**Observations on the testing of Proposed Bengali LGR Rules on corpus.**

1. Halant+C+halanta+C ( here halanta+ra one after the other)

যোগেন্দ্র্র: valid

As there is no blocking of number of consonant after a halanta this type of labels have been tagged as VALID.

2.

শু-: invalid

যোগ্য-করা: invalid

যো-হুকুমের: invalid

যদু-বংশ: invalid

All hypenated labels are tagged as INVALID as hyphens not included in the MSR, i.e. are ineligible for the root zone (digits, hyphen)

3. Nukta character (joint vs. exploded)

a. যোগমায়া: invalid

যথাসময়ে: invalid

যথানিয়মে: invalid

যাবতীয়: invalid

উদয়কাল: invalid

b. উড়োজাহাজ: invalid

 উড়ুনি: invalid

য় (YA) (U+09DF) as a consolidated character is not included in the MSR. Moreover, to have a label tagged as VALID one has to type য + nukta (U+09BC) to form য় (YA). Otherwise the labels with directly typed য় (U+09DF) will be tagged as INVALID in all tests. The same implies for ড় and ঢ়, which are to be types as ড (U+09A1 ) and nukta, ঢ (09A2) and nukta respectively.

Accommodated in WLE rules (see section 7.1, Rule no. 1(h) in proposed Bengali LGR).

4.

খানা৩:

Tagged as INVALID as numbers are not included in the MSR

5.

যাওয়া/হাত: invalid

All punctuated labels are tagged as INVALID as hyphens not included in the MSR, i.e. are ineligible for the root zone (digits, hyphen). This particular label has another possible reason for INVALID i.e. the য় (YA) issue mentioned in no. 3 above.

6. Explicit halanta followed by non-joiner (in the middle or at the end of a label)

উদ্‌: invalid

উহ্‌: invalid

উদ্‌ভ্রম: invalid

উদ্‌ভ্রান্ত: invalid

উদ্‌যাপন: invalid

শ্বন্‌: invalid (L-3965)

Explicit halanta followed by non-joiner (U+200C) is tagged as INVALID as non-joiner is not allowed. Otherwise explicit halanta at the end (without a non-joiner) is marked VALID, for e.g. উদ্ U+0989 (উ) U+09A6 (দ) U+09CD (্).

7.

Wrong typing

শৌ-িক: invalid

Probably the label is tagged as INVALID due to wrong typing.

8.

Misc incorrect typing or mistakes:

শুকম-ল: invalid

শুি-: invalid (hyphen + matra after matra)

শু-িকা: invalid