

東洋古典電子テキストの利便性向上にむけて

— 「Tibetan-Sanskrit 構文対照電子辞書 eDic」 閲覧システムの構築を通じて —

相場 徹[†]

生出 恭治[‡]

[†] 東北大学 高等教育開発推進センター

[‡] 東北大学

概要: 古典サンスクリット語 (Skt.) や古典チベット語 (Tib.) の電子テキストは、その利便性が十分に活用されているとは言い難い。本稿では、“eDic” のテキスト検索・閲覧システムの試作を通じて行った Skt. や Tib. の電子テキストの利便性向上への取り組みについて述べる。

Tib. の電子テキストも、僅かの電子情報を追加するだけで利便性向上を図ることができる。また、さまざまな電子情報をリンクさせることにより、さらに新たな電子情報の構築が可能となり、それがさらに電子テキストの利便性向上に結び付くことが期待される。

Towards the Improvement of the Usability of E-Texts in Classical Tibetan and Sanskrit: Through the Process of Constructing “eDic” Browsing System

Tooru AIBA[†]

Kyoji OIDE[‡]

[†] Center for the Advancement of Higher Education,
Tohoku University

[‡] Tohoku University

Abstract: Currently we have many more e-texts in classical Tibetan or Sanskrit, but we still cannot easily analyze them by using software tools available now. In this paper we will try to seek improvements of the usability of such e-texts through the process of implementation of “eDic” browsing system.

We realize that even a little amount of knowledge about the target language can be helpful. We also claim that various kinds of information improve the usability of Sanskrit and Tibetan e-texts. Although our attempt to automatically extract such information does not attain the practical results, it will be the basis of more intellectual processes.

1 はじめに

古典サンスクリット語 (Skt.)、古典チベット語 (Tib.) などのテキストは、電子テキスト化、すなわちデジタルアーカイブ化がかなり進んでいる。しかし、電子テキストの利便性が十分に活用されているとは言い難い。その理由として、Skt. および Tib. の複雑さ、また特殊性などが挙げられよう。また一

方では、Skt. および Tib. などに関するさまざまな電子データをリンクさせて電子テキストの利便性を向上させようとする事例が少ないため、電子テキストにどのような情報をリンクさせて使うとより便利になるかに関する具体的なイメージを、我々も含めた利用者が持ちにくいこともその一因ではないかと考える。

本稿では、我々が参加している「Tibetan-Sanskrit

構文対照電子辞書「eDic」(eDic)プロジェクト [8] のテキスト検索・閲覧システムについて述べる。プロジェクト開始当初は、単純に Skt./Tib. の対訳電子テキストのみを用いるシステムを想定していたが、システムを実際に構築する作業の最中に、またプロジェクトのメンバーの人たちにシステムを試用してもらったことにより、電子テキストそのものが持つ情報だけでは対処できない問題点や、「もっと使いやすい」ものへと向けた提案などが出るようになってきた。これは、実際に構築されたシステムの利用を通じて、電子データ利用に関する具体的なイメージを利用者が持つようになった結果といえるのではないか。

2 eDic プロジェクト 概要

インド仏教文献において原典とは Skt. 文献を指すが、その多くは現存しない。それゆえ、古典インド仏教研究は、Tib. や古典中国語の翻訳資料を用いた異訳対照比較を通じ、失われた Skt. 原典の記述を推測しながらの研究となる。

また、翻訳文献の中ではとくに Tib. 訳の文献が Skt. 原典の逐語訳に近く、現存しない Skt. 原典の内容を推測する際に最も重要な手がかりとされる。このような事情から、Skt. と Tib. の対訳辞書(蔵梵辞典)も何種類か編纂されている。しかし、既存の辞書等はほとんど単語の対応関係を示すレベルに留まっており、用例が提示されることはほとんどない。だが、文献研究においてはそれでは不十分なことが多い。eDic は、このような既存の対訳辞書に欠けている「用例」を提示するシステムの構築プロジェクトと位置付けられる。

