

Aspects of a New Morphology

– Harald Schweizer –

A. Introduction

In April 1986 I stayed for 5 days in Amsterdam, together with Winfried Bader. The date for us is easy to remember – on macabre reasons –, because at the same time the reactor of Tchernobyl exploded. Going back to the station we read the headlines and we only grasped that a severe accident happened. But we were'nt able to read and understand sufficiently the foreign language.

So, we met the *Woerkgroup Informatica*. Eep Talstra had offered to program some search procedures which I had in mind. At that time we had many theoretical ideas but not yet enough practical experiences in computer assisted analysis of the Hebrew Bible. So we gratefully accepted the invitation. Our project was to analyze all instances of the Hebrew verb QRB. And in doing this I wanted to sketch the outline of a new kind of morphology.

Some observations may help to illustrate what I'm speaking of:

1. Starting our work I asked Eep Talstra to do the search algorithms in respect to the consonantal text of the Hebrew Bible. Talstra asked: »You are not interested in our taggings?« I said: »No!«. Talstra and Postma were heavily astonished. – That incident illustrates our subject: Talstra understood and continues to understand »morphology« in the normal and widespread sense, namely as the task to describe a two sided object: we are confronted with some written or spoken elements of a language **and** we have to define, what functions of content are combined with these elements of the expression level. A computer – of course – is only able to handle expressions, to compare whether they are identical or not. A computer cannot understand what an expression means. Therefore, a competent speaker of that language must add tags to the expressions, tags which indicate the function of content which has to be combined with that element of expression. So, we may take the expression *nšym* and add by tags the function of contents like: »Plural of *šh*, plural somewhat irregular, because the ending normally standing for masculine plural here is used at a noun designating a feminine entity.« – Expression **and** semantic function: that twofold item is the basis of the morphology as we know it.¹

¹ The example I chose deliberately allows even more insights: By the example of the question of »Genus« it is to be

2. A second observation confirmed the impressions I formulated up to now: In the office of the Woerkgroup Informatica lay the grammar of LETTINGA on the table. Apparently that book was the theoretical framework for all the analyses and the growing database. – One may proceed in such a way and meanwhile there are many publications which used the amsterdamite tagged hebrew text with very interesting results.² That's a type of morphology that works, that can be used. And I think the electronic morphology of the masoretic text edited by Alan Groves is of the same type and it is regularly used – as I know – by some people in Tübingen at any time when they are tackling a new hebrew text.

In contrast to that I want to give an outline of a morphology which doesn't yet work, which at the moment is only a scientific project, an object of scholarly debate.³ The need to search for a new understanding not only of »morphology« but also of the other levels of language analysis (such as syntax, text or discourse analysis) follows from modern General Linguistics or Semiotics growing up since the beginning of that century. The most important name is FERDINAND DE SAUSSURE. The important grammars of Hebrew have been written in the 19th century. Then the conception of a 'historical grammar' predominated.⁴ And I want to contend that even concerning modern languages there aren't yet grammars which really follow the basic insights and axioms of DE SAUSSURE. So we aren't speaking here of a special problem of the Hebrew grammar but of the question how to analyze any language according to the principles of modern Semiotics.

What I'm telling you can be read in the Internet as hypertext (in German):

<http://www-ct.informatik.uni-tuebingen.de/ct/morph.html>

[There the folios can be found completing the actual text by practical examples]

seen that the traditional morphology includes problematic terms. What does »Genus« really mean? An item of the world outside language? Or purely a linguistic item? In the first case the term »masculine« is wrong, in the second case the term »Genus« is misleading and should be replaced. And further on: the question of »rules and exceptions« has to be answered. Who is allowed to label some linguistic item by »exception« when on the other side one has to concede that the phenomenon in that language has a widespread contestation? The regular plural of 'šh is nšym despite the fact that ym as ending of a »feminine« noun looks rather strange. I would like to conclude that every linguistic item in a language, which is not textcritically corrupt and contributes without any problems to the running communication is »regular« and has to be accepted by the grammarian. Therefore the term »exception« has to be forbidden.

