Computer-assisted linguistic analysis.
The Hebrew Database used in Quest.2

Eep Talstra
Vrije Universiteit Amsterdam

0. The project Quest.2

This workshop concentrates on practical information about the goal and the features of Quest.2, a cooperation project of the *'Vrije Universiteit'* (Amsterdam), the *'Universität Greifswald'*, *'Westminster Theological Seminary'* (Philadelphia), the *'Deutsche Bibelgeselschaft'* (Stuttgart) and the *'Nederlands Bijbelgenootschap'* (Haarlem). My goal is to explain the data types available in the Hebrew syntactic data base that is being produced at the *Vrije Universiteit* and that will be used by the retrieval programme 'Quest.2', that is developed at the University of Greifswald. The complete package will be distributed by the German Bible Society.[1]
This paper will also present a number of the options to choose from and the difficulties to solve when one is engaged in the production and the retrieval of lexical, morphological and syntactic data.

The emphasis in my contribution being on data types and data structures, the main focus will be on the research needed to produce linguistically analyzed Hebrew text data and the impact this research has on grammatical and exegetical research in general. As a tool for retrieval, Quest2 will allow its user to retrieve from the database of Old Testament text a variety of grammatical and lexical data. So the challenge for the data producer is: how does one define and store the following data?

- words [lexemes, word forms, combination of words, word order, part of speech, comparison with both ancient and modern translations]
- phrases [phrase types, complex phrases and their internal relations]
- clauses [clause types, parsing of constituents, patterns of verbal valency]
- clause connections [sequences of clauses, shift of text types (narrative, direct speech), segmentation of a text into paragraphs].[2]

---

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

2 See: 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.

1. The data types used by Quest.2

First, a user of a Bible text database, being a reader of an ancient text, needs to have access to all kinds of linguistic layers of the text, not only to word-level elements of the language. To clarify what that requirement implies for the definition of the data types, the presentation given here starts with an example of data analysis at text level, even when in terms of actual data production the text level, of course, is the final one and not the first one. In a way, one may even claim that a text-level database never may be really completed. For, will we ever be able to understand and analyze fully the precise balance of syntactic, lexical and rhetorical features in a text? However, for an exegete the text level is the most important one, since here it is that language as a system and a literary text as an unique composition interact.

Second, starting from the text level, this paper will inform about the various linguistic layers identified in the data and the data types identified at each of these levels.

Third, some wishes and possibilities of retrieval, present or yet to be made available in future versions of Quest.2. Clearly we will need more international cooperation in this area, since exegetes, literary analists and Hebraists, all have their particular sets of wishes, which one team never will be able to meet completely.

1.1. Description of part of an analysed text

The data base of Hebrew text used by Quest.2 is organized by two, overlapping, ways of structuring the data.[3] First, it is crucial to preserve the document structure (books, chapters, verses, half-verses, words) of the data, to allow access in terms of traditional categories. Second, the data have to represent linguistic information from morpheme-level up to sentence and paragraph-level. This is mainly the information to be presented below.
The actual process of producing syntactic information has been introduced in other publications[4]. Once produced, the actual linguistic data allow for a number of different presentations. By way of introduction I present first a short section a fully analysed text, taken from the text of I Kings 21. Next to that, an overview of the internal structure of the data themselves. These presentations I will use again in § 2.1. to explain a number of the challenges and possibilities of the computer-assisted textual analysis that is applied to

---

[3] See the descriptions of the Quest.1 data by 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.

[4] 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. See also ftn 1.

produce the data to be used by Quest.2.

An example of analyzed text: I Kings 21, 2

The text is presented in lines that equal simple clauses. In cases of syntactic embedding (not in this sample text) a line will represent only part of a clause.

```
                                    [<Co> אל נבות] [<Su> אחאב] [<Pr> ידבר] [<Cj>ו] 4  N   WayX  3sgM 02
                                                                  [<Pr> לאמר]  5  N   InfC. ---- 02
 ==========================================================+      ===
                                [<Ob> את כרמך] [<Co> לי] [<Pr> תנה] |   6  NQ  Imp.  2sgM 02
                  [<Co> לגן ירק] [<PC> לי] [<Pr> יהי] [<Cj>ו]     |   |   |   7  NQ  Wey0  3sgM 02
 [<Lo>אצל ביתי] [<PC>קרוב] [<Su> הוא] [<Cj> כי]  [  |   |   |   8  NQ  AjCl  -sg- 02
[<Ob>כרם] [<Aj> התחתיו] [<Co>לך] [<Pr> אתנה] [<Cj>ו]  |   |   |   9  NQ  Wey0  1sg- 02
                                    [<Co> ממנו] [<PC> טוב]      |   |  10  NQ  AjCl  -sg- 02
                          [<Co> בעיניך] [<PC> טוב] [<Cj> אם]  ↓   |   |  11  NQ  AjCl  -sg- 02
                           [<Ob> כסף מחיר זה] [<Co> לך] [<Pr> אתנה]   |  12  NQ  0Yqt  1sg- 02
 ==========================================================+      ===
```

