              [<Ob> עצי עלה] [<Pr> יבקע] [<Cj>ו]   |      |    17  N      Way0  3sgM 03
                                                [<Pr> יקם] [<Cj>ו]   |      |    18  N      Way0  3sgM 03
                              [<Co> אל המקום] [<Pr> ילך] [<Cj>ו]   |      |    19  N      Way0  3sgM 03
                    [<Su> האלהים] [<Co> לו] [<Pr> אמר] [<Re> אשר]      |      |    20  N      XQtl  3sgM 03
                                                   [<Ti> ביום השלישי]      |    21  N      CPen  ---- 04
                     [<Ob> את עיניו] [<Su> אברהם] [<Pr> ישא] [<Cj>ו]          |    22  N      WayX  3sgM 04
                   [<Aj> מרחק] [<Ob> את המקום] [<Pr> ירא] [<Cj>ו]   |         |    23  N      Way0  3sgM 04
                     [<Co> אל נעריו] [<Su> אברהם] [<Pr> יאמר] [<Cj>ו]          |    24  N      WayX  3sgM 05
==============================================================================+|         |        =======
                   [<Aj> עם החמור] [<Co> פה] [<sc> לכם] [<Pr> שבו] ||         |    25  NQ     imp.  2plM 05
                [<Co> עד כה] [<Pr> נלכה] [<Su> אני והנער] [<Cj>ו]     ||         |    26  NQ     WPyq  1pl- 05
                                        [<Pr> נשתחוה] [<Cj>ו]         ||         |    27  NQ     Wey0  1pl- 05
                          [<Co> אליכם] [<Pr> נשובה] [<Cj>ו]         ||         |    28  NQ     Wey0  1pl- 05
==============================================================================+|         |        =======
                    [<Ob> את עצי העלה] [<Su> אברהם] [<Pr> יקח] [<Cj>ו]          |    29  N      WayX  3sgM 06
                       [<Co><ap> על יצחק / בנו] [<Pr> ישם] [<Cj>ו]   |   |      |    30  N      Way0  3sgM 06
         [<Ob> את האש ואת המאכלת] [<Co> בידו] [<Pr> יקח] [<Cj>ו]   |   |      |    31  N      Way0  3sgM 06
```

```
                                                                                               Ln   Ttype    ClLab Vpng Vs

                                [<Mo> יחדו] [<Su> שניהם] [<Pr> ילכו] [<Cj>ו]   |        |  32  N        WayX  3plM 06
                  [<Co><ap> אביו / אל אברהם] [<Su> יצחק] [<Pr> יאמר] [<Cj>ו]   |   |    |  33  N        WayX  3sgM 07
                                                   [<Pr> יאמר] [<Cj>ו]   |   |   |    |  34  N        Way0  3sgM 07
==================================================================+|   |   |   |    |      =======
                                                       [<Vo> אבי] ||   |   |   |    |  35  NQ       Voct  ---- 07
==================================================================+|   |   |   |    |      =======
                                                   [<Pr> יאמר] [<Cj>ו]   |   |   |    |  36  N        Way0  3sgM 07
==================================================================+|   |   |   |    |      =======
                                                      [<Is> הנני] ||   |   |   |    |  37  NQ       NmCl  ---- 07
                                                    [<Vo> בני]    ||   |   |   |    |  38  NQ       Voct  ---- 07
==================================================================+|   |   |   |    |      =======
                                                   [<Pr> יאמר] [<Cj>ו]   |   |   |    |  39  N        Way0  3sgM 07
==================================================================+|   |   |   |    |      =======
                                 [<Su> האש והעצים] [<Ij> הנה] ||   |   |   |    |  40  NQ       NmCl  ---- 07
                      [<Su><sp> לעלה / השה] [<Qp> איה] [<Cj>ו]    ||   |   |   |    |  41  NQ       NmCl  ---- 07
==================================================================+|   |   |   |    |      =======
                                     [<Su> אברהם] [<Pr> יאמר] [<Cj>ו]   |   |    |  42  N        WayX  3sgM 08
======================================================================+    |   |    |      =======
             [<Ob><sp> לעלה / השה] [<Co> לו] [<Pr> יראה] [<Su> אלהים] |    |   |    |  43  NQ       Xyqt  3sgM 08
                                                       [<Vo> בני]     |    |   |    |  44  NQ       Voct  ---- 08
======================================================================+    |   |    |      =======
                                [<Mo> יחדו] [<Su> שניהם] [<Pr> ילכו] [<Cj>ו]   |    |  45  N        WayX  3plM 08
                                    [<Co> אל המקום] [<Pr> יבאו] [<Cj>ו]   |    |  46  N        Way0  3plM 09
                      [<Su> האלהים] [<Co> לו] [<Pr> אמר] [<Re> אשר]   |   |    |  47  N        XQtl  3sgM 09
              [<Ob> את המזבח] [<Su> אברהם] [<Lo> שם] [<Pr> יבן] [<Cj>ו]   |    |  48  N        WayX  3sgM 09
                                [<Ob> את העצים] [<Pr> יערך] [<Cj>ו]   |   |    |  49  N        Way0  3sgM 09
                        [<Ob><ap> בנו / את יצחק] [<Pr> יעקד] [<Cj>ו]   |   |    |  50  N        Way0  3sgM 09
[<Lo><sp> לעצים/ ממעל] [<Co> על המזבח] [<Ob> אתו] [<Pr> ישם] [<Cj>ו]   |   |    |  51  N        Way0  3sgM 09
                           [<Ob> את ידו] [<Su> אברהם] [<Pr> ישלח] [<Cj>ו]   |    |  52  N        WayX  3sgM 10
                                [<Ob> את המאכלת] [<Pr> יקח] [<Cj>ו]   |   |    |  53  N        Way0  3sgM 10
                                      [<Ob> את בנו] [<Pr> לשחט]       |   |    |  54  N        infc. ---- 10
```

