CHAPTER EIGHT

THE PROCEDURE OF THE CALAP ANALYSIS

8.1 INTRODUCTION

In the preceding chapter we have described the model of linguistic and textual analysis that has been applied in the present study. In the present chapter we will elaborate on the procedure of the computer-assisted analysis and the computer-programs used in this analysis.

Some of the computer programs were developed in the CALAP project, others were adaptations of programs that had been used previously in the WIVU for the linguistic analysis of Biblical Hebrew texts. Although the adaptation to Syriac could have been done more directly by substituting the Syriac linguistic data for the Hebrew data, one of the aims of the CALAP project was to do more than that. Both from a methodological view and because of practical considerations, we thought it more appropriate to develop language-independent tools for linguistic analysis. This explains why in the following paragraphs a clear distinction is made between language-specific auxiliary files—such as description of the morphology or a lexicon—and language-independent programs that use these auxiliary files in the linguistic analysis.

The use of auxiliary files or language-definition files, which contain grammatical and lexical information, is one of the main characteristics of the CALAP procedure of linguistic analysis. Thus rather than tagging a ‘perfect 3rd pers. masc. sing. Pael’, it is the computer program that produces such an analysis on the basis of a combination of grammatical information from the auxiliary files and the encoded text. As a consequence, the observations that led to a certain analysis can always be retrieved.

The auxiliary files are used in the analysis of documents containing the text in question and the results of the analyses at earlier stages.

---

1 For a description of these programs see Talstra–Sikkel, ‘WIVU-Datenbank’.
Accordingly, for each program the input contains files that are the output of preceding steps in the analysis. Even though each program generates new data, the old information from earlier analyses is preserved. This makes it possible at each phase of the analysis process to reconsider decisions made previously.

8.2 THE GRAPHIC TEXT

The first step in the analysis of a written text is the abstraction of the graphs that occur in the actual manuscripts or editions towards the graphemes they represent. Graphs are the written or printed realizations of graphemes. The realizations may differ, for example, in the script used (thus ḫ and ḫ represent the same grapheme in respectively an Estrangelo and a Serto font) or in their conventional form according to the place in a word (compare the Kaph in لحم and حلم).

The subsequent steps in the textual analysis require the preparation of a text that contains unique and unequivocal representations of each grapheme. In the CALAP project this is done with the computer program 'pil2wit'.

The input for this program is the so-called ‘running text’, in the case of Sirach the Syriac text from the Leiden edition (in preparation for publication) in electronic form (file-name: 'sirach'). It contains the consonants in transliteration, diacritics and interpunction, variant readings and instructions regarding the variant readings which should be accepted in the main text, and comments from the researcher (optional). For example, the running text of Sir 48:1 in 'sirach' is²

```
1 'dm' dg m 'nby' dgm' lnwr'. wmlth yqd' 'yk twr': dm$tgr [dm$tgr / + 7a1]
```

In the printed Leiden edition this will appear as

```

The output of pil2wit is the ‘graphic text’ (file-name: ‘BenSira’). This is a transliterated text according to an established format that enables

² The caret marks a point over the preceding letter. The semicolon marks the end of the verse. 7a1 repeats קב. Such a ditography occurs eight times in Syr in 7a1. The repetition will not be selected as the main text and in the Leiden edition it will appear in the first critical apparatus.