

第10章

校訂テキスト を含む 学術編集版の 構造化

永崎研宣

version 1.0

2026.3.21 作成

1. 学術編集版と校訂テキスト

テキスト資料における学術編集版 (scholarly edition) とは、あるテキストを研究目的に耐えるかたちで提示するために、**本文そのもの**だけでなく、本文が成立した根拠と編集過程を検証可能にする諸情報を組み合わせて編纂した成果物である。したがって学術編集版は、単に資料を読みやすく整えた出版物 (あるいはデジタル公開) ではなく、編纂に携わった研究者が参照した資料群、採用した方針、行った判断を明示し、第三者が追試・再評価できる状態を目指したものである。

一般に学術編集版は、少なくとも次の要素から構成される。第一に、読者が連続して読めるかたちに整えられた本文である。第二に、本文中の問題箇所について、どの資料にどの異同があるか、どの読みを採用したか、採用しなかった候補は何かを示す校異情報 (critical apparatus) である。第三に、編集方針や対象資料の性格、伝本関係、採用した規範 (正規化・表記処理・注記方針等) を説明する序論・解題に相当する部分である。必要に応じて、注釈、索引、語彙表、画像・翻刻・訳文との対応付けなどが加わることがある。対象となる研究課題や利用場面に応じて版の設計は様々である。

この意味で、校訂版 (critical edition) は学術編集版の中心的構成要素ではあるが、学術編集版そのものと同義ではない。校訂版は「本文としてどの読みを採用するか」という編集上の選択の結果であるともみることができ、それに対して学術編集版はその選択がどのような資料と判断にもとづくのかを含めて、全体として学術的に責任ある形で提示する枠組みとすることができる。TEI ガイドラインは、この枠組みをデジタル環境で実装するための記述手段であり、本文・校異・方針・参照体系を相互に結びつけて記述できる点に特長がある。

1-1. 本章の目指すところ

本章では、学術編集版を構成する諸要素のうち、とりわけ中核となる校訂テキスト (critical text) と、それを支える校異記録 (apparatus criticus) を、TEI でどのように記述するかを主に扱う。すなわち、複数の伝本 (witness) に見られる異同を比較し、編集者が本文として採用した読みを提示しつつ、採否の根拠となる情報を追跡可能な形で記述する方法を示す。伝本関係の推定や本文批判の理論史そのものは、本書では扱わない。必要に応じて最低限に触れるに留め、TEI による記述の設計と実装に焦点を絞る。

ここまで見てきたテキストデータ構築に関わる同種の作業を改めて区別しつつ、本章で目指すところを明らかにしておこう。まず、**翻刻**は、特定の伝本を可能な限り忠実に写し取る行為であり、原資料の表記や誤記を含めて記録してそれ自体を共有することに重心がある。**正規化**は、検索性や可読性のために表記を整える処理であり、同一語の表記揺れを統一するなど、利用目的に即した変換を行なうものである。そうしたものと異なり、**校訂**は、複数の伝本間の異同を踏まえ

て本文の読みを選択・補正する編集上の判断であり、その判断は、異同の提示と根拠の記録を伴って初めて学術的に検証可能となる。校訂テキストとは、この判断を反映した本文であり、校異記録とは、本文と競合する読み（variant readings, 異文、異読とも）や、判断に関わる根拠情報を体系的に記述する手法である。

以上を踏まえると、本章の目標は次の三点に整理できる。第一に、TEIにおいて本文中に校異情報を埋め込む基本的な枠組みとして、<app> (apparatus)、<lem> (採用した読み)、<rdg> (異文・異読) を用いて、本文と異同の対応を記述することである。第二に、各読みから @wit によって参照される伝本 (witness) を、teiHeader 側で定義し、本文から一貫して参照できる参照体系を構築することである。第三に、編集者による校案 (emendation) を含む校訂判断について、責任 (誰の判断か) と根拠 (どの資料に基づくか) を明示し、後の再評価や別版の作成に対応できる形で情報を保持することである。

2. 本文への校異情報組み込みの基本

TEI における校異情報の記述には以下の3種類の手法が提案されている。

- (1) Location-referenced method (位置参照手法)
- (2) Double-end-point-attached method (両端点参照方式)
- (3) Parallel segmentation method (並列分節方式)

(1) は、本文に書き込む形のものであり、校異情報の対象となるテキスト中の語句の始点か終点のどちらかに、校訂情報を書き込む方法である。

(2) は、本文以外のところに校異情報を書き込む手法であり、校異情報の対象となる本文中の語句の始点と終点にアンカーとなるタグを置き、それに対して別置された校異情報を xml:id 参照でリンクする。他の二つの手法が、校異情報を本文に書き込む inline 形式と呼ばれるのに対して、この手法は本文の外側に校異情報を置くことから stand-off 形式と呼ばれることがある。

(3) は、本文に書き込む形のものだが、校異情報の対象となるテキスト中の語句全体に対して、校訂情報を書き込む方法である。

このなかでは (3) の手法が作業上のバランスがよく理解しやすいため、これを用いたごく基本的な例を以下に見てみよう。

校訂テキストの Parallel segmentation method による記述は、本文中の当該箇所には <app> を置き、その内部に本文として採用した読みを <lem> として示し、別の読みを <rdg> として列挙する、という形をとる。加えて、どの読みがどの伝本に現れるかを @wit で示す。その参照先となる伝本 <witness> は、<teiHeader> 側の <sourceDesc> に <listWit> として定義しておく。

```

5 <TEI xmlns="http://www.tei-c.org/ns/1.0">
6   <teiHeader>
7     <fileDesc>
8       <titleStmt>
9         <title>Title</title>
10      </titleStmt>
11     <publicationStmt>
12       <p>Publication Information</p>
13     </publicationStmt>
14     <sourceDesc>
15       <listWit>
16         <witness xml:id="A"><bibl>底本A (…)</bibl></witness>
17         <witness xml:id="B"><bibl>異本B (…)</bibl></witness>
18       </listWit>
19     </sourceDesc>
20   </fileDesc>
21 </teiHeader>
22 <text>
23   <body>
24     <p>…<app>
25       <lem wit="#A">天下</lem>
26       <rdg wit="#B">天上</rdg>
27     </app>…</p>
28   </body>
29 </text>
30 </TEI>

```

2-1. 校異情報記述の実践 (a) : 般若心経を例として

ここではまず、事例として、仏典の校訂テキストとして広く用いられている『大正新脩大藏經』に No.251 として収録されている、三藏法師玄奘訳『般若波羅蜜多心経（以下、般若心経）』を採り上げる。

大正新脩大藏經はインド・中国・日本に残る、主に漢文で書かれた様々な仏典を集約して校合を行なったものである。インド・中国の仏典に関しては、大部分を 13 世紀に高麗で刊行されて東京の増上寺に所蔵されている木版の大藏經を底本としており、この『般若心経』の場合は、他の伝本として 11 世紀に南宋時代の 12 世紀に刊行された思溪版大藏經、元の時代、13 世紀に刊行された普寧寺版大藏經、を校合している。この經典は短くて取り扱いやすいため、これを題材に TEI ガイドラインに準拠した校異情報の記述の基本を見ていく。