2.1 閲覧システムに求められるもの

eDic の概念図を図 1 に示す。図 1 の Text A は、Skt. 原典が散逸してしまい Tib. 訳文献しか残っていない文献を示している。Text B, C は、Skt. 原典および Tib. 訳文献の両方が現存する文献である。ここで一般に、Text A を研究対象とする必要が生じたときは、現存する Text A の Tib. 訳文献から散逸した Text A の Skt. 原典の姿をイメージして研究を行う必要がある。しかし、こ

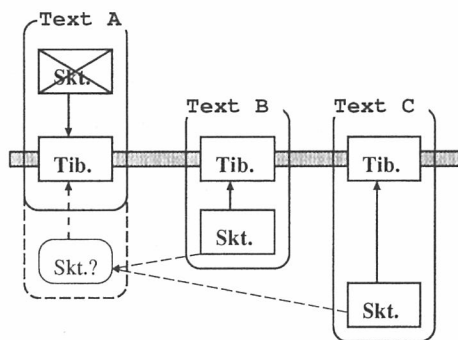


図 1: eDic の概念図

Text A の Tib. 訳文献と類似した Tib. の表現を持つ Text B, C を見つけることができた場合、これら Text B, C の Skt. 原典を参照することによって、散逸してしまった Text A の Skt. 原典の姿をイメージしやすくなる。

このように「Tib. 訳文献のみが現存するテキストの、Skt. 原典の姿をイメージする手助けをする」ことが、現状の eDic プロジェクトの主目的である。それゆえ閲覧システムの基本的な枠組は以下となる。

- 利用者が、Text A 中の Tib. 単語列を「検索文字列」として入力
- 入力された検索文字列を含む Tib. 訳の構文と、それと対応する Skt. 構文とのテキスト対 (図 1 の Text B, C の内容に相当) を表示

将来的には、単に Text B, C における相当部分を羅列して表示するだけではなく、『検索語』として入力されたチベット語表現に最もふさわしいであろう Skt. 表現を推定して提示することも想定されているが、それはかなり将来的な課題であり、現段階での課題とはなっていない。

2.2 データ構造および構築

前節で述べた枠組を実現するためには、Skt. 原典とその Tib. 語訳テキストの両方を構文(文・段落)単位で区切り、対応付けを行った電子テキストが必要と我々は考えた。また現状では、この対応付けの作業は、手作業で行わざるを得ない。その理由を以下に示す。

- 文献研究の基礎資料となることが期待されているため、対応付けされた結果は正確であることが要求されること
- Skt. 原典と Tib. 訳文を構文レベルで切り分け、かつ対応付けを自動的に行うことは現状では技術的に不可能

そこで我々は、eDic プロジェクトのメンバーで分担して図 2 で示すような Skt., Tib. の構文ごとの対応データをすべて手作業で構築している¹。構築の手順は以下となる。

- Skt. 原典と、その Tib. 訳の電子テキストとを用意
- Skt. 原典を「構文」で区切る
- 区切られた Skt. の各構文に対応するよう Tib. 文を区切り、Skt. と対にする

3 閲覧システムの試作

我々は、eDic データ構築作業を行うと同時に、テキスト検索・閲覧システムの試作にも着手した。

最初は、電子テキストの利用法としてよく用いられる検索方法、すなわち「検索語」として指定された文字列を含む行を単純に羅列して提示するシステムを試作し、WWW を通じて公開した。本節では、公開の結果得られたシステム改良に向けてのヒントや要望などを契機として、eDic 閲覧システムの利便性を向上させるために行った、電子テキストそのもの以外の情報の追加および閲覧システムの拡張について述べる。