Data presented:

Grammatical features of the predicate

3sgM	3 singular masculine	2sgM	2 singular masculine
1sg-	1 singular -	-sg-	- singular -

Parsing of Clause Constituents

<Pr>	Verbal predicate	<PC>	Nom. Complement to predication
<Su>	Subject	<Co>	Complement
<Ob>	Object	<Aj>	Adjunct
<Lo>	Locative	<Cj>	Conjunction

Clause type labels

WayX	Wayyiqtol(+<Su>) clause	InfC	Infinitive construct clause
Wey0	Weyiqtol (-<Su>) clause	0Yqt	Yiqtol (on first pos.) clause
AjCl	Adjective <=PC> clause	Imp.	Imperative clause

Text Types

N	Narrative section	Q	Quotation (direct speech) section

Clause Hierarchy

Syntactic relations are visualized by a system of indentation and vertical marks. In the majority of cases the clauses are connected upward, to foregoing clauses ['|']. Sometimes, as in line 11, a downward

connection ['↓'] is used. Internally the data have codes added to each connection of two clauses.

Participant tracking

A further development, I am actually experimenting with, is to have the programme that proposes the clause connections, also make proposals for relating textual participants to each other, e.g. וִיהִי (line 7) refers back to כַּרְמְךָ (line 6), its suffix ךָ referring back to the verb תְּנָה. This imperative in its turn refers back to the complement אֶל נָבוֹת in the previous clause (line 4). This procedure requires calculations based on person-number-gender features, but it also requires calculations to recode, e.g., the subject and the complement of the narrative clause into a first person and a second person in the dependent direct speech section. It is a complicated procedure, which, one may assume, never will run fully automatically. But the experiment as such also helps in understanding more of syntactic processes in textual structures and it contributes to a better understanding of the interacting of syntax and pragmatics. Once the data of relations between textual participants have been produced, they will be very helpful for further discourse analysis. And they may help the exegete to get access to a text in terms of the relations or tensions between its actors, and also to experiment with them in cases where more textual interpretations are an option. This line of research is becoming possible on the basis of the existing data, but its results will be present only after much more experiment in later versions of Quest.

Below, the data of a segment from I Kings 21 are presented. To demonstrate the direction of the continued research, as indicated here, the experiments with participant tracking are included.

abbreviations

LexP:	part of speech (derived from lexicon: e.g. line 8, אצל = noun)		
ls:	lexical set (subsets part of speech: אצל -3 = possible preposition)		
prf:	preformative	rtf:	rootf.-morpheme
vbe:	verbal ending	nme:	nominal ending
pnSfx:	pronominal suffix	VbFrm:	verbal form
Pers:	person	Numb:	number
Gend:	gender	State:	state
SynP:	part of speech (redefined by morphosyntax: אצל = preposition)		
Ptype:	Phrase Type		
Dtrm:	determination		
Sub-Phrase Types:	e.g.: +rgE = end of subphrase with *regens* relation -rcE = end of subphrase with *rectum* relation		
Constit:	Parsing of constituent (e.g. Predicate, Subject, Complement)		
Part.track:	phrase or suffix refers back to Line:x, phrase or suffix:y.		

1.2. I Kings 21:2 Word-level Data

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	6	conj	--	--	--	--	----	----	----	----	----	----	conj	וְ
IKg 21,02	-1	verb	J	D	0	..	....	ipfc	3rd	sing	masc	----	verb	יְדַבֵּר
IKg 21,02	0	prop	--	--	--	0	----	----	----	sing	zero	abs	prop	אַחְאָב
IKg 21,02	2	prep	--	--	--	--	....	----	----	----	----	----	prep	אֶל
IKg 21,02	0	prop	--	--	--	0	----	----	----	sing	zero	abs	prop	נָבוֹת

Form.Info: Line: 4 rel.to Line: 2 Code:372 Line: 5 Code: 64 Line:13 Code:200

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	14	prep	--	--	--	--	....	----	----	----	----	----	prep	לְ
IKg 21,02	-1	verb	0	0	0	0	....	infc	zero	zero	zero	zero	verb	אמר

Form.Info: Line: 5 rel.to Line: 4 Code: 64 Line: 6 Code:999

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	0	verb	0	0	H	..	....	imp.	2nd	sing	masc	----	verb	תֵּנָה
IKg 21,02	14	prep	--	--	--	--	1sg	----	----	----	----	----	prep	לִי
IKg 21,02	4	prep	--	--	--	--	....	----	----	----	----	----	prep	אֶת
IKg 21,02	0	noun	--	--	--	0	2sgM	----	----	sing	zero	cstr	noun	כַּרְמֶךָ