3. 伝本の記述

上記の通り、この『般若心経』の校訂テキストでは 4 つの伝本を用いている。これらは、

<teiHeader>において元になった資料の情報を記述する <sourceDesc> に伝本の一覧として記述することになる。伝本の一覧は <listWit> 要素であり、その中に <witness> として記載していく。上記のように各 <witness> には @xml:id を付与する必要がある、それを踏まえると以下ようになる。

```
<sourceDesc>
  <listWit>
    <witness xml:id="宋">増上寺宋版思溪蔵</witness>
    <witness xml:id="元">増上寺元版普寧蔵</witness>
    <witness xml:id="麗">増上寺高麗版</witness>
    <witness xml:id="明">酉蓮社明版嘉興蔵</witness>
  </listWit>
</sourceDesc>
```

ここで用いられている @xml:id の値は、XML ファイルの他の箇所からは「#」と「xml:id の値」を組み合わせることで、該当する @xml:id の値を持つ要素を参照することができる。たとえば、本文中の読みとして「増上寺高麗版」であることを示す際には「#麗」という値を用いることになる。

<witness> の中には書誌情報も記述可能である。一般的な図書資料であれば <bibl> や <biblFull>、<biblStruct> 等を利用することになり、写本などの貴重資料の場合には <msDesc> を用いて詳細な書誌情報を記述してもよいだろう。もちろん、TEI ガイドラインの他の多くの記述方法と同様に、これらの詳細記述もまた、必須ではない。データを作成する個人、あるいはプロジェクトがプロジェクトの目的に即して一貫して同じ詳細さで記述していくことが重要である。

この <witness> の情報を踏まえ、本文における記述は以下ようになる。

```
<app>
  <lem>帝</lem>
  <rdg wit="#宋 #元 #明 #東嘉">諦</rdg>
</app>
```

ここでは、本文は「帝」であり、他の伝本では「諦」と記述されていることを示している。すなわち、<lem> は本文にあたるテキストをあらわし、それに対する異文は、<rdg> を用いて記述される。そして、<rdg> の典拠となる伝本の情報は、上述の <witness> エレメントにつけられた @xml:id の値を、<rdg> に付与した @wit 属性の値として「#」を付けて参照することにな

る。複数の伝本を一つの @wit から指し示そうとする場合は、半角空白で区切る。

3-1. witness 設計の注意点：@xml:id、@wit、書誌の粒度

校訂テキストを TEI で記述する際、<app>/<lem>/<rdg> の運用そのもの以上に、後の利用可能性を左右するのは witness (伝本) の記述、とりわけその参照の設計である。本文中の @wit は、単に「A 本・B 本」といったラベルを付すための属性ではなく、校異記録全体を支える参照体系 (reference system) の一部として機能する。とりわけ、伝本の数が多い場合には、@xml:id の付与規則、@wit の記述方法、<witness> 以下に記述される書誌記述の粒度は、当初からよく意識して設計することが望ましい。

3-2. @xml:id の付け方：安定性と可読性

伝本の情報は、通常、ヘッダ内の <sourceDesc> 以下に記述される <listWit> において <witness> 要素として定義し、各 witness に @xml:id を付す。ここで重要なのは、@xml:id が版をまたいで安定的に参照される識別子である点である。編集の進行とともに伝本が追加されることはあっても、既存の @xml:id の値 を後から変更すると、本文中の参照や外部ツールの処理を壊してしまうことになる。

したがって、@xml:id の設計指針としては、少なくとも次の要件を満たすことが望ましい。

- ・ **一意性**：同一文書内で重複しないこと (XML の基本要件)
- ・ **安定性**：後から変更しないで済むようにすること (参照の崩壊を避ける)
- ・ **可読性**：編集者や利用者が、目で見ても判別できる程度の規則性を持つこと
- ・ **拡張性**：伝本の追加や分類の追加に耐えること

実務上は、簡便さを優先して A, B のような単純な ID を用いることも多いが、伝本数が増えたり、底本の変更や伝本群の再編が起きたりすると、意味の混乱を招きやすい。そこで、次のような中程度の規則性を推奨できる。

- ・ 伝本群 (所蔵機関やシリーズ) を反映した接頭辞を用いる (例：KYO_001, KYO_002)
- ・ 略号がある場合は、ヘッダ内で必ず説明する (「KYO」は何か、等)
- ・ 可能であれば、外部の安定識別子 (請求記号、所蔵 ID、DOI 等) と対応づける (ただし xml:id 自体に長大な値を埋め込む必要はなく、<witness> 等に記述しておく)

いずれの場合も、@xml:id は人間にとっての名称ではなく参照のための符丁であることは念頭に置いておきたい。名称や略称、番号等は、必要であれば別途 <abbr> や <idno>、あるいは

<bibl> の記述を通じて与えることが可能である。ただし、上記の般若心経の例のように、すでに分野のなかで十分に広まっていて安定的な略号の場合にはそのまま利用することも有効と考えてもよいだろう。

3-3. @wit の運用：参照の単位と記述規則

@wit は、<lem> や <rdg> がどの 伝本に基づくか（あるいはどの伝本に現れるか）を示すために用いる。@wit の値は、一般に #A #B のように @xml:id への参照（URI 参照）を半角空白の区切りで列挙する。したがって、@wit を利用する際は、次の事項についての解釈を統一しておく必要がある。

(1) 列挙の意味

- ・ <lem wit="#A #C"> は、「本文として採用した読みが A と C に存在する」ことを意味する。
- ・ <rdg wit="#B"> は、「別の読みが B に存在する」ことを意味する。このとき、「どの読みを採用したか」は <lem> で明示されるため、<rdg> の @wit は「その読みの分布」を表すことになる。

(2) 不確実性・推定の扱い

伝本の状態が悪く判読困難である、あるいは当該箇所が欠けている等の場合がある。そのようなケースを無理に @wit の列挙だけで吸収しようとする、意味が曖昧になる。少なくとも次の区別は、注記や分類によって明確化し、必要に応じて TEI の要素を用いて記述しておくことが望ましい。

- ・ 欠落 (lacuna)：当該箇所が伝本に存在しない

```
<app>
  <lem wit="#A">天下</lem>
  <rdg wit="#B"><gap reason="lost" unit="char" quantity="2"/></rdg>
</app>
```

- ・ 判読困難：存在するが読めない、あるいは確定できない

```
<app>
  <lem wit="#A">天下</lem>
  <rdg wit="#B"><gap reason="illegible" unit="char" quantity="2"/></rdg>
</app>
```

- ・ **推定**：編集者が補ったが、あるいは推定した

(例 1)

```
<app>
  <lem wit="#A">天下 </lem>
  <rdg wit="#B"><unclear>天下 </unclear></rdg>
</app>
```

※伝本 #B で不明瞭な文字列を補っている。

(例 2)

```
<app>
  <lem wit="#A" resp="#ed" cert="high">天下 </lem>
  <rdg wit="#B" resp="#tr" cert="low"><unclear>天下 </unclear></rdg>
  <rdg wit="#C"><gap reason="illegible" unit="char" quantity="2"/></rdg>
</app>
```

