

XMLによる仏教重要語彙定義集の作成と課題

高橋 晃 一^{†1}

インドの古典語サンスクリット語で著されていた仏教文献は、アジア各地への伝播に伴い、漢文などに翻訳されていった。また、今日では欧米でも仏教への関心が高まり、英語などへ翻訳されつつある。一方、主に漢文を通じて仏教を受容してきた日本では、仏教用語を日本語に翻訳することに対して積極的ではなかった。しかし、漢語の仏教語は、現代日本語として必ずしも意味が明瞭なものではない。そのため、サンスクリット原典に遡って、語義を再考し、現代語として有意義な訳語を提案する試みが現在進行している。この作業のために、サンスクリット原典とそれに対応する複数の言語による翻訳を、XMLを利用し、TEIの提唱する規格に従って整理する際の基本的な構造と課題について報告する。

Making a Glossary of Buddhist Technical Terms using XML: A Test Case

KOICHI TAKAHASHI^{†1}

In this short paper, I will report an experimental attempt to develop a glossary of Buddhist technical terms using XML structure according to the TEI Guidelines¹). The Buddhist canonical texts were originally written in Sanskrit, and then translated in later periods into Chinese and Tibetan. Nowadays, they are being translated into English and other modern European languages. This experimental glossary will contain these classical and modern translation equivalents, as well as the original definitions collected from the Sanskrit texts. With this as a basis, we expect that the glossary will evolve into a complex multilingual dictionary. Although the TEI Guidelines do not explicitly cover this kind of application, this report will show that these complex contents can be organized by means of the elements recommended by TEI Guidelines for making a dictionary, if some elements can be combined.

1. 仏教文献研究の概況

仏教を文献学的に研究する際、インドの古典言語であるサンスクリット語およびパーリ語が、第一次資料として用いられる。特にサンスクリット語の仏教文献は紀元前後ごろから中国へ伝えられ始め、その後何世紀にも渡って漢文に翻訳されており、また8世紀ごろにはチベットにも伝わり、チベット語に翻訳された。漢文として伝わる文献はいわゆる「大蔵経」であり、一方チベット語に翻訳されたものは「西藏大蔵経」と総称され、多くの仏典を現在まで伝えている。今日では大正新修大蔵経テキストデータベース (SAT)^{5)*1} や中華電子佛典協會 (CBETA)^{4)*2} によって漢訳「大蔵経」が電子データ化され、その膨大な内容を容易に概観できるようになったほか、チベット語訳「西藏大蔵経」も近年 Asian Classics Input Project (ACIP)^{3)*3} などにより電子データ化が進められている。

ところで、漢字文化圏に生きる日本人にとって、サンスクリット語やパーリ語、チベット語で著された文献と比べて、漢訳文献はある意味で非常に親しみ易いことは想像に難くないであろう。しかし、漢訳文献には、一般の人々にとってのみならず、仏教文献研究の専門家にとっても非常に大きな利点がある。例えば、サンスクリット語文献を日本語に翻訳する際、難解な専門用語を、敢えて現代日本語に訳さずに、古典的な漢訳語で置き換えて一応の翻訳とすることができる。これはある種の専門用語を翻訳しようとする、冗長になってしまう場合が少なからずあるため、簡潔に表現できる漢訳を便宜上利用していることによる。また、固有の創的な現代語訳を用いるよりも、符牒的に古典漢訳を利用した方が、誤解を招かずに済むことも、こうした方法が伝統的に認められてきた一因であろう。

しかし、古典的漢訳に依存しすぎることに弊害もある。漢字文化圏に生活している者にとって、漢字は特定の意味と密接に結びついているため、特殊な意味を持つ漢訳仏教語に対して、まったく異なる内容を思い起こさせてしまう場合もある。専門家同士でこのような誤解が生じることはまずないが、一般向けには配慮が必要となる。そのため、幅広い層に向けて、積極的な情報発信が求められている今日においては、漢訳語は逆に扱いが難しい点もある。