SP (Saddharmapuṇḍarīka):	
Sanskrit	(Wogihara: 22.2.15–22.2.16)
sa ca sarvasattvapriyadarśano bodhisattvo mahāsattvas tasya bhagavataḥ pravacane duṣkara-caryā abhiyukto abhūt	
Tibetan	(Lhasa: 237b3–237b4)
byañ chub sems dpaḥ sems dpaḥ chen po sems can thams cad kyis mthoñ na dgaḥ ba de ḥañ/ bcom ldan ḥdas de ḥi gsuñ rab la dkaḥ ba byed pa la brtson par gyur te/	

図 2: eDic における、一区切りのデータの例

¹本稿におけるチベット語の表記方法は、Chandra Das[4] における表記方法を用いる。

3.1 多様な転写方式への対応

チベット文字の転写方式について簡単に述べる。チベット文字の転写方式については、古くから Chandra Das[4] 等で用いられる方式や、アメリカ議会図書館 (Library of Congress) の方式など、さまざまな転写方式があった。しかし、これらの多くは “ñ” “z” など、ダイアクリティカルマーク付きアルファベットを用いたものであったため、タイプライタが使われるようになって以降、それら特殊文字を別の文字で再転写するいくつかの方法が使用されるようになった。また一方では、通常のアルファベットのみで転写できる「Wylie 方式」と呼ばれる転写方式も広く用いられるようになった。このようにチベット文字の転写方法は、さまざまな記法のもの并存している²。

我々は、公開した eDic 文献検索・閲覧システムの利用者がどのような検索文字列を入力しているかを調査したことがある。当初我々は、システム内部で用いている「相場式」と呼ぶ転写方式 (Chandra Das の転写方式を再転写した表記方法) のみを入出力に利用できるようにしていた。また、システムの WWW ページにもその旨の注意書きをしていたのだが、実際の利用者で我々の注意書きに従って「相場式」を使う人は少数派であった。その理由は、おそらく、利用者の多くは、さまざまな転写方式を読み分けることには慣れていないが、書き分けるのにはあまり慣れておらず、それゆえ無意識のうちに使い慣れた転写方式で入力してしまうのではないかと推測された。

そこで、利用者が検索単語を入力する際に、利用者が用いた転写方式を推測する仕掛けを作り、同時に結果出力の際にもその推定された方式で出力するようにした。これにより、より多くの利用者が、転写方式の相違に関するストレスを感じることなくシステムを利用できるようになったのではないかと。

このように、従来は「転写方式の違い」という壁があるため、電子テキストの利用にはある種の制約があった。しかし電子テキストに転写方式に関する情報を加えると、僅かながらではあるが、電子テキスト利用の裾野を広げることができるのではないかとと思われる。

²比較的によく用いられる転写方法については、[2] を参照。

表 1: 格助辞が名詞の音節を変化させる例

種類	名詞	助辞	例
目的 格助辞	“h” で終了 母音で終了	r r	dgah + r → dgar rta + r → rtar
作具 格助辞	“h” で終了 母音で終了	s s	bkaḥ + s → bkas ña + s → ñas
属格助辞	“h” で終了	hi	dgah + hi → dgahī

3.2 Tib. 検索に必要な文法情報

稲葉 [5, p.107] によると、Tib. における自立語には名詞・代名詞・形容詞・副詞・接続詞・間投詞・動詞の 7 種類があるが、このうち活用のあるものは動詞のみ、とされている³。

Tib. は、助辞を名詞などの後に付けることによって、各単語の文法的関係を表現する。たとえば目的格助辞 “la” が自立語 “gser (黄金)” に後続して “gser la (黄金を)” となる。多くの場合、この例のように助辞は名詞とは別の独立した音節を構成する。それゆえ、Tib. テキストを対象とした検索においては、「活用がある」とされた動詞を除けば、『検索単語』として入力された文字列と、テキスト中の文字列との単純な文字列比較をするだけでそれなりの結果が得られることが期待できる。