Form.Info: Line: 6 rel.to Line: 5 Code:999 Line: 7 Code:483 Line: 8 Code:503 Line: 9 Code:483

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	7	conj	--	--	--	--	----	----	----	----	----	----	conj	וְ
IKg 21,02	-2	verb	J	0	0	..	....	ipfW	3rd	sing	masc	----	verb	יְהִי
IKg 21,02	14	prep	--	--	--	--	1sg	----	----	----	----	----	prep	לִי
IKg 21,02	14	prep	--	--	--	--	....	----	----	----	----	----	prep	לְ
IKg 21,02	0	noun	--	--	--	0	....	----	----	sing	zero	cstr	noun	גַּן
IKg 21,02	0	noun	--	--	--	0	....	----	----	sing	zero	abs	noun	יָרָק

Form.Info: Line: 7 rel.to Line: 6 Code:483

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	11	conj	--	--	--	--	----	----	----	----	----	----	conj	כִּי
IKg 21,02	9	ppro	--	--	--	--	----	----	3rd	sing	masc	----	ppro	הוּא
IKg 21,02	-1	adj	--	--	--	0	....	----	----	sing	zero	abs	adj	קָרוֹב
IKg 21,02	-3	noun	--	--	--	0	....	----	----	sing	zero	cstr	prep	אֵצֶל
IKg 21,02	0	noun	--	--	--	0	1sg	----	----	sing	zero	cstr	noun	בֵּיתִי

Form.Info: Line: 8 rel.to Line: 6 Code:503

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	7	conj	--	--	--	--	----	----	----	----	----	----	conj	וְ
IKg 21,02	0	verb	>	0	H	..	....	ipfW	1st	sing	zero	----	verb	אֶתְּנָה
IKg 21,02	14	prep	--	--	--	--	2sgM	----	----	----	----	----	prep	לְךָ
IKg 21,02	-3	noun	--	--	--	J	3sgM	----	----	plur	masc	cstr	prep	תַּחְתָּיו
IKg 21,02	0	noun	--	--	--	0	....	----	----	sing	zero	abs	noun	כֶּרֶם

Form.Info: Line: 9 rel.to Line: 6 Code:483 Line:10 Code:108

Text	*ls*	*LexP*	*prf*	*rtf*	*vbe*	*nme*	*pnSfx*	*VbFrm*	*Pers*	*Numb*	*Gend*	*State*	*SynP*	
IKg 21,02	-1	adj	--	--	--	0	....	----	----	sing	zero	abs	adj	טוֹב
IKg 21,02	17	prep	--	--	--	--	3sgM	----	----	----	----	----	prep	מִמֶּנּוּ

Form.Info: Line: 10 rel.to Line: 9 Code:108

I Kings 21:2 Phrase and Clause-level Data

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	CjP	----	....	Conj	ו	ו	
IKg 21,02	VP	----	....	Pred	דבר	ידבר	
IKg 21,02	NPPN	det	....	Subj	אחאב	אחאב	
IKg 21,02	....	----	....	----	אל	אל	
IKg 21,02	PP	det	....	Cmpl	נבות	נבות	

Line 4; Func.Info: Sent: 3 Cl: 1 CLT:VbCl ; Phrases from: 1 to 4;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	....	----	....	----	ל	ל	
IKg 21,02	VP	idet	....	Pred	אמר	אמר	

Line 5; Func.Info: Sent: 3 Cl: 2 CLT:VbCl; Phrases from: 1 to 1; related as:Adjunct to Cl:-1;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	VP	----	....	Pred	נתן	תנה	Phr-Ln:4 phr:4
IKg 21,02	PP	----	....	Cmpl	ל	לי	Sfx-Ln:4 phr:2
IKg 21,02	....	----	....	----	את	את	
IKg 21,02	PP	det	....	Obj	כרם	כרמך	Sfx-Ln:6 phr:1

Line 6; Func.Info: Sent: 4 Cl: 1 CLT:VbCl ; Phrases from: 1 to 3;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	CjP	----	....	Conj	ו	ו	
IKg 21,02	VP	----	....	Pred	היה	יהי	Phr-Ln:6 phr:3
IKg 21,02	PP	----	....	PCmp	לי	לי	Sfx-Ln:6 phr:2
IKg 21,02	....	----	....	----	ל	ל	
IKg 21,02	....	----	+rgE	----	גן	גן	
IKg 21,02	PP	idet	-rcE	Cmpl	ירק	ירק	

Line 7; Func.Info: Sent: 5 Cl: 1 CLT:VbCl ; Phrases from: 1 to 4;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	CjP	----	....	Conj	כי	כי	
IKg 21,02	NPpr	det	....	Subj	הוא	הוא	Phr-Ln:7 phr:2
IKg 21,02	NPad	idet	....	PCmp	קרוב	קרוב	
IKg 21,02	....	----	....	----	אצל	אצל	
IKg 21,02	PP	det	....	Loc	בית	ביתי	Sfx-Ln:7 sfx:3