```
                                                                                                    Ln   Ttype     ClLab Vpng Vs

                  [<Lo> מן השמים] [<Su> מלאך יהוה] [<Co> אליו] [<Pr> יקרא] [<Cj>ו]      |        |  55   N         WayX  3sgM 11
                                                       [<Pr> יאמר] [<Cj>ו]   |   |       |        |  56   N         Way0  3sgM 11
        ==============================================================+|   |   |       |        |       =======
                                                 [<Vo> אברהם אברהם] ||   |   |       |        |  57   NQ        Voct  ---- 11
        ==============================================================+|   |   |       |        |       =======
                                                       [<Pr> יאמר] [<Cj>ו]   |   |       |        |  58   N         Way0  3sgM 11
        ==============================================================+|   |   |       |        |       =======
                                                              [<Is> הנני] ||   |   |       |        |  59   NQ        NmCl  ---- 11
        ==============================================================+|   |   |       |        |       =======
                                                       [<Pr> יאמר] [<Cj>ו]   |   |       |        |  60   N         Way0  3sgM 12
        ==============================================================+    |   |       |        |       =======
                       [<Co> אל הנער] [<Ob> ידך] [<Pr> תשלח] [<Ng> אל] |    |   |       |        |  61   NQ        Xyqt  2sgM 12
           [<Ob> מאומה] [<Co> לו] [<Pr> תעשׂ] [<Ng> אל] [<Cj>ו]   |      |    |   |       |        |  62   NQ        WLyq  2sgM 12
                                   [<Pr> ידעתי] [<Ti> עתה] [<Cj> כי]      |    |   |       |        |  63   NQ        XQtl  1sg- 12
                      [<Su> אתה] [<PC> ירא אלהים] [<Cj> כי]   |          |    |   |       |        |  64   NQ        NmCl  ---- 12
    [<Ob><ap> את בנך / את יחידך] [<Pr> חשׂכת] [<Ng> לא] [<Cj>ו]           |    |   |       |        |  65   NQ        WLQt  2sgM 12
[<Co> ממני]   =======================================================+    |   |       |        |       =======
                                [<Ob> את עיניו] [<Su> אברהם] [<Pr> ישׂא] [<Cj>ו]   |       |        |  66   N         WayX  3sgM 13
                                                        [<Pr> ירא] [<Cj>ו]   |   |       |        |  67   N         Way0  3sgM 13
                              [<PC> אחר] [<Su> איל] [<Ij> הנה] [<Cj>ו]       |   |       |        |  68   N         NmCl  ---- 13
                             [<Aj> בבקרניו] [<Co> בסבך] [<Pr> נאחז]           |   |       |        |  69   N         0Qtl  3sgM 13
                                              [<Su> אברהם] [<Pr> ילך] [<Cj>ו]   |       |        |  70   N         WayX  3sgM 13
                                        [<Ob> את האיל] [<Pr> יקח] [<Cj>ו]   |   |       |        |  71   N         Way0  3sgM 13
                        [<Aj> תחת בנו] [<Co> לעלה] [<PO> יעלהו] [<Cj>ו]       |   |       |        |  72   N         Way0  3sgM 13
                         [<Ob> שׁם המקום ההוא] [<Su> אברהם] [<Pr> יקרא] [<Cj>ו]   |       |        |  73   N         Way0  3sgM 14
        ==============================================================+|       |       |        |       =======
                                              [<Pr> יראה] [<Su> יהוה] ||       |       |        |  74   NQ        Xyqt  3sgM 14
        ==============================================================+|       |       |        |       =======
                                      [<Ti> היום] [<Pr> יאמר] [<Cj> אשׁר]       |       |        |  75   ND        Xyqt  3sgM 14
        ==========================================================+   |       |       |        |       =======
                                     [<Pr> יראה] [<Lo> בהר יהוה] |   |       |       |        |  76   NDQ       Xyqt  3sgM 14
          ------------------------------------------------------------+       |       |        |       -------
```

