

人文学向け電子テキスト構築の国際ガイドラインに 日本語セマンティクス（ルビ）が導入される

1. 発表者

下田 正弘（東京大学大学院人文社会系研究科 インド哲学仏教学専門分野 教授／
次世代人文学開発センター人文情報学部門長）

2. 発表のポイント

- ◆ 2021年2月25日に公表されたTEIガイドライン(Text Encoding Initiative Guidelines)P5 version 4.2.0において、日本語のルビがガイドラインの基本構造の一つに組み込まれました。
- ◆ これによって、非欧米圏の日本語資料が有する固有のセマンティクスが、1987年から欧米の人文学研究者を中心に策定されてきた人文学のための国際標準のデータ形式として、初めて認知されました。
- ◆ この成果により、データ駆動型研究へと向かう国際的な人文学の動向において日本語の資料が適切に扱われるのみならず、多言語を横断する多様な様態の資料の研究がより精確に実現できることとなります。

3. 発表概要

東京大学大学院人文社会系研究科の下田教授が研究代表をつとめる一連の科学研究費補助金事業（JP22242002, JP15H05725, JP19H00516）の成果として、2021年2月25日に公表されたTEIガイドライン(Text Encoding Initiative Guidelines) P5 version 4.2.0において、日本語のルビがガイドラインの基本構造の一つに組み込まれました。これにより、非欧米圏の日本語資料が有する固有のセマンティクス（意味論）が、人文学のための国際標準のデータ形式として、初めて認知されるに至りました。

TEIガイドラインは、人文学のあらゆる分野で使用するテキストに内在する多様な暗黙的構造をデジタル媒体上に記述する体系的な方法として、欧米の人文学における標準規格となっています。このガイドラインは、標準化を目指す意図から、個々の文化伝統のもつ特異性からは、距離を取る志向を持っていました。これに対し下田正弘教授らは仏教学におけるデジタル研究を通じ、東アジアおよび日本語資料の独自のセマンティクスを国際標準化する活動を進めてきました。2016年、TEIガイドラインを策定するTEI協会内にEast Asian/Japanese専門分科会を設立し、日本語圏に向けてTEIの普及活動を進めるとともに、TEI協会に対して日本語ルビ導入の提案を行いました。

その結果、2021年2月25日に公表されたTEIガイドラインP5 version 4.2.0において提案が承認され、ガイドラインの基本方針の一つとなりました。日本語資料が有する固有のセマンティクス（意味論）が、非欧米圏で初めて人文学の国際標準のデータ形式として承認されたことは、国際的にデータ駆動型研究が推し進められるなか、文化伝統の多様性を前提として人文学を進めるための重要な端緒となるものです。

4. 発表内容

研究代表者・下田正弘は、一連の科学研究費補助金事業において、仏教学研究を通し、人文学全体のデジタル研究環境の構築を図り、文字、テキスト、画像に関し、Unicode、TEI ガイドライン、IIIF という、国際標準の策定活動に携わってきました。本発表は、これらのうち、テキストにおける TEI ガイドラインに関する成果になります。

1987 年以来、TEI 協会 (<https://tei-c.org/>) が策定する TEI ガイドライン (<https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>) は、人文学の研究対象となる資料に内在する多様な暗黙的構造をデジタル媒体上に記述するための体系的な方法として、現在、事実上の国際標準規格となっています。同時に、このガイドラインは、主として欧米の人文学に向けて策定されるとともに、標準化を目指す意図から、個々の文化伝統のもつ特異性からは、距離を取る志向を持っていました。

こうした状況を踏まえ、本事業は、仏教学におけるデジタル研究を通じ、東アジアおよび日本語資料のもつ独自のセマンティクスの国際標準化を進めてきました。2016 年、本事業の研究分担者 Charles Muller 教授（当時）及び、永崎研宣・客員研究員により TEI Guidelines を決定する TEI Consortium 内に、East Asian/Japanese 専門分科会が設立され、両氏を共同議長として、日本語圏に向けて TEI の普及活動が進められ、そこで得られた知見をもとに、TEI 協会に対して日本語ルビのセマンティクス導入の提案が行われました。

この活動の結果、2021 年 2 月 25 日に公表された TEI ガイドライン P5 version 4.2.0 において、提案が承認され、ガイドライン全体の基礎構造の中にルビが組み込まれました。日本語資料が有する固有のセマンティクスが、人文学の国際標準のデータ形式として、非欧米圏で初めて承認されたことは、文化伝統の多様性を前提として人文学を進めるための重要な端緒となるものです。

今後も、日本の人文学における厚い学術的蓄積を、デジタル学術基盤における国際標準に反映させ、グローバルな文化多様性の実現に取り組んでまいります。

5. 問い合わせ先

東京大学大学院人文社会系研究科・次世代人文学開発センター・人文情報学部門

教授 下田正弘（しもだ まさひろ）

03-5841-3754（インド哲学仏教学研究室）

shimoda@l.u-tokyo.ac.jp

6. 添付資料

3.4.2 Ruby Annotations

The word *ruby* (or *rubi*) refers to a particular method of glossing runs of text which is common in East Asian scripts. In horizontally-oriented text, ruby annotations typically appear above the text being glossed, while in vertical runs of text they may appear to the left or right, or both, also oriented vertically. An English example of a ruby annotation might look like this:

ti: i: ai
TEI

In Japanese, furigana (振り仮名) ruby annotations are often used to provide pronunciation guidance for readers; characters from the largely phonetic hiragana or katakana syllabaries accompany Chinese characters, like this:

だいがく にゅうがくしけん きょうつう
大学の入学試験の「共通テスト」が、
にち にち
16日と17日にありました。

Figure 3.1. The first line of a news story from NHK News Web Easy intended for Japanese learners, in which every word composed of Chinese characters has a ruby gloss.

Pinyin ruby annotations are also used in Chinese to provide pronunciation guidance, and Zhuyin (注音) phonetic symbols (commonly known as *bopomofo*) are used in Taiwan for the same purpose.

The TEI schema provides many different ways of encoding glosses and annotations, from the simple and flexible [note](#) element to a native implementation of the Web Annotation Data Model ([16.11 Annotations](#)). However, ruby is a particular, distinct, and widely-used form of annotation that appears in script, print, calligraphy, and web pages, and the TEI therefore provides specific elements for it:

- `<ruby>` (ruby container) contains a passage of base text along with its associated ruby gloss(es).
- `<rb>` (ruby base) contains the base text annotated by a ruby gloss.
- `<rt>` (ruby text) contains a ruby text, an annotation closely associated with a passage of the main text.

« 3.4.1 Terms
and Glosses
» 3.4.3
Equivalents and
Descriptions
Home

図：TEI ガイドライン P5 version 4.2.0 の全体構造を扱う章におけるルビの説明

詳細は下記の URL をご参照ください。

<https://www.tei-c.org/Vault/P5/4.2.0/doc/tei-p5-doc/en/html/C0.html#COHTGRB>