Line 8; Func.Info: Sent: 6 Cl: 1 CLT:NmCl Phrases from: 1 to 4;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	CjP	----	....	Conj	ו	ו	
IKg 21,02	VP	----	....	Pred	נתן	אתנה	Phr-Ln:8 sfx:4
IKg 21,02	PP	det	....	Cmpl	לך	לך	Sfx-Ln:6 phr:1
IKg 21,02	PP	det	....	Adjt	תחת	תחתיו	Sfx-Ln:8 phr:2
IKg 21,02	NP	idet	....	Obj	כרם	כרם	

Line 9; Func.Info: Sent: 7 Cl: 1 CLT:VbCl ; Phrases from: 1 to 5;

Text	*PType*	*Dtrm*	*Sub-Phrase Types*	*Constit*	*Lexeme*	*Word*	*Part.track*
IKg 21,02	NPad	idet	....	PCmp	טוב	טוב	
IKg 21,02	PP	----	....	Cmpl	מן	ממנו	Sfx-Ln:9 sfx:4

Line 10; Func.Info: Sent: 7 Cl: 2 CLT:NmCl Phrases from: 1 to 2; related as:Attrib to -2 Phr.

2. Linguistic layers

Virtually all of the existing databases of Biblical Hebrew text do not offer any data beyond word level. That is understandable, since beyond word level linguistic material is increasingly getting more difficult to analyze and to store in well defined data types. If from the viewpoint of philology one wants, for example, lists of 'phrases' or of 'clauses', one actually assumes a quite complicated set of operations. The idea underlying Quest.2 is, that the results of these operations should be present in the data. A user should not have to be aware of all of them. However, when preparing syntactically analyzed data, a study of the categories implied and the data types needed to store them, cannot be avoided. For example, asking for 'a subject', what is it exactly a user wants? A pronominal suffix (as in וְאֵינֶךָּ verse 5); a phrase (as נָבוֹת הַיִזְרְעֵאלִי in verse 4); an implied subject (as in וַיֹּאמֶר verse 4)? Or, when asking for 'an object', does one also want object clauses (כִּי מֵת נָבוֹת as in verse 16 or even in a fronted position as in verse 24 הַמֵּת לְאַחְאָב בָּעִיר in verse 24)?
Also important for the construction of the data for Quest.2 is the attempt to make the data as little as possible reflect a particular linguistic theory. The material should be organized in such a way that it could be re-organized and labeled according to the parameters of various linguistic types of research.
For that reason the database has both 'distributional data', i.e. linguistic data segmented and presented in sequential order and a 'functional labeling' of these data, i.e. the data are also recombined in an hierarchical order and labeled accordingly.

To clarify the distinction between distributionally analyzed linguistic data and functionally labeled linguistic data, I present a survey of the linguistic layers defined. From sentence-level and higher not all of them yet have been implemented in the data base. At those levels the construction of a text-linguistic database is a matter of continuing research and experiment. The examples of data types presented below may clarify to what extent grammatically analyzed data have been produced at the actual stage of the research.

The two types of linguistic analysis applied can be described as follows:
Distributional: defines a linguistic unit in terms of its composition from lower-level components; no gapping is allowed. These units are called 'atoms'.
Functional: defines a linguistic unit in terms of its function in a higher-level linguistic unit; gapping is allowed. These units are called: phrase, clause, sentence, paragraph.

Resulting data types and linguistic layers:
I. Phrase level
Phrase Atom

- a distributional unit: array of lexemes, conditioned by word order, part of speech and morhological features; no gapping;

Phrase

- 1 .. n phrase_atoms that are part of the same clause constituent; gapping (embedding of clauses) is allowed.
  Internal parsing: relations of subphrases (attributive, regens-rectum, etc.);

II. Clause level

Clause Atom

- a distributional unit: array of phrase_atoms, conditioned by order, phrase_atom types and lexical features; no gapping;

Clause

- 1 .. n clause_atoms that constitute one predication; gapping (embedding of clauses) is allowed.
  Internal parsing: relations of clause constituents (subject, object, etc.);

III. Sentence level

Sentence Atom

- distributional unit: array of clause_atoms, conditioned by order, clause_atom types, particular relators (אֲשֶׁר or הַ) ; no gapping

Sentence

- One clause, extended by 1 .. n clauses that function as one of its constituents, or repeat some of its constituents; gapping by embedding (e.g. a direct speech section) is allowed.
  Internal parsing: attributive clause, subject clause, object clause, etc.