※伝本 #B で不明瞭なものを #tr の責任で (@resp) 補っているが確信度 (@cert) は低いので low としている。伝本 #C では 2 (@quantity=2) 文字 (@unit=char) あることはわかるが判読不能。

TEI の要素設計は複数の流儀がありうるが、本章の範囲では、まずは <note> を併置して明示する方針でもよい。重要なのは、@wit の列挙が「確定した分布」を表すのか、「推定を含む」のかを混同しないことである。

(3) 伝本の集合を簡潔に扱う方法

伝本が多数になると、@wit の値を長く列挙することになる。これを避けるため、伝本の集合 (witness group) を別途定義し、まとまりとして参照する設計を採ることがある。たとえば「宋版群」「高麗版群」のように、研究上意味のある群を前提にする場合である。

(4) 本文中の <app> の粒度との関係

@wit の運用は、校異情報の粒度と直結する。字単位の <app> を大量に置けば、@wit の列挙も増え、ヒューマンエラーが起きやすい。逆に、長い範囲を一つの <app> で扱うと、分布の記述は簡略化できても、比較や検索が難しくなる。したがって、@wit の運用規則は、校異情報の単位設計と併せて決めるべきである。

3-4. 書誌粒度：<witness>に何を書くか

<listWit> 内の <witness> にどの程度の書誌情報を書くか（粒度）は、プロジェクトの規模と目的によって変わる。ただし、校訂テキストの最小実装としては、次の二層に分けて考えると整理しやすい。

- ・ **必須層（最小）**：xml:id と、識別に足る短い書誌（<bibl> 一つ）
- ・ **拡張層（望ましい）**：所蔵・請求記号・刊写年・版種・デジタル画像の所在等を、構造化して記述

最小形の例はすでに示した通りである。

```
<listWit>
  <witness xml:id="A"><bibl>底本 A (…)</bibl></witness>
  <witness xml:id="B"><bibl>異本 B (…)</bibl></witness>
</listWit>
```

しかし学術編集版としての再利用性を高めるには、少なくとも次の情報があると望ましい。

- ・ **所蔵情報**（機関名、コレクション名）
- ・ **請求記号・管理番号**（将来の照合に必須）
- ・ **資料の性格**（写本／版本、版種、刊写年代の概略）
- ・ **参照先**（デジタル画像・目録・外部 ID 等）

これらをどこまで構造化するかは、教科書の別章（メタデータやヘッダ）との分担にも依存する。本章では、少なくとも「<bibl> に詰め込むだけで済ませる場合でも、後から構造化できるよう、記述項目の順序と表記規則を版内で固定しておく」ことを推奨しておくといよい。すなわち、最初から完璧な書誌を目指すのではなく、**参照可能性を担保する最小限の核**を確実に保持し、拡張は段階的に行うという設計である。

3-5. 小括：伝本記述の設計は校異情報の前提

校訂テキストの TEI 記述は、表面的には <app>/<lem>/<rdg> の運用に見える。しかし、その実体は、本文中の読みと、資料群（witness）とを、参照体系として結びつける設計にある。xml:id を安定させ、@wit の意味を統一し、書誌の粒度を最低限確保することによって、校訂に関する判断は初めて後から検証可能となる。次節では、この基盤上で、複数候補読みの記述、校案（emendation）の明示、責任表示の付与へと進む。

4. 校異情報の粒度の設計：どの単位で異同を記述するか

校異情報を TEI で記述する際、実務上まず問題となるのは、どの範囲を一つの <app> として扱うか、すなわち校異情報の粒度である。<app> は、ある本文箇所に対して複数の読みを対応づけるための基本単位であるが、その単位は自動的に与えられるものではない。文字単位で扱うこともできれば、語・句・文、あるいは一まとまりの表現全体を一つの変異箇所として扱うこともできる。したがって、校異情報の粒度は、対象資料の性格、研究目的、作業量、将来の再利用方法を見越して設計する必要がある。

最も細かい方法は、文字単位で <app> を設定するものである。たとえば、一文字だけ異なる場合には、その文字ごとに <lem> と <rdg> を対応させることで、どの文字がどのように異なるかを明示的に示すことができる。この方法は、検索や集計、異同の機械的比較に向いており、とりわけ字形や字種の違いを厳密に追跡したい場合には有効である。しかしその反面、異同箇所が多い資料では <app> の数が急増し、本文の可読性が下がり、入力・校正の負担も大きくなる。さらに、実際には複数文字がまとまって一つの語句の差異を構成しているにもかかわらず、それを文字ごとに分断してしまうと、編集上あるいは解釈上のまとまりが見えにくくなることがある。

これに対して、語あるいは句の単位で <app> を置く方法は、異同を意味のまとまりとして把握しやすい。たとえば、一つの熟語全体が別の熟語に置き換わっている場合や、ある一定の句が脱落・挿入されている場合には、その範囲全体を一つの校異情報として扱う方が自然である。この方法は、人間の読解に即しており、異同の意味を理解しやすい一方で、同一の <app> の内部に複数の差異が含まれることになるため、どこがどのように違うのかを細かく自動処理するには不利になることがある。

さらに大きな単位として、文や節、あるいは一定範囲のまとまり全体を一つの <app> で扱うことも考えられる。これは、ある伝本群が長い語句をまとめて欠く場合や、語順が大きく入れ替わっている場合、あるいは複数箇所の差異が相互に関連していて、一体として把握すべき場合に有効である。しかし、あまり大きな範囲を一つにまとめると、校異情報が「この範囲には何らかの違いがある」ということだけを示す粗い記録になり、比較可能性や検索性が著しく低下する。したがって、大きな単位でまとめる方法は、必要性が明確な場合に限り採るべきである。

このように考えると、校異情報の粒度設計には、少なくとも次の三つの観点に関わる。第一に、異同をどこまで細かく観察したいかという研究上の観点である。字形差や個々の文字の置換が重要であれば、文字単位の校異情報が適する。第二に、異同をどの程度人間にとって理解しやすく提示したいかという提示上の観点である。意味のまとまりとしての差異を示したいなら、語や句の単位が適する。第三に、入力・校正・再利用にかかる負担という運用上の観点である。校異情報を細かくしていくと精密さは増すが、作業コストと誤記のリスクも増大することには留意が必要である。

したがって、実務上は、原則として「研究上意味のある最小単位」を校異情報の単位とする、

という考え方が有用である。すなわち、機械的に常に一文字ごとに分けるのでもなく、逆に可読性だけを優先して大きくまとめすぎるともなく、校異として記録したい差異がどのレベルで成立しているかを見極め、その単位で <app> を設定するのである。たとえば、一文字の異同が本文理解に影響する場合には文字単位でよいが、複数文字が一体となって別表現を形成しているなら、そのまとまり全体を一つの校異情報として扱う方がよいだろう。