それゆえ、我々が最初に検索の仕組みを考える際には、入力された『検索文字列』と一致する文字列を含む Tib. 構文と、その Skt. 原典との組を提示するというだけで十分だろうと考えていた。しかし実際にシステムを試作してみるとすぐに、一部の助辞の扱いに配慮する必要があることに気付いた。本節では、助辞に関する以下の 2 つの対処について述べる。

- 格助辞の一部が、先行する名詞末尾の音節に直接に添接する場合への対処
- 『検索文字列』と格助辞との関係への対処

3.2.1 格助辞が名詞に直接添接する場合

助辞の一部には、先行する自立語の最後の音節に直接に添接してしまうものが存在する。たとえば名詞 “rta (馬)” に目的格助辞 “r” が接続すると、助辞は名詞の音節に入り込んで “rtar (馬に)” となる。

³本稿における Tib. 文法の理解は、主に稲葉 [5] に従う。

表 1 に、格助辞が名詞の音節に入り込んでしまう例を示した。このような事例が起こるのは、表 1 のとおり名詞が添後字 “h” が母音で終了するとき限定される。それゆえ、検索文字列の末尾がこれらのパターン、すなわち母音で終了するか、添後字 “h” で終了している場合には、検索文字列が名詞である可能性を考慮して、以下のように対処することとした。

- 検索文字列が母音で終了する場合は、その末尾に “r” あるいは “s” が付けられた文字列も検索
- 検索文字列が “h” で終了する場合は、末尾にある “h” を欠落させたものに “r” あるいは “s” を接続させた文字列も検索

このように対処することにより、利用者が求める検索結果が得られるようにした。

ただし、このような対処では問題が生じる可能性がある。表 1 には “dgah (喜び)” と目的格助辞 “r” とで “dgar (喜びに於て)” となる例を挙げているが、辞書 [4] には別の意味 (地名) の “dgar” が見出し語として挙げられており、理論的には、ここで我々が採用した検索語の拡張の方法は、結果的に誤った検索結果もヒットさせてしまう可能性を生じさせてしまうのである。

このように、検索文字列の拡張を行うことにより、別の単語も検索対象になってしまう事例にはどのようなものがあるかの事例収集、またそのような事例への対策等については、今後の課題である。

3.2.2 『検索文字列』と格助辞との関係

助辞に関しては、別の点でも注意すべき点があった。

我々は検索文字列として入力されるのは多くが名詞などの内容語であり、助辞などの機能語が検索文字列においては重要性を持つことがないという仮定を無意識のうちに立てていた。だがその後、逆に、ある特定の格にこだわった検索結果が欲しい状況があるので助辞もちゃんと検索できるようにしてほしいとの指摘を受けた。

これは具体的には、たとえば Tib. の名詞 “saṅs rgyas (ブツダ)” に作具格助辞 “kyis” が付いた “saṅs rgyas kyis” は、Skt. の “buddha” の Inst. 格 (具格) である “buddhena” の訳語以外にも、Nom. 格

(主格)である“buddhaḥ”の訳語として用いられることがあるが、このような使い分けには何か理由があるのかを調べたいことがある、とのことであった。

この要求に対応するため、前節で述べた検索文字列の拡張に対して以下のような制限を加えた。

- 検索文字列末尾の音節が、あらかじめ用意した助辞のリストのどれかと一致するときは、検索文字列の末尾に助辞が指定されたと見なし、検索文字列の拡張は行わない
- “\$”という記号を導入し、検索文字列末尾に“\$”が指定されたときは検索文字列を拡張しない。

本節で述べた対処については、電子テキストとリンクさせた情報量も些細なものであり、また対応も非常に簡単なものである。しかし、この程度の文法情報を追加するだけでも、電子テキストの利便性が多少は向上する例になっているのではないかと思われる。