IV. Paragraph level

Paragraph Atom

- distributional unit: array of sentence_atoms; part of a paragraph due to interruption by embedding (direct speech sections, or sub-paragraphs)

Paragraph

- functional unit: clustering of sentences. The first sentence introduces or renominalizes a new subject. Or, the first sentence changes the balance of actors substantially, e.g.: Complement becomes Subject and other Complements are introduced. The following clauses continue the pattern of actors.
  Internal parsing: participant tracking.

2.1. Phrase level-data

distributional data: phrase_atom

```
21,17,28   דְּבַר יְהוָה          [Noun + PrNoun ]
21,18      אַחְאָב               [PrNoun]
21,18      מֶלֶךְ יִשְׂרָאֵל     [Noun + PrNoun ]
21,08      בְּשֹׁם אַחְאָב       [Prep + Noun + PrNoun]
```

functional labeling of combinations: phrase

```
21,17,18   דְּבַר יְהוָה          [NP <Subj>]
21,18      אַחְאָב מֶלֶךְ יִשְׂרָאֵל   [NP + NP (apposition) <Subj>]
```

functional labeling: internal parsing of subphrase relations

```
21,18      מֶלֶךְ יִשְׂרָאֵל          [Noun{reg.} + PrNoun {rec.}]
21,08      בְּשֵׁם אַחְאָב            [Prp + Noun{reg.}+ PrN.{rec.}]
```

Sometimes phrases can have a rather complicated internal structure. See verse 22, where one finds in a PP: 4 regens-rectum contructions, 2 appositions and a parallel construction:

{ { {בֵּית יָרָבְעָם } {בֶּן נְבָט} } { { {כְּ {בֵּית בַּעְשָׁא} {בֶּן אֲחִיָּה} } { {כְּ {בֵּית יָרָבְעָם } {בֶּן נְבָט} } ו {כְּ {בֵּית בַּעְשָׁא} {בֶּן אֲחִיָּה}

Retrieval
At phrase level data retrieval is made possible of: lists of phrases, phrase types, or internal phrase relations. This means that from the perspective of the user of Quest.2 it is possible to search for phrases as such, in terms of their formal type (NP, VP, etc) or in terms of their functional labeling (Subject, Predicate, etc). Additionally the user should have the occasion to search phrases with a particular subphrase relation, e.g. apposition, attribute, genetive [regens-rectum].

2.2. Clause Level

distributional data: clause_atom

```
21,02      כִּי הוּא קָרוֹב אֵצֶל בֵּיתִי  [Conj + PronP + Adj + PP]

21,21  a        הַמֵּת לְאַחְאָב בָּעִיר [Rel + VP + PP + PP
       b        יֹאכְלוּ הַכְּלָבִים    [VP + NP]
       c                 וְ            [Conj]
       d             הַמֵּת בַּשָּׂדֶה  [Rel + VP + PP]
       e   יֹאכְלוּ עוֹף הַשָּׁמָיִם:    [VP + NP]
```

functional labeling of combinations: clause

```
21,02      כִּי הוּא קָרוֹב אֵצֶל בֵּיתִי  [AjCl]

21,21  a        הַמֵּת לְאַחְאָב בָּעִיר [PtcA]
       b        יֹאכְלוּ הַכְּלָבִים    [0Yqt]
       c                 וְ            -
       d             הַמֵּת בַּשָּׂדֶה  [PtcA]
       c   יֹאכְלוּ עוֹף הַשָּׁמָיִם:    [WxYq]
```

functional labeling: parsing of clause constituents:

```
21,02      [Cj + Subj + PredCompl + LocRef]
21,21  a   [Rel. + PredCompl + Adjunct + LocRef]
21,21  b   [Pred + Subj]
21,21  c   [Cj + Pred + Subj]
21,21  d   [Rel. + PredCompl + LocRef]
```

Retrieval
At clause level retrieval is made possible of: lists of clauses, clause types, or constituent types. This means that from the perspective of the user of Quest.2 it is possible to search for e.g. verbal clauses, nominal clauses, clauses with explicit subject, particular verbs with objects of particular complements.

2.3. Sentence level

The higher levels of syntactic analysis are present in the data, though in a number of respects still very much on an experimental basis and restricted to a number of clear cases. For example, in case of clauses taking the position of a constituent in an other clause (Subject or Object clauses) and in case of clauses that are part of a constituent (attributive clauses or clauses in a genetive connection), these clauses are connected to each other into a sentence atom. The actual state of the art means that at sentence level, and also at paragraph level, only distributional data are produced, i.e sentence_atoms and paragraph_atoms. A combination into functional units, i.e. sentences and paragraphs, has to wait until research has developed more insight into these categories and how to label them in terms of text-syntactic functions.

distributional data: sentence_atom

Composed by attributive clauses and constituent-level clauses:

```
21,21  A   הַמֵּת לְאַחְאָב בָּעִיר  יֹאכְלוּ הַכְּלָבִים
       B   וְ  הַמֵּת בַּשָּׂדֶה  יֹאכְלוּ עוֹף הַשָּׁמָיִם׃
```