5. 東アジア・日本語テキストにおける留意事項

東アジア・日本語のテキスト資料においては、1文字でまとまりや意味を持つ漢字で構成される漢文資料や、一方で、連綿体で記述されるくずし字資料などがある。前者の場合は、上述の般若心経の例のように1文字単位で扱うことも十分に有用である。後者の場合には、1文字単位というよりは複数の文字をまとめた語や句として扱った方がよい場合がある。いずれにしても重要なのは、そのテキストデータを作成している研究者や研究グループにとって有用性の高い区切り方をしておくことである。

すなわち、校異情報の粒度設計とは、どこに <app> を置くかという技術的な問題であると同時に、何を一つの異同として見なすかという編集上・研究上の判断でもある。少なくとも一つのテキストのなかでは <app>/<lem>/<rdg> の運用を安定させる必要があり、そのためには、各例をその場その場で処理するのではなく、版全体を通じて一貫した粒度方針を定め、それを必要に応じて <encodingDesc> や <editorialDecl> 等で説明しておくことが望ましい。校異情報は、その粒度が統一されてはじめて、検索・比較・再解釈に耐える学術的基盤となる。

6. 読みのグループ化：<rdgGrp> の使い方

伝本が多くなると、校異情報に現れる読みもまた複雑になる。とりわけ、ある箇所において複数の伝本が互いに近い読みを示しつつも、その細部には揺れがある場合、各 <rdg> を単純に横並びに列挙するだけでは、異同の構造を捉えにくいことがある。そのような場合に用いることができるのが、TEI の <rdgGrp> 要素である。<rdgGrp> は、ある校異情報の記述において、複数の読みを「系統関係や類縁性があるもの」としてまとめるための要素であり、同じ属性を共有する読みや、一つのまとまりとして理解すべき異読の系列をグループ化するためにも用いることができる。

ここで重要なのは、<rdgGrp> が**伝本 (witness) の群そのもの**を定義する要素ではない、という点である。<rdgGrp> がまとめるのは、あくまで一つの <app> の中に置かれた**読み**

(reading) の群である。したがって、上述の般若心経の例で言うなら「宋版群」「高麗版群」のような伝本群を文書全体で恒常的に定義するのとは性格が異なる。<rdgGrp> は、ある特定の異同箇所において、「これらの読みは正書法上の異同として同じ群に属する」「これらの読みは実質的には同一方向の異読として扱える」といったことを、その校異情報記述の内部で示すために使われる。

TEI ガイドラインでは、たとえば語の綴字上の違いだけを示す複数の読みを、`type="orthographic"` を付した <rdgGrp> にまとめる例が示されている。つまり、各伝本の読みを別々の <rdg> として保持しつつ、それらが「綴字上の異同」という共通の性格を持つことを、グループとして表現するのである。

たとえば、概念的には次のように書ける。

```
<app>
  <lem wit="#A"> 諸法 </lem>
  <rdgGrp type="orthographic">
    <rdg wit="#B"> 諸濃 </rdg>
    <rdg wit="#C"> 諸澁 </rdg>
  </rdgGrp>
  <rdg wit="#D"> 諸法門 </rdg>
</app>
```

この例では、#B と #C の読みは <rdg> として個別に保持されているが、それらが lemma に対して意味上は同内容だが正書法上の差異に属することを、<rdgGrp type="orthographic"> によってまとめて示している。これに対し、#D の読みは別種の読みであるため、グループの外に独立した <rdg> として置かれている。このように <rdgGrp> を用いると、校異情報は単なる読みの列挙ではなく、異同の型やまとまりを明示する構造になる。

<rdgGrp> のもう一つの利点は、**属性をグループ単位で共有できること**にある。TEI では、<rdgGrp> は `att.textCritical` に属する属性を持つことができ、そこに与えた `@type`、`@cause`、`@varSeq`、`@hand`、`@resp` などの値は、内部の <rdg> や <lem> に継承されると説明されている。したがって、群に属する複数の読みが同じ性格を持つ場合には、各 <rdg> に同じ属性を繰り返し付けるより、<rdgGrp> にまとめて付与する方が簡潔で一貫した記述になる。

たとえば、複数の読みがいずれも「校案による支持読み」であるとか、「異体字レベルの差異」であるとか、「同一の筆写過程に由来する変異」であると判断される場合、それを <rdgGrp> 単位で示すことができる。必要であれば、グループの内部にさらに <rdgGrp> を入れ子にすることも可能であり、TEI の定義上、<rdgGrp> は <rdg> だけでなく入れ子の <rdgGrp> も含むことができる。

さらに、TEI では、`<rdgGrp>` は「ひとまとまりの異読」を示すためにも使えるとされている。これは、単に似た読みを並べるだけでなく、「この群は同じ方向の変異の微細な差を表している」といった構造を校異情報の中に持ち込めることを意味する。たとえば、ある lemma に対して複数の witness が概ね同じ読みを支持しているが、その表記や程度に細かな差がある場合、それらを一つの支持群としてまとめ、その中に個別の `<rdg>` を置くことで、「大きな意味では同じ側に属する読みだが、細部には揺れがある」ことを表現できる。

そのような使い方は、たとえば次のように表せる。

```
<app>
  <lem xml:id="lem1">般若波羅蜜多</lem>
  <rdgGrp type="supports">
    <rdg wit="#S1">般若波羅蜜多</rdg>
    <rdg wit="#S2">般若波囉蜜多</rdg>
  </rdgGrp>
  <rdgGrp type="opposes">
    <rdg wit="#K1">波羅蜜多</rdg>
    <rdg wit="#K2">般若蜜多</rdg>
  </rdgGrp>
</app>
```

ここでは、`type="supports"` や `type="opposes"` は TEI が定義する語彙ではなく、プロジェクト側で意味を定義して使うことになる。構造としては「lemma を支持する側の読みの群」と「それに対立する側の読みの群」を分けて示している。TEI ガイドラインの一次資料に見られる実務的議論でも、こうした使い方は、校異情報の表示構造を明確にするのに有用だと述べられている。

ただし、`<rdgGrp>` の使用には注意も必要である。まず、`<rdgGrp>` は便利だからといって無条件に導入すべきものではない。TEI では、読みのグループ化は **完全に任意** であり、必須ではないと明記している。実際のところ、単純な校異情報の記述では `<rdg>` をそのまま並べるだけで十分であり、無理にグループ化するとかえって読みにくくなることがある。

また、`<rdgGrp>` は witness の実在的な系統関係を自動的に証明するものではない。ある群にまとめたからといって、それが厳密な系統上の一群であることを意味することにはならない。あくまで、その校異情報記述において、編集者がそれらの読みを「同じ型に属する」と見なしていることを構造化して示すにとどまる。ここでグループ化することの意味は `@type` や解題、あるいは `<editorialDecl>` で説明しておくことが望ましい。

これに加えて、`<rdgGrp>` を使った場合でも、個々の伝本への対応は失ってはならない。各読みは依然として `<rdg wit="...">` によって個別の伝本に結びついているべきであり、

<rdgGrp> はその上に追加される中間層である。つまり、最小単位はあくまで伝本に支えられた読みであり、グループ化はその読みを整理するための構造と考えるのが適切である。これは、後から表示形式を変更したり、群を再編したりする際にも重要である。