なお、現在のところ、自立語の中で「活用があるもの」に分類されている動詞への対処については今後の課題である。

3.3 文中単語の同定に向けて

検索結果表示の際に、利用者が入力した Tib. の検索文字列を強調表示して視認性を上げることはすでに早い段階で行っていた。これに対し、検索文字列だけを強調表示するのではなく、もし可能であれば Tib. の検索文字列に対応した Skt. 文字列も一緒に強調表示すると視認性が上がって良いのではないかと、との要望があった。

この要望自体は、電子テキストの利便性向上とは直接関係がない内容に思われる。しかし、すでに述べたとおり、eDic プロジェクトにおける我々の将来的な目標の一つに「『検索文字列』として入力されたチベット語表現にもっともふさわしいであろう Skt. 表現を推定して提示すること」がある。電子テキストを対象とした、そのような知識処理を今後推進していくためには、「文」をいかにして「単語列」に切り分けるかという課題や、処理の対象となる電子テキスト中に含まれる単語列をいかにして「単語語幹(辞書の見出し語)と文法情報など各種文法情

表 2: 併合された Tib./Skt. 対訳辞書

Tib.	Skt.
byañ	uttara
byañ chub	sambodhi; bodha; bodhi
byañ chub chen po	mahā-bodhi; mahā-bodha
byañ chub sems dpaḥ	bodhisattva
byañ chub dam pa	parama-bodhi

報の羅列」に自動変換するかという課題などを克服していく必要がある。我々は、先の閲覧システムの視認性向上に関する要望に向けた取り組みは、そのまま今後克服すべき2つの課題への取り組みの端緒となるのではないかと考えた。

本節では、これらの課題に対する我々の取り組みについて述べる。

3.3.1 語彙集を用いた Tib./Skt. の対訳語の推定

本稿で用いる語彙集 本稿では、eDic の電子テキストに、既存の対訳語彙集の情報をリンクさせることによって、対訳語の自動推定などを試みる。ここで使用するのは、電子化され公開されている以下の対訳語彙集である。

- **Mahāvvyutpatti**[6]: 9世紀にチベットの Khri lde sroṅ brtsan 王の勅命により作成された欽定翻訳語彙集で、術語が中心。電子版は、榊 [7] をベースとしている。見出し語項目数は 9,581 項目⁴。
- 瑜伽師地論索引 [9]: 『瑜伽師地論 *Yogacārabhūmi*』の Skt., Tib. および漢訳版を網羅した索引 [10] の電子版。本稿では、このうち Skt./Tib. の対訳部分のみ⁵ 22,003 項目を用いる。

これら2つの対訳語彙集を併合した約 27,600 項目の語彙集を以後、辞書として用いる。この辞書の一部項目を表 2 に例示する。

作業手順 現在我々が行っている訳語対推定の手順は簡単なもので、以下のとおりである。

⁴各項目の中には「A あるいは B」のように記述されたものもある。このように列挙された各要素を別項目として数えると約 13,000 項目。

⁵漢訳の項目を除去し、Skt. 項目と Tib. 項目のどちらかが欠ける項目を除去し、さらに重複した項目を除去したもの。

- 図 2 のような構造の対訳テキストの Tib. 部分に対し、先頭音節から順に辞書見出しとの一致を調査。図 2 の例では、まず Tib. 文の先頭の “byañ chub sems dpaḥ sems dpaḥ …” を対象とした Tib. の辞書引きを行う。ここで表 2 を辞書とすると、“byañ”, “byañ chub”, “byañ chub sems dpaḥ” の 3 項目が Tib. 単語列候補として得られる。
- 辞書見出しとの一致が取れた場合、辞書にある対応する Skt. の項目を引く。上の項目で挙げた例および表 2 に従うと、順に “uttara”, “sambodhi; bodha; bodhi”, “bodhisattva” といった Skt. の対訳語候補が得られる。
- 図 2 のようになっている Skt. 文の中から、Skt. の各対訳語候補にマッチする単語を探す。図 2 の例では、“bodhisattva” のみが “bodhisattvo” の語形で出現している⁶。その他の “sambodhi” 等は見つけることができない。
- 対訳語候補が文中に見つかったときは、それら単語対が「訳語対候補」となる。図 2 の例では、Skt. 文に “bodhisattva” を発見できた “byañ chub sems dpaḥ” と、その対訳語 “bodhisattvo” のみが訳語対候補となる。
- 次の音節から始まる文字列を対象にして、これまでと同様の処理を行う。図 2 の例では、Tib. 文の先頭文字 “byañ” の次の音節から、“chub sems dpaḥ sems dpaḥ …” に対する処理を、ここまでの手順と同様に行う。