This procedure of combining clauses into sentence is already becoming quite complicated in case a text uses additional clauses that grammatically continue one of the dependent clauses in a sentence. This is demonstrated in verse 4.d (see below), where אֲשֶׁר דִּבֶּר is continued by וַיֹּאמֶר.
These complications are compensated for by the coding system that is used to store the clause relations of a text. In the analyzed data, clause_atoms with כִּי

and אִם or וְ are connected to foregoing clauses with the help of a system of three digits. For example: if a clause_atom is connected to a foregoer by code 521, the first digit indicates the conjunction (כִּי = 5), the second one the verbal tense of the depending clause (qatal = 2) and the third one the verbal tense of the foregoing clause (yiqtol = 1). In this way clauses are combined already, even when much more research is required to label the various clause_atom combinations as sentences with appropriate syntactic categories.

functional labeling of combinations: sentence

As indicated above, at this moment the data will present the user only with encodings of the relations between clause_atoms, not yet with labels indicating syntactic function. Therefore, a functional unit 'sentence' as such has not yet been labeled in the data. The database itself should be used for the continuing research into a systematic functional labeling of sentence constructions. For example: one can search for clause_atoms connected with the code '521' and then test on what further conditions this connection could be labeled 'causal' or should be labeled otherwise.
Thus, for the time being the data type 'sentence' is equal to the data type sentence_atom.

functional labeling: parsing of clause relations

21,21	A		יֹאכְלוּ הַכְּלָבִים [Obj.Cl] הַמֵּת לְאַחְאָב בְּעִיר
	B	:יֹאכְלוּ עוֹף הַשָּׁמָיִם	[Obj.Cl] הַמֵּת בַּשָּׂדֶה [Embedding] וְ

As a result, even when a general labeling of sentences according to grammatical caregories has been postponed, a restricted number of clause relations has been defined: the majority of them being object clauses, subject clauses, adjunct clauses and attributive clauses.

Retrieval
At sentence level retrieval is restricted to the retrieval of clause combinations where one of them fills a particular constituent slot of a clause: subject, object, complement, adjunct, or in case a clause has a relation to a constituent of the higher level clause: attributive, or genetive relation. Other cases are clauses that resume a preceding *casus pendens*.

2.4. Paragraph level

At the paragraph level a similar, distributional procedure of data production has been applied. Even when not all combinations of clauses into sentences yet have been labeled in terms of functional syntax, it is possible to connect larger chains of clause_atoms into higher levels segments of text, i.e.: paragraph_atoms. The production programme assumes the beginning of a new paragraph_atom when certain clause types are present: Wayyiqtol + Subject; W-X-Qatal (X being the subject). If earlier subjects remain present, for example by pronominal reference, the new paragraph_atom is regarded a sub-paragraph. Below, the example of I Kings 21 verse 4 and 5.

distributional data: paragraph_atom

verse 4-5

	sent.atoms		paragraph_atoms
4.a	s.a 1	וַיָּבֹא אַחְאָב <Subj> אֶל בֵּיתוֹ	§ 1
4.b	attrib.	סַר וְנָעַף עַל הַדָּבָר	
4.c	attrib.	אֲשֶׁר דִּבֶּר אֵלָיו נָבוֹת <Subj> הַיִזְרְעֵאלִי	
4.d	continued	וַיֹּאמֶר	
4.e	s.a 2	לֹא אֶתֵּן לְךָ אֶת נַחֲלַת אֲבוֹתָי	§ 1.0
4.f	s.a 3	וַיִּשְׁכַּב עַל מִשָּׁתוֹ	§ 1
4.g	s.a 4	וַיָּסֵב אֶת פָּנָיו	
4.h	s.a 5	וְלֹא אָכַל לָחֶם:	
5.a	s.a 6	וַתָּבֹא אֵלָיו <pron.ref> אִיזֶבֶל <Subj> אִשְׁתּוֹ	§ 1.1
5.b	s.a 7	וַתְּדַבֵּר אֵלָיו	
5.c	s.a 8	מַה זֶּה רוּחֲךָ סָרָה	§ 1.1.0
5.d	s.a 9	וְאֵינְךָ אֹכֵל לָחֶם:	

Paragraph atoms are coded by digits indicating their levels of dependency, e.g. 1 (§ 1) , 1.1 (one level of embedding within § 1), etc. In case of a direct speech section a "0" is added to the code, e.g. 1.0 (a direct speech paragraph within § 1).

functional labeling of combinations: paragraph

The construction of paragraphs from paragraph_atoms, is even more a matter if experiment than it is the case with sentences. The principle guiding the actual research is to define a paragraph as consisting of a sequence of sentences, grammatically connected, and characterized by a consistent set of participants. For example, in case of the paragraph_atoms constructed in verse 4 and 5, one sees a set of three interacting participants:

אַחְאָב	-	-
ו(sfx)	נָבוֹת	
ו(sfx)	-	אִיזֶבֶל

These data may be helpful to the translator and the exegete to reflect on procedures, active in the source text, of the explicit marking of participants, or of deliberate avoiding them in a particular context. How to interprete these, or how to render these phenomena in a translation?