実務的な指針としては、<rdgGrp> は次のような場合にとくに有効である。(1) 異体字・綴字差・表記差のような、同種の小差異が複数並ぶ場合。(2) lemma との関係で、支持側・反対側・周辺の変異のようなまとまりを示したい場合。(3) 一つの校異情報記述の中で、下位区分を持つ異読の系列を視覚的にも構造的にも整理したい場合。逆に、単純な二者択一の異同しかない箇所や、群の意味を説明しにくい場合には、無理に <rdgGrp> を使わない方がよいだろう。

つまり、<rdgGrp> は、伝本を群にするための要素ではなく、**読みの側に中間構造を与えるための要素**である。これによって校異情報記述の内部に「異同の型」や「変異のまとまり」を表現できるようになり、単なる列挙よりも解釈しやすい校異記録を構築できる。とくに伝本が多く、しかも異同の分布が単純ではない場合には、<rdgGrp> を適切に使うことで、異読情報の可読性と分析可能性を高めることができる。TEI による校異記録をもう一歩進めるうえで、この要素は重要な選択肢の一つである。

7. 編集者による訂正をどのように記述するか

校異情報において <lem> は本文として採用した読みを示すが、その読みが常に既存のいずれかの伝本にそのまま見られるとは限らない。複数の伝本を比較した結果、いずれの伝本の読みもそのままでは採用しがたく、編集者が本文批判上の判断にもとづいて補正・再構成した読みを本文として立てる場合がある。これが校案である。学術編集版において重要なのは、編集者が校案を行なうこと自体を避けることではなく、それがどこで行なわれ、どのような根拠と責任においてなされたかを明示することである。

たとえば、伝本 A には明らかな誤写と思われる文字列があり、伝本 B・C にもそれぞれ問題があるが、文脈や並行箇所、引用関係などを踏まえると、編集者としては別の形を本文として採るべきであると判断することがある。このような場合、<lem> に置かれた本文は、単なる伝本の分布の要約ではなく、編集者の判断を含んだ結果となる。そのため、<lem> を「採用本文」と説明するだけでは不十分であり、必要に応じて、それが伝本支持のある読みなのか、編集者による校案なのかを明確にしておく必要がある。

TEI では、このような判断の責任を記述するために、@resp 属性を用いることができる。@resp は、その読みや注記、判断が誰の責任によるものかを示す属性であり、通常は teiHeader 側で定義した編集者や翻刻担当者などの識別子を参照する。たとえば、resp="#tarou_yamada" は、その読みが編集者 #tarou_yamada の判断によることを示す。これにより、校異情報は単なる異文の集積ではなく、「誰がどの判断を下したのか」を確認できる構造となる。

また、その判断がどの程度確実なものかを示すために、@cert 属性を併用することもできる。たとえば、ある校案について編集者が高い確信を持つなら cert="high"、可能性の一つと考えるなら cert="low" という形で記述する。これにより、本文中の校訂判断をすべて同じ重みで扱うのではなく、確信度の差を保持したまま記録することが可能になる。とりわけ共同研究や長期継続型のプロジェクトでは、この区別が後の再検討にとって重要となる場合がある。

たとえば、伝本 A には「波羅蜜多」とあり、伝本 B には「波囉蜜多」、伝本 C は当該箇所が判読困難だが、編集者 (tarou_yamada) が他資料も考慮して「般若波羅蜜多」を補って本文としたとしよう。そのような場合、校異情報は次のように記述することになる。

```
<app>
  <lem resp="#tarou_yamada" cert="medium">般若波羅蜜多</lem>
  <rdg wit="#A">波羅蜜多</rdg>
  <rdg wit="#B">波囉蜜多</rdg>
  <rdg wit="#C"><gap reason="illegible" unit="char" quantity="6"/></rdg>
</app>
```

この例では、<lem> に置かれた本文が、いずれかの伝本にそのまま存在する読みではなく、編集者 #tarou_yamada の判断によって採られたものであることを示している。さらに必要であれば、<note> を併置して、その判断理由、たとえば「他の般若心経関連の諸本との比較による」などを補足してもよい。

```
<app>
  <lem resp="#tarou_yamada" cert="medium">般若波羅蜜多</lem>
  <rdg wit="#A">波羅蜜多</rdg>
  <rdg wit="#B">波囉蜜多</rdg>
  <rdg wit="#C"><gap reason="illegible" unit="char" quantity="6"/></rdg>
  <note resp="#tarou_yamada">他の関連資料および定型句との比較に基づいて補った。</note>
</app>
```

このように、校案の記述において重要なのは、編集者が本文に介入したという事実を明示することである。校訂テキストはしばしば「確定した本文」のように見えるが、実際にはその背後に複数の判断がある。TEI によって校異情報を構造化する意義の一つは、そうした判断の所在を記録し、後の編集者や利用者が再評価できる状態を保つ点にある。したがって、校案を扱う際には、単に <lem> と <rdg> の使い分けに留まらず、「採用本文の成立事情」と「その責任主体」を併

せて記述することが、学術編集版の透明性にとって不可欠なのである。

8. 脱文 / 衍文 / 転倒の扱い：単純な置換ではない異同の記述

校異記録の基本例としては、「本文では A と読むが、他の伝本では B と読む」という置き換え型の異同がもっとも理解しやすい。しかし実際の資料比較においては、異同はこのような単純な置き換えばかりではない。ある伝本では特定の語句が欠けていることもあれば、逆にある伝本にのみ語句が付加されていることもあり、また語順や句の配列が入れ替わっている場合もある。したがって、校異情報記述の設計においては脱文 (Omission)、衍文、(addition)、転倒 (transposition) をどのように扱うかを明確にしておく必要がある。

まず脱文は、ある伝本において本文相当箇所が欠けている場合である。ただし、その「欠け」が何を意味するかは注意して区別しなければならない。伝本の本文構成上、本来あるべき語句が脱落している場合もあれば、物理的損傷のためにその部分が現存しないだけの場合もある。前者は本文の違いとしての脱文であり、後者は資料の状態としての欠落 (lacuna) と捉えるべきである。TEI では、こうした区別を、必要に応じて <rdg> の中で <gap> を用いることで記述できる。

たとえば、本文では「照見五蘊皆空」であるが、ある伝本では「五蘊」が脱けている場合、次のように記述する。

```
<app>
  <lem wit="#A #B">五蘊</lem>
  <rdg wit="#C"/>
</app>
```

これは、伝本 C において当該箇所に対応する語が存在しないことを簡潔に示す一つの方法である。ただし、空の <rdg/> は意味が読み取りにくい場合があるため、必要に応じて <note> を添えるか、あるいは脱文であることを明示する表現を採る場合もある。

一方、当該箇所が脱落しているのではなく、資料の破損や汚損のために読めないのであれば、本文の相違ではなく資料の状態として扱うべきであり、たとえば次のようになる。

```
<app>
  <lem wit="#A #B">五蘊</lem>
  <rdg wit="#C">
    <gap reason="lost" unit="char" quantity="2"/>
  </rdg>
</app>
```

ここでは、伝本Cに対応箇所はあるはずだが、資料の状態のために読めなくなっていることを示している。このように、脱文を扱う際には、「本文として欠く」のか、「現状では確認できない」のかを混同しないことが重要である。