なお、このような手順で単語の対応付けを行うと、一つの音節が複数の Skt. 訳語対候補に含まれてしまうことがある。このときは、現在のところは以下のように選択している。

- 一方が他方を全部含んでいるときは、より大きい方を選択 (例: 候補 “byañ chub” を構成する音節が全部、候補 “byañ chub sems dpaḥ” の一部であれば後者を選択)
- 前後で重なるときは、前にある方を選択 (例: 候補 “byañ chub sems dpaḥ” の後ろ 2 音節と、

⁶文中単語の “bodhisattvo” は辞書項目 “bodhisattva” が格変化し、さらに連声により音韻変化した結果である。このように Skt. では言語知識をまったく持たない状況での単語の探索はほぼ不可能と言ってよい。ここでは、別稿 [3] にて我々が言及した Skt. の連声・単語活用解析システムの簡略版を利用して対応単語を探すようになっている。また Skt. 文においては、文中の、空白で区切られた各要素を「単語」として扱っている。

表 3: 辞書を用いた対訳語の発見

SP-22		ALL:2215	
A		B	
1193 (53.7%)		1335 (60.3%)	

“sems dpaḥ chen po” の前 2 音節が重なったときは、前者を選択)

3.3.2 評価

我々は、前節で示した作業を eDic のテキストデータに対して行い、対訳語の自動推定の作業を行った。作業の内容に関する評価を行うため、評価用の実験を行うこととした。

- どの程度で対訳語対の推定を行うことができたか
- 「文」をどの程度「単語列」に切り分けられたか

実験は、*Saddharmapuṇḍarīka* (SP) 第 22 章のテキストを用いて行う。まず、SP 第 22 章のテキストを以下のように 2 種類用意する。

- A) 図 2 のような、普通の対訳形式のもの
- B) A) と同じ形式のデータに対し、Skt. 文については、それぞれの文中単語を語幹 (辞書見出し) に書き換え、また Tib. 文については単語ごとに区切りを入れたもの

この 2 つのテキストを対象として、それぞれ前節で述べた手順での対訳語推定の実験を行い、結果を比較する。

辞書を用いた対訳語の発見 前節の B) のテキストは、Tib. はすでに単語で切られており、また Skt. も語形解析をする必要がない状態になっていることから、テキスト中の単語を単純に辞書引きするだけで訳語対候補が決まる。それゆえ B) の実験結果は、現行の辞書と枠組を用いた場合での最も理想的な結果と見ることができる。

実験の結果を表 3 に示す。これによれば、SP 第 22 章に含まれる Tib. の 2215 単語のうち、Skt. の対訳語を見つけることができたのが A) では約 55%、B) でも約 60% にすぎないことがわかる。B) で 60%

表 4: Tib. 文の単語分割の結果

SP-22		ALL:1932	
自動区切り数	正解	再現率	適合率
1268	1129	58%	89%

程度の結果であったが、これが我々が今回用いた辞書の情報量の限界であろう。

また A) と B) の結果には特段の差がないことがわかる。これはすなわち、Tib. の単語については、あらかじめ単語区切りを入れた状態でなくても、訳語対の推定作業にはあまり支障がないということを示しているように思われる。