^{†1} 東京大学人文社会系研究科

Graduate School of Humanities and Sociology, Tokyo University

*1 <http://21dzk.l.u-tokyo.ac.jp/SAT/> (April 20th 2010)

*2 <http://www.cbeta.org/> (April 20th 2010)

*3 <http://www.asianclassics.org/> (April 20th 2010)

このように日本における仏教文献研究は古典漢訳語の恩恵に与りながらも、その弊害を引かずっているという状況にある。これに対して、欧米の研究者にとっては、漢訳語は完全な外国語であり、それ自体が直感的に意味を明示することはない。当然のことながら、彼らが仏教用語を自国語に翻訳する際にも、言語体系の全く異なる漢訳語をそのまま翻訳語として使用することはありえない。そのため、欧米の研究者は、サンスクリット語やパーリ語、またはその翻訳である漢文やチベット語の文献資料を、言語学、文献学などの成果に基づいて分析し、その内容を英語を始めとする欧米諸語に翻訳する努力を続けてきた。こうした努力は漢訳語を符牒的に利用してきた日本の研究者以上に、欧米の仏教研究者が仏教用語に造詣を深めていく可能性を包含している。

こうした現状を鑑み、仏教の重要語に関して、有意味な現代語訳を考案することが求められている。

2. 分析対象とするテキスト

さて、仏教要語の現代語訳を考えるにあたり、まず術語の定義的な用例をサンスクリット語原典から採集し、それらを XML を用いてデータベース化したうえで、語彙集を作成することを現在試みている。その際、TEI¹⁾の規格に従ってデータ整理を行っている*1。なお、今回は語彙と定義文を採集するに当たり、試験的に『アビダルマ・コーシャ』⁶⁾というテキストを使用している。

整理法に関して述べる前に、この『アビダルマ・コーシャ』をテキストとして使用する理由について簡単に触れておく。そもそも仏教思想の根幹であるブッダの教説は「アーガマ」と総称される聖典群であり、その内容は哲学的な場合も多いが、必ずしも体系化されたものではない。その「アーガマ」に対して解釈を加え、様々な概念を関連付けながら教理体系としてまとめたものが「アビダルマ」である。この「アビダルマ」の教理について、数世紀に渡る歴史の中で膨大な文献が著され、それに伴ってその内容は複雑化していった。そうした状況の中で、5世紀ごろ「アビダルマ」の煩瑣な教理を集大成した著作『アビダルマ・コーシャ』が登場する。「コーシャ」とは平たく言えば「辞書」のことであり、実際この語は現代ヒンディー語ではその意味で使われている。英語のシソーラス (thesaurus) が「倉庫」を表す語でありながら、「知識の宝庫」という意味で「百科辞典」を指すようなものである。

このように、『アビダルマ・コーシャ』は教理学辞典であり、仏教の哲学的術語の多くを解説している。現代的な辞典のように見出しがついたものではないが、比較的容易に仏教特有の術語とその定義を抽出することが可能であろう。さらに、全体がサンスクリット語原典で残っており、その校訂テキスト⁶⁾が出版されている他、漢訳⁷⁾⁸⁾、チベット語訳⁹⁾が現存している。このような点で、『アビダルマ・コーシャ』は文献資料に基づいて語彙を分析するための条件が整っていると言える。

3. 語彙の選定

しかし、『アビダルマ・コーシャ』が煩瑣な教理を比較的要領よくまとめた著作とは言っても、その全体を一度に扱うのは容易ではない。そこで、今回は「五位七十五法」という体系にそって重要な術語を選定し、語彙集を作成することとする。「五位七十五法」とは、教理上重要な75の概念を五つの範疇に分類したもので、種々ある教理体系のうちでも、比較的多岐に渡る語彙を含んでいる。したがって、この75の概念を見出し語に立てることによって、実験的に語彙集を作成する上では、十分な語彙を抽出することができる。

