ISO/IEC JTC/1 SC/2 WG/2
Universal Multiple-Octet Coded Character Set (UCS)
Secretariat: ANSI

Title:	**Requirements for Coded elements**
	(Proposal: Addition of new annex onto ISO/IEC 15285)
Doc. Type:	National body contribution
Source:	**Japan and MLIT** projects -tks
Project:	02.18
Status:	For discussion at Beijing meeting
Date:	2000-03-13
Distribution:	JTC 1/SC 2/WG 2
Reference:	WG2 N2148

This document proposes an annex to ISO/IEC 15285 (Character/Glyph model) as annex-X.

The annex discuss about the coding of syllabic characters.

The proposed annex –X includes four parts.
1. Problem description and sample
2. Requirements
3. Possible solution
4. Other requirements

1. Problem statement
WG2 document N2148 describes about the potential problems associated with Syllabic Characters. While Beijing WG2 meeting, WG2 members are welcome to exchange the information and idea about this issue with Dr. Tashiro and Prof. Kataoka.
If required, Japan will provide further contribution about the problem details.

The document discusses about the ambiguity problem of the coded character sets for syllabic script in general by taking a sample cases.

2. Requirements
As a conclusion, there are two basic requirements for coded elements and their strings of syllabic scripts. (to eliminate an ambiguity)

- The string should be normalized.   No different sequences for one syllable.

- The discrimination of the syllable (or equivalent) should be defined only one resultant group.

   (for render and process)

- Once syllable is defined, the glyph should be defined without any ambiguity.

- In addition some more requirements are expressed in the N2148.

The discrimination and glyph shape issues are highly related with the code design, while sequence is related to the usage of the code.

3. Possible Solution
Resolving the coding issue, there are a couple of recommended methods to design the code that meets with the requirements.

3-1 If sequence of the elements defined by handwriting tradition does not have any discrimination problem.   Use them as they are.   No trick is needed.
3-2. If Not. Then there are choices;
3-2-1   Try to assign multiple codes for same element.   For example, one for independent form and another for other presentation forms.   (Not fully, if fully different codes, it is not per the TR 15285.   Thus for very limited selected one.)
3-2-2   Try to assign a code for selected combined elements.   This method has been useful for Korean Hangul Jamo.
3-2-3.   Try to create a discriminator or equivalent such as form selector.   This method is similar to Mongolian case.   The discriminator may be pure discriminator, joiner, modifier and/or terminator (whatever easy to use and implement)
3-2-4.   If all of above are not practical, then assign a code for the combined result.   This method is used for CJK ideograph and Hangle.
3-2-5.   What else?
3-2-6   Language dependency problem might be open issue.

4. Other requirements
In addition to above major points, the annex X may cover several points that the paper by Dr. Tashiro recommend in N2148.

One of the possible objection for the proposal would be human-unfriendliness of the proposed coded elements and their strings.
Another proposal for annex-Y is discussing this problem and proposing possible solution.

----end---