衍文はこれと逆に、ある伝本にのみ本文にない語句や文字列が付加されている場合である。たとえば、本文では「遠離一切顛倒夢想」であるが、ある伝本では「遠離一切諸顛倒夢想」と「諸」が加わっている場合、次のように記述できる。

```
<app>
  <lem wit="#A #B">一切顛倒夢想</lem>
  <rdg wit="#C">一切諸顛倒夢想</rdg>
</app>
```

このような場合、衍文は別の読みとして<rdg>に表現されることになる。ただし、添加された部分だけを校異情報として切り出すのか、前後を含むまとまり全体を一つの<app>にするのかは、上述の粒度設計の問題と関わる。意味上の差異を把握しやすいように意味のまとまりを損なわない範囲で校異情報を区切るのか、あるいは、文字単位で記述して異なる箇所のみを確認しやすくするのか、プロジェクトの全体的な方針を定めた上で対応すべきである。

転倒は、語や句の内容自体はほぼ同じであるが、その配列順序が異なる場合である。たとえば、本文では「色不異空 空不異色」であるのに対し、ある伝本ではその順序が入れ替わっている場合が考えられる。この種の異同は、単純な置き換えとして扱うことも不可能ではないが、内容の差よりも順序の差に着目したい場合には、そのことがわかるよう注記を加えるのが望ましい。

```
<app>
  <lem wit="#A #B">色不異空 空不異色</lem>
  <rdg wit="#C">空不異色 色不異空</rdg>
  <note>語順の転倒が見られる。</note>
</app>
```

TEIには校異情報における転倒を表現するための専用のエレメントが用意されていることはなく、基本的には<lem>と<rdg>の対応のなかで表し、必要に応じて<note>によって差異の性質を補足することになる。転倒が多発する場合には、校異情報を本文に埋め込むのではなく本文の外にstand-offと呼ばれる形で記載する方法が扱いやすいこともある。

このように、脱文・衍文・転倒はいずれも、校異情報において頻繁に現れ得る異同の型であるが、いずれも単に「別の読み」として処理するだけでは、その性格が見えにくくなることがある。

したがって、TEIで記述する際には、まず<lem>と<rdg>によって本文との対応関係を確保しつつ、必要に応じて<gap>や<note>を補い、その差異が欠落・添加・転倒のいずれに当たるのかを明示することが望ましい。さらに言えば、<note>の中でそれぞれを記述する際に、記述の仕方を統一しておくことで、後々の利活用可能性を高めておくことも考えるとよいだろう。校異記録は異同を列挙するだけでなく、異同の性質を理解可能な形で残すことによって、後の研究利用によりよく資するものとなる。

9. editorialDecl: 校異情報のルールを明示する

ここまで見てきたように、TEIによる校訂テキストの記述では、本文中に<app>、<lem>、<rdg>を配置し、必要に応じて@wit、@resp、@cert、<gap>、<note>などを用いて異同と編集判断を構造化していくことになる。しかし、これらの記述だけでは、その版がどのような原則にもとづいて校異を採録しているのかは必ずしも明らかにならない。たとえば、異体字は校異情報に含めるのか、単純な字体差は無視するのか、校案はどのような場合に行なうのか、欠字と判読不能はどのように区別するのか、といった方針が本文中の個々の<app>だけから自明になるとは限らない。そこで重要になるのが、teiHeader内におけるeditorialDeclでの記述である。

<editorialDecl>は、当該TEI文書において採用した編集方針を説明するための要素であり、校訂テキストを含む学術編集版においてはきわめて重要である。校異情報は一見すると本文中の局所的な記述の集まりに見えるが、実際には「何を異同として採るか」「編集者の判断をどこまで明示するか」といった全体方針の上に成り立っている。したがって、本文中に現れる校異情報の解釈規則は、可能な限り<editorialDecl>において版全体の原則として記述しておくべきである。

たとえば、少なくとも次のような事項は<editorialDecl>で説明しておくことが望ましい。第一に、どの種類の異同を校異情報に採録するかである。すべての字形差・異体字を採録するのか、あるいは本文理解に関わる異同のみを対象とするのかによって、校異情報の密度も意味も大きく変わる。第二に、<lem>がどのような本文を表すのか、すなわち底本としたテキストの読みを原則とするのか、複数伝本の比較に基づく校訂本文なのか、編集者による校案を含みうるのかを明記しておく必要がある。第三に、@respや@certの運用方針、すなわち誰の判断をどの単位で記録し、確信度をどのような基準で付与するかを示しておくこと、後の利用者にとって解釈しやすい。

このような説明は、異同の記述に対する理解を正確なものにするだけでなく、編集作業上どうしても避けがたい誤記が生じた場合にも、誤記であるかどうかを判断する上で役立つという効率上の理由もある。たとえば、<lem>が底本としたテキストの読みを原則とするのであれば、それとは異なる読みがあれば、誤記と判断できる。あるいは、<lem>に校案が存在しないというルー

ルが明示されていたなら、どの伝本にも存在しない読みはやはり誤記であると判断し得る。

また、脱落 (omission)・付加 (addition)・転倒 (transposition) の扱いについても、本文中の実例だけで理解させるのではなく、版全体としての方針を記述しておくといだろう。たとえば、「物理的欠損による判読不能は <gap> により記述し、本文差としての脱落とは区別する」「転倒は原則として一つの <app> の中で <lem> と <rdg> によって示し、必要に応じて <note> で補う」といった形である。こうした方針が明示されていれば、読者は一貫した編集ルールに基づくものとして個々の校異情報を理解できるようになる。

以上を踏まえると、<editorialDecl> には次のような記述を置くことができる。

```
<encodingDesc>
  <editorialDecl>
    <p> この文書では、本文理解に関わる異同を中心として校異情報に採録し、単純な字体差・字形差は原則として採録していない。</p>
    <p> 本文中の <gi>lem</gi> は校訂本文を示し、編集者による校案を含んでいる。校案の場合には、責任主体は <att>resp</att> 属性により、確信度は <att>cert</att> 属性により示す。</p>
    <p> 物理的欠損や判読不能箇所については <gi>gap</gi> により記述し、本文差としての脱落とは区別する。語順の転倒や語句の添加については、原則として一つの <gi>app</gi> 内で対応づけ、必要に応じて注記を付す。</p>
  </editorialDecl>
</encodingDesc>
```

この例は最小限のものにすぎないが、それでも、ここでの校異情報の理解に必要なルールを一定程度明示できている。本文側にどれほど精密な <app> を配置しても、その背景にある編集ルールがヘッダで説明されていなければ、版の設計思想は利用者にうまく伝わらない場合があるということには十分に留意しておきたい。

したがって、校訂テキストの TEI 記述においては、本文とヘッダを分離して考えるべきではない。本文中の <app>/<lem>/<rdg> は個々の異同を示し、<listWit> はそれを支える資料参照体系を与え、<editorialDecl> はその全体を貫く編集ルールを説明する。この三者が結びついてはじめて、校訂本文の成立事情とその検証可能性が十分に担保されるのである。学術編集版の透明性は、本文の細部だけでなく、こうした方針記述を含めた全体設計によって支えられていることを、ここでは確認しておきたい。