4. XML による整理方法

4.1 辞書項目記述法の概略

こうして選定した語彙を、TEI が提唱する辞書項目の整理法にしたがって分析する*4。今回の試みでは、選定した語彙について、最低限、以下の内容を採集している。

- (1) サンスクリット語原典に見られる対象語句に対する説明文または定義文
- (2) 上記のサンスクリット語原典に対する和訳（基本的に採集者の責任において翻訳）
- (3) 上記のサンスクリット語原典に対応する漢訳（二種）およびチベット語訳（一種）
- (4) サンスクリット語原典またはその漢訳かチベット語訳に対応する既存の欧米語訳

これらの内容を TEI ガイドライン¹⁾が提唱する Dictionary の記述法に従って整理していくのだが、具体的なサンプルを示す前に、まずガイドラインの関連する項目（9.3.3、特に9.3.3.1-2）の概略をまとめておく。同ガイドラインに従えば、XML によって辞書に単語

*1 TEI ガイドラインに準拠した辞書作成の試みとしては、Charles Muller 博士の先駆的研究がある。最新の研究成果は、文献 2) として公表される予定とのことなので、そちらを参照のこと。

*4 Burnard, and Bauman (2010) P5:TEI Guidelines¹⁾, 9.Dictionary 参照。

を登録する際には、`<entry>` 要素を一つの単位し、その中の子要素として、単語の基本語形を `<form>` 要素、それに対する定義や説明などの意味情報を `<sense>` 要素として示すことになっている。ところで、一般的に辞書には国語辞典のような「単一言語辞書」と、英和辞典のような「多言語辞書」がある。また単語の意味を提示する形式も様々であって、例えば、同義語で置き換えたり、他の言語に翻訳したり、様式化された定義を与える場合もある。TEI の規格ではそれら諸形式を特に区別せず、基本的に `<def>` (definition) タグによって、ある語に対する定義を示すことになっている。TEI ガイドラインの例を見る限り、これは「単一言語辞書」で語彙を説明する場合に相当する。一方、「多言語辞書」の場合、`<cit>` (cited quotation) タグによって他の言語への翻訳語を提示することができる。

以上が TEI に準拠した辞書作成の概略だが、この規格を今回の語彙集作成の試みに当てはめた場合、次のような問題が考えられる。

今回の語彙集は、サンスクリット語の術語を見出し語とし、それに対する説明文または定義文を示すことが基本的な構造となっている。その意味では、「単一言語辞書」の構造を持っている。しかし、これに基づいて現代語訳を考案するという目的のために、「サンスクリット語で記述された説明文」に対して、古典語と現代語による複数の翻訳文を付すことになる。それは文字通りの「多言語辞書」となる。したがって、最終的には「単一言語辞書」を複数の言語によって翻訳した「多言語辞書」が出来上がることになる。すなわち、XML によって記述する場合、`<sense>` 要素の下に `<def>` 要素と `<cit>` 要素を並存させ、それぞれに言語情報を属性として付加することになる。以下、サンスクリット語で「信仰心」を意味する術語 *śraddhā* を例にサンプルを示す。

4.2 単一言語辞書型の構造

まず、見出し語となるサンスクリット語の術語と、それに対するサンスクリット語の定義文を上記の規格に従って記述すると次のようになる。

```
<entry>
  <form >
    <orth xml:lang="sa">śraddhā</orth>*1
  </form>
```

*1 `<orth>` タグは orthographic form すなわち見出し語の正書法に適った綴りを示す。また、言語の種類は属性 `@xml:lang` により、ISO 639-1 に従って示した。言語情報については P5:TEI ガイドライン¹⁾ vi.Languages and Character Sets 参照。

```
<def>tatra śraddhā cetasaḥ prasādaḥ</def>
</entry>
```

