Title: Addendum to: A system of control characters for Ancient Egyptian hieroglyphic text (L2/16-210)
From: Mark-Jan Nederhof & Vinodh Rajan & Johannes Lang (University of St Andrews, UK),
Stéphane Polis & Serge Rosmorduc (Ramses Project, Université de Liege, Belgium & CNAM, Paris),
Tonio Sebastian Richter & Ingelore Hafemann & Simon Schweitzer (Thesaurus Linguae Aegyptiae, Berlin-Brandenburgische Akademie der Wissenschaften, Berlin)
To: UTC
Date: 2016-08-04

See Section 9 of L2/16-210 for more details.

A Structure of hieroglyphic encoding (variant with operator precedence)

For a sequence of signs and control characters to have their intended meanings, it should comply with the following (Backus-Naur) specification. Lower-case non-bold names are classes. Bold-face names represent characters, with upper-case boldface names representing particular characters, and **sign** representing any hieroglyph. The pipe symbol | separates alternatives. Square brackets [ ] indicate optional elements, round brackets followed by an asterisk ( )* indicate repetition zero or more times, and round brackets followed by a plus symbol ( )$^+$ indicate repetition one or more times.

In the below we need nonterminals for different levels of operator precedence. The variable of i can stand for one of the levels **1**, **2** or **3** (more powerful systems could allow for **4** and **5** as well). Also different control characters exist for these values. So we have for example the binary operators $\mathbf{HOR_1}$, $\mathbf{HOR_2}$, $\mathbf{HOR_3}$, all for horizontal grouping. These may be realised in different ways, e.g.

- Each of $\mathbf{HOR_1}$, $\mathbf{HOR_2}$, $\mathbf{HOR_3}$ is a different (single) character.

- $\mathbf{HOR_2}$ is represented by twice $\mathbf{HOR_1}$, $\mathbf{HOR_3}$ is represented by three times $\mathbf{HOR_1}$.

- There are dedicated characters that raise the level of binding, one for each level, if they are appended behind $\mathbf{HOR_1}$.

- There is one dedicated character, which is appended once or twice behind $\mathbf{HOR_1}$ to obtain $\mathbf{HOR_2}$ or $\mathbf{HOR_3}$.

The following states that a fragment of hieroglyphic text consists of one or more groups of the lowest level, and that a general group may be a basic group, a horizontal group or a vertical group.

fragment ::= (group$_\mathbf{1}$)$^+$
group$_i$ ::= basic_group$_i$ | horizontal_group$_i$ | vertical_group$_i$

The following states that horizontal and vertical groups consist of subgroups separated by normal separators or by joining separators. A subgroup of a horizontal group may not be another horizontal group and a subgroup of a vertical group may not be another vertical group. If a horizontal group contains a vertical group, the latter must be one level of precedence up.

horizontal_group$_i$ ::= hor_subgroup$_i$ (hor_separator$_i$ hor_subgroup$_i$)$^+$
hor_subgroup$_i$ ::= basic_group$_i$ | vertical_group$_{i+1}$
hor_separator$_i$ ::= $\mathbf{HOR}_i$ | $\mathbf{HORJOIN}_i$
vertical_group$_i$ ::= vert_subgroup$_i$ (vert_separator$_i$ vert_subgroup$_i$)$^+$
vert_subgroup$_i$ ::= basic_group$_i$ | horizontal_group$_i$
vert_separator$_i$ ::= $\mathbf{VERT}_i$ | $\mathbf{VERTJOIN}_i$

A basic group is empty, or it is a core group, surrounded by optional insertions. A core group is a sign by itself, or is a stacking of two signs or horizontal or vertical groups. The subgroups must all be one level of precedence up.

basic_group$_i$ ::= **EMPTY** |
                [group$_{i+1}$ **INSERT_T_L$_i$**] [group$_{i+1}$ **INSERT_B_L$_i$**] core_group$_i$
                    [**INSERT$_i$** group$_{i+1}$] [**INSERT_T_R$_i$** group$_{i+1}$] [**INSERT_B_R$_i$** group$_{i+1}$]
core_group$_i$ ::= **sign** | core_subgroup$_i$ **STACK$_i$** core_subgroup$_i$
core_subgroup$_i$ ::= **sign** | horizontal_group$_{i+1}$ | vertical_group$_{i+1}$

A different way of looking at the above is that for the same level of precedence, stacking binds more tightly than the insertion operators, which bind more tightly than the horizontal separators, which bind more tightly than the vertical separators.

Practical implementations may restrict themselves to a subset of the fragments that are possible in principle. In particular, if a core_subgroup is a horizontal_group or vertical_group, then one may assume the subgroups therein are all individual signs.

B   The characters

The following defines the individual characters used in the above description of the structure.

HOR: "normal separation in horizontal group"

HORJOIN: "joining separation in horizontal group"

VERT: "normal separation in vertical group"

VERTJOIN: "joining separation in vertical group"

EMPTY: "a zero-width and zero-height (empty) character"

INSERT, …: "insertion of a group into the bounding box of another, at the indicated position"

STACK: "stacking of two groups"