Retrieval:
Paragraph_atoms can be used as definition of context. For example, one could search for particular clauses to be present within the same paragraph_atom. Also the user should have the option to ask for a presentation of the syntactic hierarchy of a segment of text.

3. The Quest.2 data and Retrieval.

The previous section has demonstrated the line of actual research performed in preparation of a syntactic database of Hebrew text. It has also made clear where this research still is very much experimental in nature. Distributional and functional data types as presented above, have not yet been completed for all books of the Hebrew Bible. Until now, a full analysis has been produced for the Pentateuch. It will be available in the data of Quest.2. For the time being, the data of the other biblical books will be present with lower levels of syntactic analysis:

	syntactic text structure of individual chapters; text types				
	morpheme	lexeme	phrase	clause	sentence
Jos. - Kings:	morpheme	lexeme	phrase	clause	-
Isa. - Chron.:	morpheme	lexeme	phrase	clause_at.	-

The level of analysis present means that with the actual version of Quest.2 one can perform in the entire Hebrew Bible searches for data beyond word level: phrases and clauses (or clause_atoms). Joshua - Kings have clauses with also the internal parsing of constituents included.
Since the research on sentences and paragraphs is still going on, the Quest.2 programme will not yet allow the retrieval of clause relations, or the presentation of textual structures, in areas other than the Pentateuch.

Below I will present some questions of exegesis or translation with the kind of retrieval the exegete or the translator may require to have material helpful for textual interpretation. I will show some searching done with queries used in the actual Quest.1 programme[5] and I will suggest where more options are present or would be needed in Quest.2.

Continued research in retrieval options, available, or to be made available in Quest.2 concentrates on three topics:

First: how could one search on a more abstract grammatical level?
Second: how to present to the user segments of text according to their grammatical structure?
Third: how to allow the user to make an instruction, without the user having to know too much of a special query language? In Quest.1 a special language

---

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

has been designed for the composition of queries: QML (Quest menu language). In Quest.2 it should be possible to point at particular textual elements, asking the programme to describe the grammatical and lexical elements they have, thus allowing the user to select from them elements needed for a query. As long as one is interested in 'words' such procedures are not too difficult to program. But beyond word-level such is a quite complicated task for a programmer. From the user's perspective, however, this approach will be much more comfortable than the actual approach of demanding the user to compose queries in an independent way.

3.1. Some examples of questions and searches

A translator of I Kings 21 will find some rather exceptional constructions of attribute clauses being used in verse 8 and 11.

8 וַתִּשְׁלַח ... אֶל הַזְּקֵנִים וְאֶל הַחֹרִים אֲשֶׁר בְּעִירוֹ הַיֹּשְׁבִים אֶת נָבוֹת׃ 8
11 וַיַּעֲשׂוּ אַנְשֵׁי עִירוֹ הַזְּקֵנִים וְהַחֹרִים אֲשֶׁר הַיֹּשְׁבִים בְּעִירוֹ 11

The interesting point here is the combination of הַ and אֲשֶׁר referring to the same Noun phrase. Moreover one can observe a difference in syntactic construction between verse 8 and verse 11. The user of Quest.2 should have occasion to ask for a presentation of these structures on screen.
For example:

I Kings 21,8

```
[<Co> אל הזקנים ואל החרים] [<Ob> הספרים] [<Pr> תשלח] [<Cj>ו]  45  N  Way0  3sgF 08
                                [<PC> בעירו] [<Re> אשר]  |   |    |    46  N  NmCl  ---- 08
                        [<Aj> את נבות] [<PC> ישבים] [<Re> ה]  |   |    47  N  ptcA  -plM 08
                              [<Co> בספרים] [<Pr> תכתב] [<Cj>ו]  |    |    48  N  Way0  3sgF 09
                                                  [<Pr> לאמר]  |   |    |    49  N  Infc. ---- 09
```

I Kings 21,11

```
        [<Su><ap> אנשי עירו / הזקנים והחרים] [<Pr> יעשו] [<Cj>ו]  59  N  WayX  3plM 11
                                  [<Lo> בעירו] [<PC> הישבים] [<Re> אשר]  |    |    60  N  NmCl  ---- 11
      [<Su> איזבל] [<Co> אליהם] [<Pr> שלחה] [<Re> כאשר]  |    61  N  xQtl  3sgF 11
```