```
                                                                                                 Ln   Ttype      ClLab Vpng Vs

 [<Lo> מן השמים] [<Mo> שנית] [<Co> אל אברהם] [<Su> מלאך יהוה] [<Pr> יקרא] [<Cj>ו]      |         |  77  N          WayX  3sgM 15
                                                         [<Pr> יאמר] [<Cj>ו]      |         |  78  N          Way0  3sgM 16
  ==========================================================================+      |         |      =======
                                       [<Pr> נשבעתי] [<Co> בי] |      |         |  79  NQ         XQtl  1sg- 16
                                     [<PC> נאם יהוה]   |   |   |      |         |  80  NQ         NmCl  ---- 16
                                              [<Cj> כי]   |   |      |         |  81  NQ         Defc  ---- 16
                  [<Ob> את הדבר הזה] [<Pr> עשית] [<Cj> יען אשר]   |   |   |      |         |  82  NQ         XQtl  2sgM 16
      [<Ob><ap> את בנך / את יחידך] [<Pr> חשכת] [<Ng> לא] [<Cj>ו]   |   |   |      |         |  83  NQ         WLQt  2sgM 16
                             [<PO> אברכך] [<Mo> ברך] [<Cj> כי]   |   |      |         |  84  NQ         Xyqt  1sg- 17
        ככוכבי השמים] [<Ob> את זרעך] [<Pr> ארבה] [<Mo> הרבה] [<Cj>ו]   |   |      |         |  85  NQ         Xyqt  1sg- 17
 [<Aj> וכחול            [<PC> על שפת הים] [<Re> אשר]  |   |   |      |         |  86  NQ         NmCl  ---- 17
              [<Ob> את שער איביו] [<Su> זרעך] [<PC> ירש] [<Cj>ו]   |   |      |         |  87  NQ         WeyX  3sgM 17
         [<Su> כל גויי הארץ] [<Co> בזרעך] [<Pr> התברכו] [<Cj>ו]   |   |      |         |  88  NQ         WQtl  3pl- 18
                         [<Co> בקלי] [<Pr> שמעת] [<Cj> עקב אשר]   |      |         |  89  NQ         XQtl  2sgM 18
  ==========================================================================+      |         |      =======
                              [<Co> אל נעריו] [<Su> אברהם] [<Pr> ישב] [<Cj>ו]         |  90  N          WayX  3sgM 19
                                                    [<Pr> יקמו] [<Cj>ו]   |         |  91  N          Way0  3plM 19
                         [<Co> אל באר_שבע] [<Mo> יחדו] [<Pr> ילכו] [<Cj>ו]   |         |  92  N          Way0  3plM 19
                            [<Co> בבאר_שבע] [<Su> אברהם] [<Pr> ישב] [<Cj>ו]         |  93  N          WayX  3sgM 19
                                    [<Ti> אחרי הדברים האלה] [<Pr> יהי] [<Cj>ו]  94  N          Way0  3sgM 20
                                          [<Co> לאברהם] [<Pr> יגד] [<Cj>ו]     95  N          Way0  3sgM 20
                                                       [<Pr> לאמר]   |   |    96  N          infc. ---- 20
  ==============================================================================+   |   |        =======
 [<Co><ap> לנחור /אחיך] [<Ob> בנים] [<Su><sp> מלכה / גם הוא] [<Pr> ילדה] [<Ij> הנה] |   |   |    97  NQ         0Qtl  3sgF 20
       [<Ob><ap> את עוץ / בכרו / ו/ את בוז / אחיו / ו/ את קמואל / אבי ארם]        |   |   |    98  NQ         Ellp  ---- 21
          [<Ob> את כשד ואת חזו ואת פלדש ואת ידלף ואת בתואל] [<Cj>ו]        |   |   |    99  NQ         Ellp  ---- 22
  ==============================================================================+   |   |        =======