Tib. 文を単語分割 訳語対候補が見つかった Tib. の単語候補、たとえば先に例であげた “byañ chub sems dpañ” などを単語として扱うのであれば、その先頭にある “byañ” の直前に単語区切りが入っていると考えることができる。すなわち、訳語対候補を推定する作業とは、同時に Tib. 文に単語区切りを入れる作業であったとも見ることができる。そこで、テキスト A) に対する訳語対候補推定の作業の結果生じた単語区切り箇所と、テキスト B) においてあらかじめ手作業で行った単語区切り箇所との比較を行った。

比較の結果を表 4 に示す。自動的に区切られた箇所と、手作業によって区切られた箇所とが一致するときを「正解」とした。また、手作業による区切り箇所の個数に対する「正解」の個数の比率を「再現率」、自動的な区切り箇所の個数に対する「正解」の個数の比率を「適合率」とした。

この結果を見ると「再現率」の値が 60% 程度になっている。これは、表 3 で示された、自動で正しく得られた訳語対の、テキスト中全単語に対する割合が 50~60% 前後であることと関係していると思われる。一方の「適合率」が 90% 近くに達した点については、今後さらに知的処理を交えて Tib. 文を自動的に単語に区切る作業を行う時にも、現在用いている語彙集から得られる情報はかなり役立ちそうなことが期待できる。

現状の枠組では、推定できた訳語対の数が少なかったこともあり、今回の実験で得られた結果をそのままさらに何らかの処理の基礎データとすることはできない。しかし、eDic 形式の電子テキストと、

対訳語彙集をリンクさせることにより、さらに新たな、そしてある程度の信頼性を持った電子情報が構築できることが確認できた。

3.3.3 対訳語関係の閲覧システムへの取り入れ

表 3 の結果が示すとおり、今回の実験結果によって推定できた単語対の数は eDic データの全単語の約半分程度に留まっている。現状の結果では、実用的な成果が得られたとは言えない。しかし、閲覧システムの補助として、対応付けができた単語組をそれとなく示す程度であれば、多少精度が低いデータであっても、利用者の助けになるのではないかと思われる。それゆえ、目立たない程度に、本節で述べた自動的な単語対応付けの成果を eDic 閲覧システムの画面表示に取り入れることとした。

本節の成果を文献閲覧システムに取り入れる際に問題となるのは、それがあくまで単語レベルの情報にすぎない点である。すなわち、たとえば同じ構文中に、同じ単語が複数登場したときに、どの単語が対応関係にあるのかを判断するための情報がない。また、Tib. 文に同じ単語が複数存在するにもかかわらず、Skt. の対応語が一つしか発見できなかったときも単語の対応関係の絞り込みができない。我々は、同じ単語・対応単語とされた単語群を「グループ」とし、WWW ブラウザ上で、あるグループに属する単語上にマウスを置いたときに、同じグループに属する単語の文字色を変化させ、強調表示させる仕組みを取った⁷。

本節における成果は、現状ではまだまだ電子テキストの利便性向上に直接結び付くものではない。しかし、電子テキストの閲覧のしやすさに役立つという点で、閲覧システムの利便性を向上させ得たのではないか。

3.4 未対応の要望

ここまで述べてきた以外の要望で、現段階で十分な対応ができていないものとして主に以下がある。

- それぞれの検索結果について、類似する構文が提示されると非常に良い

⁷マウスの位置によって文字色を変化させる仕組みは、簡単な JavaScript を用いて実装した。それゆえ、すべての WWW ブラウザに対応する仕組みとはなっていない。

- eDic 本来の目的とは異なるが『Skt. では同じ単語を使っているのに、Tib. で状況によって翻訳語を使い分けている』ことがある。この時は Skt. をキーにしたテキスト検索が有効なので、Skt. をキーにした検索機能があるとよい