No doubt, a Hebrew scholar or a translator would like to see more cases of such constructions with both הַ and אֲשֶׁר, in order to find whether they are unusual indeed, and also wether the difference in construction between verse 8 and 11 requires a variation in the translation. Possibly they also expose a linguistic variation or a tendency of linguistic change in classical Hebrew.
The user of Quest.1 can search for parallels by composing (with QML) a query with precise definitions of the words and features, indicating also the context required. For example:

```
% filename: AsjerHa.TOP
% START QUERY
     definitions                              (where to search)
          domain bible
     topograph                                (what to search)
          [verse
               [word lexeme ">CR" ]           (אֲשֶׁר)
               [word <- 2 ]                   (distance max. 2)
               [word lexeme "H" ]             (הַ)
               [word verbal_tense participle]
          ]
     commands                                 (how to present)
          mark( 2, word 1)                    אֲשֶׁר
          mark( 3, word 3)                    הָ
;
```

Some results:

I Ki 12:8 וַיַּעֲזֹב אֶת עֲצַת הַזְּקֵנִים אֲשֶׁר יְעָצֻהוּ וַיִּוָּעַץ אֶת הַיְלָדִים אֲשֶׁר גָּדְלוּ אִתּוֹ הָעֹמְדִים לְפָנָיו׃

Ez 20:29 וָאֹמַר אֲלֵהֶם מָה הַבָּמָה אֲשֶׁר אַתֶּם הַבָּאִים שָׁם

IIChr 10:8 וַיַּעֲזֹב אֶת עֲצַת הַזְּקֵנִים אֲשֶׁר יְעָצֻהוּ וַיִּוָּעַץ אֶת הַיְלָדִים אֲשֶׁר גָּדְלוּ אִתּוֹ הָעֹמְדִים לְפָנָיו׃

IChr 11:10 וְאֵלֶּה רָאשֵׁי הַגִּבֹּרִים אֲשֶׁר לְדָוִיד הַמְתְחַזְּקִים עִמּוֹ בְמַלְכוּתוֹ . . .
[cf.the parallel text II Sam 23,8: no ... הַמְתְחַזְּקִים]

IIChr 36:8 וְיֶתֶר דִּבְרֵי יְהוֹיָקִים וְתוֹעֲבֹתָיו אֲשֶׁר עָשָׂה וְהַנִּמְצָא עָלָיו
[cf. II Kings 24,5 no parallel]

One can observe that the use of הַ . . . אֲשֶׁר increases at later stages of Biblical Hebrew, which may point at linguistic variation being visible in these data.
From a practical point of view, it is clear that the user of Quest.1 needs some special skills for preparing the searching. One needs to compose a Query based on ones own reading and grammatical understanding of the text of I Kings 21, 8 and 11. The new version, Quest.2, will be more helpful in these matters. The user of the programme will be guided by the grammatical analysis of the texts as demonstrated on the screen, and will have occasion to compose a Query by making selections from the linguistic elements and features presented. For example, search for: Nominal clause with אֲשֶׁר followed by participle clause with הַ.

A comparable example is the special clause construction in I Kings 21, 23:

וְגַם לְאִיזֶבֶל דִּבֶּר יְהוָה לֵאמֹר הַכְּלָבִים יֹאכְלוּ . . .

In stead of searching for words, e.g.: גַּם followed at some distance by a verb in Qatal form, one should now in Quest.2 have the option to search for verbal clauses with the elements: **גם** + PP + Qatal.

Some results (as yet produced with the help of Quest.1):

jeremi 52: 10 וישחט מלך בבל את בני צדקיהו לעיניו וגם את כל שרי יהודה שחט ברבלתה .

II_chr 36: 13 וגם במלך נבוכדנאצר מרד אשר השביעו באלהים ויקש את ערפו ויאמץ את לבבו משוב אל יהוה אלהי ישראל .

A wish, I still have, is to construct a more intelligent database that would by itself present to the user suggestions of interesting literary parallel texts. For example, Jeremiah 52,10 is related to II Kings 25,7, where one, hoewever, does not find a parallel **וגם**-clause. II Chronicles 36,13 is parallel to Kings 24,20, where one, also does not find a parallel construction using **וגם**, but a wayyiqtol-clause. Is this a literary difference, or does it suggest linguistic change?

These few examples may have demonstrated that the research in constructing a textlinguistic database and its use for continued textual and linguistic analysis together constitute a fascinating discipline of scholarly work.
I do hope this work will continue to serve both the linguists and the exegetes in their tasks: analyzing language and understandig Scripture.

Quest.2 will be the first software package to allow access to these data to a broader community of students of Bible. In the mean time the team in Amsterdam will proceed with computer-assisted linguistic analysis, using the new data made for continued linguistic and exegetical research.

Literature

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

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,
'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,
'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

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

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

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*

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

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