                              [<Ob> את רבקה] [<Pr> ילד] [<Su> בתואל] [<Cj>ו]   |   100  N          WXQt  3sgM 23
                [<Co><ap> לנחור / אחי אברהם] [<Su> מלכה] [<Pr> ילדה] [<Ob> שמנה אלה]   101  N          XQtl  3sgF 23
                                                  [<Fr> פילגשו] [<Cj>ו]   102  N          CPen  ---- 24
                                      [<Su> ראומה] [<PC> שמה] [<Cj>ו]   |   103  N          NmCl  ---- 24
             [<Ob> את טבח ואת גחם ואת תחש ואת מעכה] [<Su> גם הוא] [<Pr> תלד] [<Cj>ו]   104  N          Way0  3sgF 24
```

3. Searching and Presenting of linguistic data. Some examples.

By way of illustrating the variety of options the Hebrew textual database offers for presenting, searching and collecting linguistic data, I present here a number of examples. Of course, access to the options available in the data can only be realised by a search engine that is able to exploit in an effective way all the textual features present. Creating a database is one thing, using it is something else. With the queries below I will indicate where the options of the Amsterdam Hebrew database are not yet fully addressable by the search engine developed so far. I am pleased by the fact that the Amsterdam Hebrew data base can be used now in the context of the SESB. It is also clear, however, that the first version of the user interface presented in the SESB package for access to the Amsterdam Hebrew data base, needs further development. The queries presented below are ordered from word level up to text level.

3.1. Words level data: lexical and grammatical searching

1. *Task:*
   The names Manasse and Efraim, in any order, in a clause, or in a verse.
   Query:

   ```
   manasse:
   Clause
        Word: X 1 מנשה
        Word: between 1 and 5
        Word: X 1 אפרים
   ```

   Result: Gen 46:20 48:1 48:5 48:17 ...

2. *Task:*
   imperative - Weqatal (identical person, number); function= request?
   Query:

   ```
   imp + weqatal:
   Clause: Imp
        Word: Verb 2 M
   Clause: W.Qat
        Word: Verb 2 M
   ```

   Result: Gen. 19:2 27:43-44 44:4 45:9 ...

3. *Task:*
   Weqatal - Weqatal (person 2 > person 3); function= consecutive?
   Query:

   ```
   weqatal2 + weqatal3:
   Clause: W.Qat
        Word: Verb wePf 2
   Clause: W.Qat
        Word: Verb wePf 3
   ```

   Result: Gen 8:17 29:27 Ex 7:19 ...

3.2. Phrase level data:

1. *Task:*
Clause, with phrase = Subject, with lexeme = David.

Query:

```
David subject:
Clause
     Phrase: Sub
          Word: {1} דוד
```

Result: 1 Sam 16:21, 22, 23; 1 Sam 17:12, 15, 20, 22, 23, 26, 29, 31, 32, 34, 37, 39, etc.

2. *Task:*
Nominal phrase with an internal genitive relation;

Query:

```
noun cstr + God:
Phrase
     Word: Noun Adj Cons
     Word: {3} יהוה יה אלהים Name Noun
```

Result: Genesis 1:2, 27 3:8 4:16, 26 5:1 6:8 9:6 etc.

3. *Task:*
Nominal phrase containing an adjective in attributive relation

Query:

```
attrib adj:
Phrase:
     Word: Noun Abs
     Word
     Word: Adj Abs
```

Result: Genesis 1:16 (3x), 21; 2:13, 14 (2x); 6:9; 7:2, 8, 19; 8:5, 20; 9:10 etc.

4. *Task:*
Nominal phrase containing a participle in attributive relation

Query:

```
attrib ptc:
Phrase:
     Word: Noun Abs
     Word
     Word: Ptc(a) Abs
```

Result: Gen 32:16 41:33 41:35 ...

3.3. Sentence and clause level data:

1. *Task:*
Search the Tenach for clauses with constituents of the type: Subject, Object, Predication, in this order.

Query:

```
SOV-clause:
Clause
     Phrase: Sub Phrase: Obj Phrase: V.Pred
```

Result: *Exodus 19:17, *35:21; Leviticus 7:18, 20:11, 21:13, 26:8; *Deuteronomy 5:33, etc.

Clearly the present search engine does not skip embedded clauses. The effect is that it also accepts cases where one of the constituents requested actually is part of a different, embedded clause. These cases are indicated here by '*'.

2. *Task:*
Search the Tenach for clauses with a particular verb and its satelites, e.g. all cases of הלך and its Complements with preposition אל.

Query:

```
HLK >L:
Clause
     Word: {1} הלך
     Phrase: Compl
          Word: {1} אל
```

Result: Gen 12:1 13:3 22:2 22:3 22:19 ...

3. *Task:*
Search the Tenach for clauses with a form of היה as its verbal predication combined with a participle marked as <PC>.

Query:

```
Clause
     Phrase: V.Pred
          Word: {1} היה
     Phrase: N.Pred
          Word: Ptc(a) Ptc(p)
```

[NB in the user interface the label <PC> (= Predicate Complement) has been altered into <N.Pred>

Result: Genesis 1:6, 4:2, 14, 20, 21; 19:14; 21:20; 34:25; 39:22; 42:31; etc.

Unfortunately the search engine does not yet allow for free order of the units requested. It means that one has to built a separate query for the reversed order 'ptc' + היה. [Deuteronomy 9:7,22,24]

4. *Task:*
Search the Tenach for clauses with a form of שׁמע as its verbal predication, rootformation Hif'il, and with complements of the type את or ל or ב
Query:

```
Query 4:  Clause
                Phrase: V.Pred
                        Word: {1} שׁמע Verb Hifil
                Phrase: Compl Supp
                        Word: {3} ב ל את
```

`Result:` `None.` It turns out that the lexeme שׁמע can not be retrieved. Does this have to do with the fact that Hebrew Shin and Sin are composite characters in Unicode?

5. *Task:* [translation problem of Psalm 67:7f.]
Search the Tenach for a combination of two clauses, *i.e.*:
clause 1: non-determinated noun followed by Qatal 3 p.sing;
clause 2: starts with a yiqtol.
Query:

```
Qatal-Yiqtol order
Clause: S.Qat X.Qat
        Word: Noun Word: Verb Pf 3 S
Clause:  0.Yiq  S.Yiq  X.Yiq  SX.Yiq  W.Yiq.0  W.Yiq.S
                WS.Yiq WX.Yiq
```

Result: (some) Genesis 21:6 Isaiah 9:9 Amos 3:8 Psalm 67:7.
The problem is that many other texts listed do match with the query in terms of the two clauses required, but nevertheless fail grammatically since they have no syntactical relation. The actual search engine does not allow for asking syntactical relations encoded in the data.

3.4 Text level data

1. *Task:*
Present the text of Exodus 33:7-14 divided according to its clauses and its clause types. Mark the verbs and mark the constituents with the function <Subject>
Query: Not yet possible in the current version of the search engine
Result:

2.a. *Task:*
Search the Tenach for clauses of the type:
W-X-Yiqtol with Text Type: ND
This means: clauses with the order Subject - imperfect tense in a segment of narrative text (not in a direct speech section). The goal is to search for grammatical exceptions.
Query: Not yet possible in the current version of the search engine
Result:

2.b. *Task:*

Search the Tenach for clauses of the type:
W-Qatal with Text Type: ND
This means: cases similar to 2.a., now with WeQatal (consecutive perfect)

Query: Not yet possible in the current version of the search engine

Result:

3. *Task:*

Search the Tenach for clauses of the type:
* W-Qatal (verbal lexeme = היה; person = 3; number = singular), connected to clauses of the type:
* Infinitive construct in VP (verbal phrase) with preposition כ)
connected to clauses of the type:
* 0-Yiqtol

This means, searching for particular cases of starting a paragraph of text, i.e. והיה + כ+inf.cstr. + Yiqtol

Query:

```
wehayah + k-inf + 0yiq:
Clause: W.Qat
     Word: 1 היה Verb 3 M S
Clause: Inf.cs
     Phrase: Verb
          Word: 1 כ Prep
Clause: 0.Yiq
```

Result: Exodus 33:8,9 Joshua 8:8

Again one wants the option to ask for syntactical clause relations. The actual query demands for three clauses in the order listed, allowing no elements in between. One would however prefer to ask for a syntacical relation between clause 1 and 3. Then the distance between them could be free.

Literature

F.I. Andersen,
The Sentence in Biblical Hebrew, (Janua Linguarum, Series Practica, vol. 231), The Hague, 1974.

W.R. Bodine,
'How Linguists Study Syntax', in: W.R. Bodine (ed), *Linguistics and Biblical Hebrew*, Eisenbrauns: Winona Lake, Indiana, 1992, p. 89 - 107

J.W. Dyk and Eep Talstra,
'Paradigmatic and Syntagmatic Features in Identifying Subject and Predicate in Nominal Clauses', in: Cynthia L. Miller (ed.), *The verbless Clause in Biblical Hebrew. Linguistic Approaches. [LSAWS: Linguistic Studies in Ancient West Semitic, Volume 1]*, Winona Lake: Eisenbrauns, 1999, p. 133- 185

C.J. Doedens,
Text Databases. One Database Model and Several Retrieval Languages (Language and Computers, Studies in Practical Linguistics, 14) (Doctoral dissertation University of Utrecht), Amsterdam: Rodopi, 1994

W. Eckardt,
Computer Gestützte Analyse althebräischer Texte. Algorithmische Erkennung der Morphologie, (ATSAT 29) St. Ottilien, 1987.

E. Gülich and W. Raible,
'Überlegungen zu einer makrostrukturellen Textanalyse': J. Thurber, The Lover and His Lass', in: *Untersuchungen in TextTheorie*, Göttingen, 1977, p 132 - 175.

C. Hardmeier, E. Talstra,
'Sprachgestalt und Sinngehalt. Wege zu neuen Instrumenten der computergestützten Textwahrnehmung', *ZAW* 101 (1989) 408 - 428;

C.H.J. van der Merwe, J.A. Naudé, J.H. Kroeze,
A Biblical Hebrew Reference Grammar, (Biblical Languages: Hebrew 3), Sheffield: Sheffield Academic Press, 1999

C.H.J. van der Merwe,
'Discourse Linguistics and Biblical Hebrew Grammar', in: Bergen, Robert D.(ed.), *Biblical Hebrew and Discourse Linguistics*, Winona Lake: Eisenbrauns, 1994, p. 13 - 49

W. Richter,
Grundlagen einer althebräischen Grammatik. B. Die Beschreibungsebenen. III. Der Satz (Satztheorie), (ATSAT 13) St.Ottilien, 1980;

W. Richter,
Biblia Hebraica transcripta BH^{t} 3. Numeri, Deuteronomium (Arbeiten zu Text und Sprache im Alten Testament 33.3), Münchener Universitäts-Schriften, EOS Verlag: St. Ottilien, 1991. Review by E. Talstra in: *JSS* 39 (1994) 290 - 299.

J.H. Sailhamer,
'A Database approach to the analysis of Hebrew Narrative, *Maarav 5-6* (1990) 319-335;

J.H. Sailhamer,
'2 Samuel 13: 1-4 and a Data Base Aproach to the Analysis of Hebrew narrative', in: *Actes du troisième Colloque International Bible et Informatique: Interprétation, Herméneutique, Expertise*, Tübingen 28-31 aout 1991, CIB Maredsous (ed.), Paris/Genève, 1992, p. 99 - 122;

W.-D. Syring,
'QUEST 2 - Computergestützte Philologie und Exegese', *Zeitschrift für Althebraistik* 11 (1998) 85-89

E. Talstra,
'Hebrew Syntax: Clause Types and Clause Hierarchy', in: K. Jongeling, H.L. Murre-van

den Berg, L. van Rompay (ed.), *Studies in Hebrew and Aramaic Syntax presented to Professor J. Hoftijzer*, Leiden, 1991, 180-193;

E. Talstra,
'Text Grammar and Biblical Hebrew: The Viewpoint of Wolfgang Schneider', *Journal of Translation and Textlinguistics* 5 (Dallas, USA) (1992) 269-297.