10. 校異情報の配置形式：inline と stand-off

TEI で校異情報を記述する手法としてここまで見てきた例では、校異情報を本文中の該当箇所に直接埋め込む形で `<app>` を配置してきた。このような方法は、inline の校異情報と呼ぶことができる。すなわち、本文の流れの中に `<app>` を置き、その内部に `<lem>` と `<rdg>` を記述することで、本文と異同との対応関係を本文の中で直接表現する方式である。

inline の方式の最大の利点は、本文と校異情報との対応が直観的に分かりやすいことである。どの箇所にどの異同があるのかが本文上で即座に確認でき、教育用の教材や小規模な学術編集版、あるいは試作段階のデータではとりわけ扱いやすい。また、TEI の基本的な校異情報の理解にも適しており、本章で扱ってきた `<app>/<lem>/<rdg>` の関係をもっとも素直に記述できる。さらに、本文と校異注を対応づけた表示画面にするために XSLT やその他のプログラミング的な処理を用いる場合でも、比較的単純な方法で作成しやすいという利点もある。

しかしその反面、inline の方式には限界もある。校異情報の数が増えると本文そのものが複雑になり、可読性が下がる。とりわけ文字単位の異同を多く記録する場合には、本文の連続性が見えにくくなり、データの保守や校正も難しくなる。また、一つの本文に対して複数種類の校異情報を重ねたい場合、たとえば語彙注・校異・異本ごとの翻刻対応・画像領域とのリンクなどを同時に持たせようとする、inline だけでは構造が過密になりやすい。さらに、同一の本文に対して異なる編集方針による校異情報を並存させたい場合にも、本文中にすべてを埋め込む方法はオーバーラップしてしまうなどして記述が困難になってしまいがちである。

これに対して、校異情報を本文から切り離し、別の箇所あるいは別の構造として管理する方法がある。これは、テキスト構造化の議論においては、stand-off 的な記述と呼ばれ、校異情報の記述に際しては Double end-point attached method という手法が提供されている。ここでは、本文それ自体は比較的単純な形で保持し、どの位置にどの異同が対応するかを、別の要素群や別ファイルに記述された異同の情報と `xml:id` 参照によって結びつける。この方法の利点は、本文の可読性を保ちやすいこと、複数の注釈層を独立に管理できること、また一つの本文に対して異なる編集ビューや複数版の校異情報を重ねて提供しやすいことである。大規模コーパスや長期運用のプロジェクト、あるいは画像・翻刻・校訂本文・注釈を多層的に連携させるデジタル学術編集版では、こうした設計が有効になることが少なくない。

一方で、stand-off 的な記述は参照機構が増える分だけ設計が複雑になり、人の目にとっての可読性が下がってしまうことがある。Oxygen XML Editor のような一般的な XML エディタで記述する場合には、記述にやや難しさがある。そして、どの位置とどの校異が結びつくのかを、要素の入れ子ではなく `xml:id` やアンカーによって管理する必要があるため、データ作成時の整合性管理を一般的な XML のバリデーションとは別に用意する必要が生じる。したがって、最初はまず inline の校異情報を基本として理解した上で、本文が大規模になったり、校異以外の注釈層との組み合わせが必要になったりした場合に、stand-off 的な方式を導入する選択肢があると考え

のが適切だろう。

実務的には、小規模な教育用サンプルや、校異箇所が比較的限定的な校訂テキストであれば、本文中に <app> を埋め込む inline 方式で十分である。一方で、多くの伝本翻刻を並行して保持し、さらにそれらと校訂本文・画像・索引・注釈を相互参照させるような構成を目指す場合には、最初から stand-off 的な管理も視野に入れて設計した方がよい。inline と stand-off では、どちらが優れているということではなく、本文の規模、注釈層の数、想定される表示形式、将来の拡張可能性に応じて適宜選択することになる。ただし、inline から stand-off への変換は自動的にこなすことが容易な場合が多いが、stand-off から inline への変換はしばしば容易ではなく、その点には留意する必要がある。

つまり、校異情報の配置形式は、単なる技術上の選択ではなく、学術編集版全体の設計方針と結びついている。本文中に直接埋め込むか、別管理とするかによって、可読性、保守性、再利用性は大きく変わる。本章では説明の明快さを優先して inline の例を中心に扱ってきたが、デジタル環境における学術編集版の発展を考えるならば、stand-off 的な構成もまた重要な選択肢であることを意識しておきたい。

11. 日本語や漢文資料における校異記録の注意点

TEI における校異情報の記述は、欧州語の資料に限らず、日本語資料や漢文資料にも適用することができる。しかし、その際には、表記体系や書記慣習の違いに由来するいくつかの注意点がある。日本語や漢文資料では、何を本文の差異として扱い、何を単なる表記差・字形差として扱うかの判断が、とりわけ重要になるからである。

まず問題となるのは、文字の扱いの章でも扱ったように、異体字・俗字・略字・新旧字体などの扱いである。たとえば、字体が異なっても本文批判上は同一の文字と見なす場合がある一方で、その字体の差異自体が資料の系統や成立事情を考える上で重要になることもある。したがって、すべての字体差を校異情報に採録するのか、それとも本文理解に関わる差異のみを対象とするのかを、最初の方針として定めておかなければならない。とりわけ漢文資料では、一字の差がそのまま典拠の違いに関わる場合もあるため、単純に「字体差だから無視してよい」とは言い切れない場合もある。また、闕画のように、文書の成立年代の特定に役立つ字形差もあるが、この情報は闕画であることを識別できる人でなければ記述できないため、最初からそれを前提として記述しようとするよりは、後からの修正を前提として考えるのがよいだろう。

日本語資料の場合には、仮名遣い、送り仮名、振り仮名、返り点、訓点、句読点など、本文そのものと周辺的な読解補助情報との境界が問題になる。たとえば、返り点や訓点を本文の一部として校異情報の対象に含めるのか、それとも別層の注記情報として扱うのかによって、記述設計は大きく異なる。漢文訓読資料や注釈付き古典籍では、本文・注記・訓点が同一面上に重なり合っ

て現れ、複数のテキストが混在するような形になる場合もあるため、何を校訂本文の対象とし、何を補助情報として別に構造化するかを明確にする必要がある。

版本や写本では、改行、闕字、抬頭、空格、欄外注、行間書入れなど、版面や書写面の特徴が本文理解に関わることがある。とくに古典籍の翻刻から校訂本文へ進む場合、翻刻段階では版面由来の情報をできるだけ忠実に記録することになるが、校訂本文ではそれらをどの程度整理・抽象化するかを決めなければならない。つまり、日本語・漢文資料の校異情報では、本文差だけでなく、翻刻段階で記録された本文以外の様々な特徴との関係を意識して設計する必要がある。

漢文資料や仏典資料では、一定の定型句や典拠の反復が多く見られるため、一見すると些細な異同が、実は他の経典や注釈書との関係を示す重要な手がかりになることがある。したがって、校異情報の粒度や校訂の判断は、単なる文字列比較だけでなく、当該分野の文献学的知識や引用関係の理解を前提とすることが多い。TEIでは、どの判断がどのような根拠を持ちどの責任主体によるものかを記録できるため、そうした枠組みを活かして、解釈に関わる情報を積極的に記述していくことも有用である。