² The last work I got acquainted is that of N. WINTER-NIELSEN.

³ Cf. SCHWEIZER, H (1991QRB) ch. 3.

⁴ There were only revisions of these grammars in our century (cf. GESENIUS-KAUTZSCH) or reuses of the old conception (cf. BERGSTRÄSSER, MEYER) or amplifications in respect to the material used (cf. BROCKELMANN – with at least some slight methodological innovations). The grammar of LETTINGA wasn't written for scholarly purposes but as textbook for teaching purposes.

B. The essence of my position

1. The twofold and inhomogeneous object of the old morphology has to be separated. The result will be two homogeneous levels of linguistic research:
 - 1.1 expression level (without any semantic function): now the computer will be able to analyze automatically the expression level of any given language by distribution analysis: which morph can be distinguished from another and of what kind are the combinations among these morphs?⁵
 - 1.2 content analysis (semantics / Pragmatics): Normally grammars do not offer chapters dealing with an explicit content analysis. Contrarily, some elements of content analysis are included in morphology or syntax. The results sometimes are ridiculous due to a perspective taken over without reflection from the greco-roman tradition.⁶
 - 1.3 Several authors from different linguistic schools could be cited expressing the conviction that the old morphology has to be split up.
 - 1.4 The gain would be a twofold one because the levels of research now are separated and therefore of homogeneous manner each one. And they do not continue to cat back each other:
 - 1.4.1 Level of expressions: The genuine ability of the computer to read, write, modify, sort and count expressions – whatever be their meaning – can be used extensively. By algorithms of decomposition the whole vocabulary of a language can be reduced to 'atoms', i.e. a composite structure beyond the level of pure letters. The letters as such do not characterize an individual language. But the expressions in the composite structure of 'atoms' do.⁷

⁵ I told you, that Winfried Bader and I were not able to read sufficiently the newspapers in Amsterdam. Why not? Because we didn't know the conventions of the expression level in Dutch. Our problems lay not on the level of contents, but in the question: which morphs are used in Dutch and in which combinations are they accepted and to which contents these morphs or agglomerations of them are pointing? – A computer could give answers at least to the first question: morphs/agglomerations.

⁶ E.g. – as I quoted – in Hebrew the »wives« are masculine, whereas the plural of »father« is feminine. »to die« – according to RICHTERS grammar – is an action, whereas I would say, it is a process. The terms »finite« and »predicate« usually are confused, the same with »accusative« and »object« and other terms. The important field of »modalities« is only activated when explicit modals can be found in a clause. It is overlooked that they form an independent and differentiated category in the perception of utterances. The opposition »active vs. passive« often is overestimated whereas many authors of standard grammars are not even able to differentiate the types of »static« relations. Mostly »existential clauses« are not dealt with and it is even not seen that there is a fundamental difference between them and ordinary predications. The question of »definiteness« often is confused with the presence of an article, or the interrelation with »quantification« is not seen. So »every game« is not »indefinite singular« but »definite distributive plural«. – We could continue to cite many more of such 'highlights' of a content analysis in deficit.

⁷ With the background of a distributional analysis we can say that the string *ction* certainly does not belong to a german word, nor does *sh*.

1.4.2 Level of contents: we now are released to develop a forced content analysis, much more precise and coherent than what usually is offered in traditional grammars. – I only want to mention that point. It is not the subject of our meeting.⁸

2. The structure of the algorithm

During the last two years I restarted working on morphology. As corpus we chose a novel from Günter Grass: »Die Blechtrommel«. A student wrote the programs in C + +. Several seminars and colloquia dealt with our results. The work is not yet finished. But I'm shure that at least we reached the »point of no return«.

So I could offer you an illustration which is a little bit dangerous for me. For I didn't want to list our results and insights valid for the German language. Therefore I tried to simulate the algorithm and its results applied to the English language. So there is not a real computer analysis behind but only the search in my brain or in my dictionary. But I'm convinced that the main principles of our distributional analysis can be illustrated in such a way.

[Folios in *morph.html*]

⁸ But for an extensive illustration what I'm speaking of see our analysis of the Joseph story: Schweizer (1995).