Title: Proposal for new provisional Unihan database property: **kJapanese**
Author: Ken Lunde
Date: 2022-08-23

The existing provisional kJapaneseKun and kJapaneseOn properties provide 11,296 and 13,177 Japanese readings, respectively, which together cover 13,395 unique CJK Unified Ideographs. These readings are expressed using uppercase Latin characters based on a modified Hepburn transliteration system.

The proposed provisional Unihan database property, **kJapanese**, provides one or more Japanese readings that are expressed in hiragana or katakana, and covers 51,582 unique CJK Unified Ideographs. The proposed property data is entirely based on Version 006.01 of the Moji Jōhō Kiban database (文字情報基盤データベース), which the Unicode Consortium licensed from CITPC (*Character Information Technology Promotion Council* 文字情報技術促進協議会), the organization that currently owns the database.

The following table provides the fields for this property as reflected in UAX #38:

Field	Text
Property	`kJapanese`
Status	`Provisional`
Category	`Readings`
Introduced	`TBD`
Delimiter	`space`
Syntax	`[\x{3041}-\x{3096}\x{3099}\x{309A}\x{30A1}-\x{30FA}\x{30FC}]+`
Description	The Japanese readings(s) for this character expressed in Kana. Readings expressed in Hiragana are generally considered Kun-yomi (訓読み), and readings expressed in Katakana are generally considered On-yomi (音読み). The Moji Jōhō Kiban database and its Japanese readings are owned by CITPC (Character Information Technology Promotion Council 文字情報技術促進協議会), and are used under license.

Coverage

A total of 51,582 CJK Unified Ideographs are assigned this proposed property. The following table indicates—on a per-block basis—the number of CJK Unified Ideographs that are assigned this proposed property, along with the percentage that this number represents within each block as of Unicode Version 15.0:

	URO	Compatibility	Extension A	Extension B	Extension C	Extension D	Extension E	Extension F
Characters	18,178	12	5,773	25,392	368	116	478	1,265
Percent	87%	3%	88%	59%	9%	52%	8%	17%

I propose that at first we include only readings that are provided in the Moji Jōhō Kiban database, which will serve as a very stable baseline for future property changes and additions.

Data file

The *kJapanese-data.txt* data file, which is a PDF attachment, provides everything that is necessary for adding this property to the Unihan Database, and for adding a full description of the property to UAX #38, *Unicode Han Database (Unihan)*.

Future plans for the kJapaneseKun and kJapaneseOn properties

The current idea is to eventually remove the provisional *kJapaneseKun* and *kJapaneseOn* properties from the Unihan database. By keeping those properties for now, Japanese readings enthusiasts will therefore be able to compare the readings between those properties and the *kJapanese* property where they overlap, to include the handling of 141 CJK Unified Ideographs that have *kJapaneseKun* or *kJapaneseOn* property values, but lack a *kJapanese* one.

That is all.