これらの対処については、今後の検討課題としたい。

4 おわりに

電子テキストの利便性を向上させるためには、他のさまざまな情報とリンクさせると良いはずである。しかし、電子テキストにどのような情報をリンクさせると電子テキストの利便性が向上するのか — この問いへの我々の取り組みは、いまだ十分な成果を得るには至っていない。しかし、実際に eDic 閲覧システムを構築で、実際に文献研究者の人たちに使ってもらうことにより、利用者の人たちも、従来よりも「どんな情報があればもっと便利になるか」が具体的にイメージしやすくなってきているのではないかと考える。それゆえ、eDic 閲覧システムの改良などを通じ、今後も引き続き電子テキストのよりよい利用形態についての模索を継続していきたい。また、eDic 形式の電子テキストと、対訳語彙集をリンクさせることにより、さらに新たな、そしてある程度の信頼性のある電子情報が構築できることが確認できたが、この試みをさらに推進させ、Skt. や Tib. に関する電子情報の構築に取り組んでいきたい。

我々が試作したシステムは [1, 8] の URL からアクセス可能である。最初に eDic 閲覧システムを公開してから、本発表で述べた要望が集まるまでに、数年もの歳月を要した。システムの改良に関する要望が集まるのにここまで時間を要した理由の一つには、eDic 閲覧システムが想定する利用者層の絶対数の少なさがあると思われる。前述のとおり、eDic 閲覧システムは、対象となる古典仏教文献や Skt. および Tib. に対して相当に高度な理解力および問題意識を持ち、辞書に載っている語釈文を見ただけでは満足しないような、ごく一部の文献研究者向けのものである。しかしこれでは、我々の「電子テキストのよりよい利用形態についての模索」が遅々として進まないというのはある種当然のことと言える。それゆえ今後は、もし可能であれば、本システムを

より多くの人、なかでも Skt. あるいは Tib. の初学者にとっても便利なものにしていく、あるいは別の言語の電子テキスト、たとえば古典中国語なども扱うことのできるようなシステムにしていき、利用者の裾野を拡大していければと考えている。

参考文献

- [1] T. Aiba. eDict Viewer. WWW. (Nov 10, 2005) URL: <<http://www-asia.human.is.tohoku.ac.jp/demo/eDv/html/>>.
- [2] T. Aiba. Table of Transliteration Schemes for Sanskrit and Tibetan. WWW. (Nov 10, 2005) URL: <<http://texa.human.is.tohoku.ac.jp/aiba/codes/table/>>.
- [3] 相場徹, 生出恭治. 語幹情報に注目したサンスクリット文献閲覧システムの試作. 情報処理学会研究報告, Vol. 2004-CH-63-5, pp. 31-38, 2004.
- [4] Sarat Chandra Das. *Tibetan-English Dictionary*. Calcutta, 1902 (Reprint in Kyoto, 1969).
- [5] 稲葉正就. チベット語古典文法学. 法蔵館, 京都, 1979 (改訂版第 5 刷).
- [6] 蜜波羅鳳洲. The digitalized text of *mahāvīyūtpatti* based on Sakaki's edition. WWW. (May. 1, 2002) URL: <<http://texa.human.is.tohoku.ac.jp/aiba/archive/mvyut/open/>>.
- [7] 榊亮三郎. 梵藏漢和四訳対校 訳名義大集. 真言宗京都大学, 京都, 1916.
- [8] 鈴木隆泰. Tibetan-Sanskrit 構文対照電子辞書プロジェクト eDic. WWW. (Nov. 10, 2005) URL: <<http://suzuki.ypu.jp/edic/>>.
- [9] K. Yokoyama and T. Hirokawa. Chinese-sanskrit-tibetan table of buddhist terminology based on the *yogacarabhūmi*. WWW. (Aug. 1, 2000) URL: <<http://www.buddhist-term.org/yoga-table/>> (Nov. 2005 現在、サイト消失).
- [10] 横山絃一, 廣澤隆之. 漢梵藏対照 瑜伽師地論總索引. 山喜房佛書林, 1996.