以上を踏まえ、日本語・漢文資料の校異情報を記述する際には、まず何を本文の差異として採録するかを慎重に定め、その方針を<editorialDecl>などに明示しておくことがとくに重要である。異体字、訓点、返り点、送り仮名、追記、割注、欄外注などをどの層に属する情報として扱うのかを曖昧のままにすると、校異情報の意味も不安定になる。言い方を変えれば、こうした複雑さを持つ資料だからこそ、TEIによる構造化は、本文と周辺情報を整理しつつ、それぞれの関係を保持するための有効な手段となるのである。

12. 校訂テキスト作成のワークフロー

ここまで述べてきた校異情報の記述方法や伝本記述の設計は、実際の作業の流れの中で理解するとより把握しやすい。そこで最後に、TEIによる校訂テキスト作成のおおまかなワークフローを整理しておこう。実際のプロジェクトでは資料の性格や規模に応じて手順の前後や細部は異なってくるが、基本的な流れはおおむね共通している。

最初の段階では、対象とする伝本（版本・写本等）を確定し、どの資料を比較対象とするかを決める。これは単に手元にある資料を集めるというだけでなく、どの版を底本の候補とするか、どの資料を主要な伝本として扱うかを見通した上で、比較可能な資料群を選定する作業である。この段階で、資料の所在情報、請求記号、刊享年、版種、画像の有無などを整理しておく、後の伝本記述が安定する。

次に、各資料について翻刻を行なう。校訂本文は複数資料の比較によって成立するため、各伝本の本文が参照可能な形で存在しなければならない。もちろん、すでに信頼できる翻刻が利用できる場合にはそれを用いてもよいが、翻刻方針が異なるデータを混用する場合には注意が必要

である。異体字や改行、訓点、欠損、追記などをどの程度まで翻刻段階で保持するかは、後の校訂方針と整合するように決めておく必要がある。

その上で、伝本一覧を <listWit> として定義し、各伝本 (<witness>) に安定した @xml:id を与える。この作業は一見すると単純だが、校異情報全体を支える参照体系を形成することになるため、実務上きわめて重要である。可能であれば、この段階で編集責任者や翻刻担当者なども teiHeader 側で識別子として定義しておく¹⁾、後に @resp を用いた責任表示がしやすい。

次の段階で、各伝本の本文を比較し、異同箇所を抽出する。このとき、何を異同として採録するか、校異情報の粒度をどこに置くか、異体字や字形差をどう扱うかといった方針が必要になる。異同の抽出は、単なる機械的比較によってある程度支援できる場合もあるが、最終的には研究目的と資料理解に基づく判断が必要である。ここで、脱落、付加、転倒、判読不能、後筆の追記など、異同の性質を区別しながら整理していくことが重要となる。

本文として一つの善本に忠実に従うという選択を採らない場合には、次に、校訂本文を立てることになる。すなわち、各異同箇所について本文としてどの読みを採用するかを決め、必要に応じて校案を行なう。この作業では、複数伝本の比較、文脈、定型句、並行資料、研究史などを踏まえて判断することが求められる。採用本文と異文の関係は <app>/<lem>/<rdg> によって記述し、必要に応じて <rdgGrp> でまとめたり、@resp、@cert、<note> を用いて責任と確信度を明示する。

さらに、校異情報の運用や翻刻、校案の方針、字体差の扱い、欠損の表現方法などを <editorialDecl> を中心として teiHeader に記述する。ここで方針を明示しておくことにより、個々の <app> の意味づけが版全体として統一される。本文だけを精密に記述しても、方針がヘッダに示されていないければ、利用者はその記述を十分に解釈できないことには改めて留意されたい。

最後に、必要に応じて本文・翻刻・画像との対応付けを行ない、表示や検索のための変換・整形を行なう。具体的な画像との対応付けの方法については第9章を参照されたい。デジタル学術編集版では、校訂本文だけを提示するよりも、各伝本の翻刻や画像を並行して参照できる方が望ましいことが多い。そのため、実務のワークフローは、単なる XML データの作成で終わるのではなく、閲覧・検証・再利用までを含めた全体的な設計の中で考える必要がある。

このように、校訂テキストの TEI による構造化は、一つひとつの <app> を記述する作業の集積であると同時に、資料選定、翻刻、比較、判断、責任表示、方針記述を連結する一連の編集作業でもある。したがって、実務上は、個々のタグの用法を覚えるだけでなく、学術編集版全体をどのような手順で構築するのかを見通しておくことが重要である。

13. 校訂本文・翻刻・画像の関係

デジタル環境における学術編集版では、校訂本文だけを単独で提示するよりも、その背後にある翻刻や画像との関係を併せて示すことが望ましい。校訂本文は、複数の伝本を比較し、編集者の判断を経て成立した本文である。したがって、その妥当性を検証可能にするためには、各伝本の翻刻、さらに可能であれば元画像へ遡れる構造を持つことが重要である。

この関係を模式的に表すと、まず最下層に原資料としての画像がある。写本や版本のデジタル画像や IIF マニフェストがこれに当たる。その上に、各伝本ごとの翻刻が置かれる。翻刻は、特定の資料をできるだけ忠実に文字列化し、版面上の情報や欠損、追記、注記などを必要に応じて保持する層である。そして、それら複数の翻刻を比較した上で、編集者が本文として提示するのが校訂本文である。校異情報は、この校訂本文と各伝本の読みとの関係を示す媒介層として機能する。

言い換えれば、校訂本文は直接に画像から生まれるのではなく、各伝本の翻刻と読解を経由し、それらの比較と判断の結果として成立する。したがって、デジタル学術編集版においては、少なくとも概念上、次のような関係があると考えられる。

画像 (Image / Facsimile)

↓

各伝本の翻刻 (Diplomatic Transcription of Witness A, B, C ...)

↓

異同の比較 (Collation)

↓

校異情報の記述 (Apparatus Criticus) — 解釈の記述を含む

↓

校訂本文 (Critical Text)

ただし実際には、これらは単純な一方向の流れではない。校訂本文を検討する過程で再び翻刻や画像を見直すこともあれば、翻刻段階での判読の修正が校異情報や本文の判断に影響することもある。その意味で、画像・翻刻・校訂本文は階層関係を持ちながらも、相互に往還可能になっている必要がある。TEI は、これらの各層の関係を記述するための枠組みを提供しているのである。

学術編集版における校訂本文は、それ単独で完結するものではない。画像、翻刻、校異記述という複数の層に支えられて初めて、検証可能な本文として成立する。TEI による構造化の利点は、まさにこの層間の関係を明示的に記述し、利用者が必要に応じて本文から翻刻へ、翻刻から画像へ、また画像から本文の判断へと辿ることのできる編集版を構築できる点にある。

注

- 1 この場合、titleStmt において respStmt として編集者の名前を記載し、xml:id を付与しておくのが比較的簡素である。