これが今回の語彙集を作成する上で、最も基本となる部分であり、見かけ上はサンスクリット語の「単一言語辞書」の様相を呈することになる。なお、TEI ガイドライン¹⁾ に準拠すれば、`<def>` 要素は `<form>` 要素などと異なり、単独でより下位の要素を支配することはない^{*2}。通常の辞書であれば、定義文には出典などは必要ないので、これで充分であろう。しかし、今回の語彙集では、定義は『アビダルマ・コーシャ』から採集したもので、すなわち引用したもののなので、出典を明示できないのは都合が悪い。出典を示すためには `<def>` 要素を `<cit>` 要素の子要素とする方法が考えられる。`<cit>` 要素は書誌情報と共に他の文書からの引用を示すために用いられ、辞書では例文や翻訳、用例を示す^{*3}。したがって、`<cit>` を用いて次のように記述すれば、定義文の出典を示すことができる。

```
<entry>
  <form >
    <orth xml:lang="sa">śraddhā</orth>
  </form>
  <sense>
    <cit>
      <def>tatra śraddhā cetasaḥ prasādaḥ</def>
      <bibl>*4
        <title>Abhidharmakośabhāṣya of Vasubandhu</title>
        <imprint>
          <date>1967</date>
          <biblScope type="pages">55</biblScope>
          <biblScope type="lines">6</biblScope>
```

*2 Burnard, and Bauman (2010) P5: TEI Guidelines¹⁾9.3.3.1 参照。

*3 Burnard, and Bauman (2010) P5: TEI Guidelines¹⁾ 9.3.3.2 Translation Equivalents の `<cit>` の解説を参照。

*4 `<bibl>` タグは比較的大雑把に書誌情報を提示することができる。文献 1) 3.11 Bibliographic Citations and References 参照。ここでは「書名」「出版年」「頁」「行」を含めた。

```
</imprint>
</bibl>
</cit>
</sense>
</entry>
```

なお、TEI ガイドライン自身はこのような場合を想定しているわけではないが、<cit> の配下に <def> を置く例を紹介している*1。

4.3 多言語辞書型との統合

これまでの作業でサンスクリット語の術語をサンスクリット語で説明するタイプ、すなわち「単一言語辞書型」の語彙集のモデルができたが、次に、これに対する古典訳および既存の現代語訳の情報を追加し、新たな現代日本語訳を提案するための資料として整備していかねばならない。その際、定義文に対する翻訳の情報は、やはり <cit> タグを用いて示すことになる。すでに見たように、<cit> タグは出典を伴って例文や翻訳を提示するためのものなので、問題なく使用することができるだろう。

```
<entry>
  <form >
    <orth xml:lang="sa">śraddhā</orth>
  </form>
  <sense>
    <usg type="dom">abhidharma</usg> *2
    <cit type="example" xml:lang="sa">
      <def>tatra śraddhā cetasaḥ prasādaḥ</def>
      <bibl>
        (書誌情報省略)
      </bibl>
    </cit>
```

*1 Burnard, and Bauman (2010) P5: TEI Guidelines, 9.3.3.2 Translation Equivalents p.273 参照。
*2 <usg> タグは辞書項目の用例情報を示す。@type 属性の値 dom (domain) は「分野・領域」を示しており、ここでは「仏教教理学 (アビダルマ) の領域に関わる用例」の意味で abhidharma を要素として指定している。

```
<cit type="translation" xml:lang="ja">
  <quote>このうち<term type="translation">信仰心</term>は心を清らかに
  することである.</quote>
</cit>
<cit type="translation" xml:lang="zh">
  <quote>此中<term type="translation">信</term>者. 令心澄淨.
  <bibl>玄奘訳,19b2--4</bibl></quote>
</cit>
<cit type="translation" xml:lang="bo">
  <quote>de la <term type="translation">dad pa</term> ni sems dang
  pa'o// <bibl>Peking edition, 72b4--</bibl></quote>
</cit>
<cit type="translation" xml:lang="en">
  <quote><term xml:lang="sa">Śraddhā</term> or
  <term type="translation">faith</term> is clarification of the mind.
  <bibl>tr. by Leo M.Pruden, 1988, 191</bibl></quote>
</cit>
</sense>
</entry>
```