E. Talstra, C. Hardmeier, J.A. Groves (ed.),
Quest. Electronic Concordance Applications for the Hebrew Bible (data base and retrieval software), Haarlem: NBG, 1992 [Manual: J.A. Groves, H.J. Bosman, J.H. Harmsen, E. Talstra, *User Manual Quest. Electronic Concordance Application for the Hebrew Bible, Haarlem, 1992*

E. Talstra, 'Demonstration ECA database and retrieval software. A preliminary Report',
in: *Actes du Troisième Colloque International Bible et Informatique: "Interprétation, Herméneutique, Expertise", Tübingen 28-31 aout 1991* (Paris / Genève 1992) p.605-611.

E. Talstra,
'Desk and Discipline. The Impact of Computers on the Study of the Bible'. Opening Adress of the 4th AIBI Conference, in: *Proceedings of the Fourth International Colloquium Bible and Computer: Desk and Discipline, in Amsterdam August 15-18 1994*, Paris/Geneva, 1995, p 25 - 43

E. Talstra,
'Clause Types and Textual Structure. An experiment in narrative syntax', in: *Narrative and Comment. Contributions to Discourse Grammar of Biblical Hebrew, presented to Wolfgang Schneider on the occasion of his retirement as a lecturer of Biblical Hebrew at the "Kirchliche Hochschule" in Wuppertal*, Amsterdam/Kampen, 1995, p. 166 - 180

E. Talstra,
'A Hierarchy of Clauses in Biblical Hebrew Narrative', in: E.J. Van Wolde (ed.), *Narrative Syntax and the Hebrew Bible. Papers of the Tilburg Conference 1996 (Biblical Interpretation Series 29)*, Leiden: Brill, 1997, p. 85-118

E. Talstra
'Workshop: Clause Types, Textual Hierarchy, Translation in Exodus 19, 20 and 24', in: E.J. Van Wolde (ed.), *Narrative Syntax and the Hebrew Bible. Papers of the Tilburg Conference 1996 (Biblical Interpretation Series 29)*, Leiden: Brill, 1997, p. 119-132

E. Talstra,
'Reconstructing the Menorah on Disk. Some Syntactic Remarks', in: Marc Vervenne, ed., *Studies in the Book of Exodus* (BETL) 126, 1996, 523 - 533;

E. Talstra
'Deuteronomy 31: Confusion or Conclusion? The story of Moses' threefold succession', in: M Vervenne, J. Lust (eds), *Deuteronomy and Deuteronomic Literature. Festschrift C.H.W. Brekelmans*, (Bibliotheca Ephemeridum Theologicarum Lovaniensium [BETL] 133), 1997, 87-110

E. Talstra
'Tense, Mood, Aspect and Clause Connections. A Textual Approach', *JNSL [Journal of NorthWest Semitic Languages]*, 23 (1997) 81-103

E. Talstra
'From the Eclipse to the Art of Biblical Narrative. Reflections on Methods of Biblical Exegesis' in E. Noort (ed.), *Perspectives on the Study of the Old Testament and Early Judaism. A symposium in Honour of Adam S. van der Woude on the Occasion of His 70th Birthday, Groningen 1997*, Leiden: Brill, 1998, 1-41

E. Talstra - C. Sikkel,
'Genese und Kategorienentwicklung der WIVU-Datenbank, oder: ein Versuch, dem Computer Hebräisch beizubringen', in: Christof Hardmeier, Wolf-Dieter Syring, Jochen D. Range, Eep Talstra (eds.), *Ad Fontes! Quellen erfassen - lesen - deuten. Was ist*

Computerphilologie? [Applicatio 15], Amsterdam: VU University Press, 2000, p. 33-68

E. Talstra,
Oude en Nieuwe lezers. Een inleiding in de Methoden van Uitleg van het Oude Testament (Ontwerpen 2), Kampen: Kok, 2002

A.J.C. Verheij,
Grammatica Digitalis I. The Morphological Code in the "Werkgroep Informatica" Computer Text of the Hebrew Bible. Applicatio 11, VU Uitgeverij, Amsterdam, 1994.

B. Waltke - M. O'Connor,
An Introduction to Biblical Hebrew Syntax, Winona Lake, 1991, p. 53-55.