単一言語辞書型の構造を説明した際と同様に、翻訳文に関する書誌情報もタグによって整理の方が望ましいが、煩雑になるので今は詳細は省く。ここで注意すべき点は、<def> の親要素としての <cit> と、その兄弟要素であり、定義文に対する翻訳を提示するための <cit> を区別する必要があることだろう。現時点では、@type により、前者を「定義的 用例」という意味で example、後者を「それに対する翻訳」という意味で translation とし て識別している。また、それぞれに@xml:lang によって言語情報を与えてある。なお、見出し語であるśraddhā—=a に対する各翻訳文中の訳語を <term> タグを付けて識別してあるが、これが適切な方法か否かは今後検討を要する。

以上のようにして、一つの意味を示すためのサンスクリット語原文とそれに対する複数の翻訳を整理することができる。なお、現時点ではまだ試みていないが、今後、異なる種類の文献から定義文を採集する場合を想定し、意味情報の使用範囲を示す <usg> (usage) 要素

を意味情報に付加しておくのが望ましいと思われる。

5. 今後の課題

今回分析の対象としている『アビダルマ・コーシャ』は、古典的ではあるがそもそも辞書として体裁を持っている。そのため、TEIのガイドラインが提唱するDictionaryの構造に沿って比較的容易に整理できる。しかし、このような例は特殊なものであり、すべてのサンスクリット語およびパーリ語の仏教文献に適用することができない可能性もある。今後、語彙集の用例を拡充していくに当たり、物語の体裁を持つ文献や戒律を扱う文献も視野に入れていくことを予定しているが、そのようなタイプの異なる多様な文献を、XMLによってどのように分析していくか、検討する必要がある。

謝辞 本研究は科学研究費基盤(A)「仏教用語の『日英基準訳語集』構築に向けての総合的研究」(課題番号:19202002,代表:斉藤明 東京大学大学院人文社会系研究科・教授)の助成を受けたものである。なお、Charles Muller 博士(東京大学)には、XMLによる辞書作成に関して助言をいただいたほか、英文要旨も校正していただいた。また永崎研宣氏(人文情報学研究所所長)と松田訓典氏(東京大学東洋文化研究所助教)より、技術的な助言をいただいた。謹んで感謝の意を表したい。

参 考 文 献

- 1) Burnard, L. and Bauman, S.eds.: *TEI P5: Guidelines for Electronic Text Encoding and Interchange P5*, Version 1.6.0 February 12th 2010. TEI Consortium. <http://www.tei-c.org/Guidelines/P5/>(April 20th 2010), (2010).
- 2) Muller, C., Nagasaki, K. and Soulat, J.: The XML-Based DDB: The DDB Document Structure and the P5 Dictionary Module; New Developments of DDB Interoperation and Access, *Chán Buddhism- Dūnhuáng and Beyond: Texts, Manuscripts, and Contexts*. Anderl, C. and Wittern, C.(eds), Forthcoming in *Abhandlungen fuer die Kunde des Morgenlandes*. Wiesbaden, Harrassowitz (2011).
- 3) Asian Classic Input Project (ACIP), <http://www.asianclassics.org/> (April 20th 2010)
- 4) 中華電子佛典協會 Chinese Buddhist Electronic Text Association (CBETA), <http://www.cbeta.org/> (April 20th 2010)
- 5) 大正新脩大藏經テキストデータベース Saṃgānikīkṛtaṃ Taiśotripiṭakam The SAT Daizōkyō Text Database (SAT), <http://21dzk.l.u-tokyo.ac.jp/SAT/> (April 20th 2010)
- 6) Prahan, P: *Abhidharmakośabhāṣya of Vasubandhu*, Tibetan Sanskrit Works Series,

- vol VIII, K.P.Jayaswal Research Institute, Patna, (1967).
- 7) 玄奘(訳): 阿毘達磨俱舍論, 大正新修大藏經 no.1558, (1926).
- 8) 真諦(訳): 阿毘達磨俱舍積論, 大正新修大藏經 no.1559, (1926).
- 9) *Chos mngon pa'i mdzod kyi bshad pa*, 影印北京版西藏大藏經 115 卷 no.5591, 西藏大藏經研究会, (1957).
- 10) Pruden, L.M.(tr.):*Abhidharmakośabhāṣyam by Louis de La Vallée Poussin*, vol 1. (1988).