

OTDO における古チベット語文書の オンライン共同編集作業

松田 訓典¹ 星 泉¹

概要: Old Tibetan Documents Online (OTDO, <http://otdo.aa.tufs.ac.jp/>) ではこれまで wiki ベースで関連資料の編集作業を行ってきたが、そこにはいくつかの利用上の問題点があった。この状況を改善するために現在新しいインターフェースへの移行を試みている。本稿では本プロジェクトの概要と従来の wiki ベースの編集作業での問題点、その改善にむけての取り組みと展望について報告する。

キーワード: OTDO, チベット語コーパス, オンライン共同編集, データベース

Online Collaborative Editing in the Old Tibetan Documents Online

KUNINORI MATSUDA¹ IZUMI HOSHI¹

Abstract: In the Old Tibetan Documents Online (OTDO) project, the wiki-based system was used to collaboratively prepare the texts and related information, but certain problems or unsatisfactory points in the wiki-based system were observed. To alleviate these problems, we are developing another collaborative editing system. In this paper, we report the outline of our project, problems we encountered with the wiki-based system, and merits of the new system and areas for improvement.

Keywords: OTDO, tibetan corpus, online collaborative editing, database

1. はじめに

Old Tibetan Documents Online (OTDO) では、およそ七世紀から十二世紀と推定される敦煌写本、碑文、その他関連する資料を、アルファベット転写し、校訂を加えた古チベット語文書のコーパスとして作成・公開している。

このプロジェクトは、1970年代にフランス国立科学センター (CNSR) の今枝由郎氏によって開始された古チベット語文書の集成に端を発する。それが1980年代になって、当時東京外国語大学アジア・アフリカ言語文化研究所 (AA研) が保有していた大型計算機のチベット語入力システムを利用して敦煌文書のデータを入力するようになり、その成果が1990年、文献 [1] として出版された。さらにその

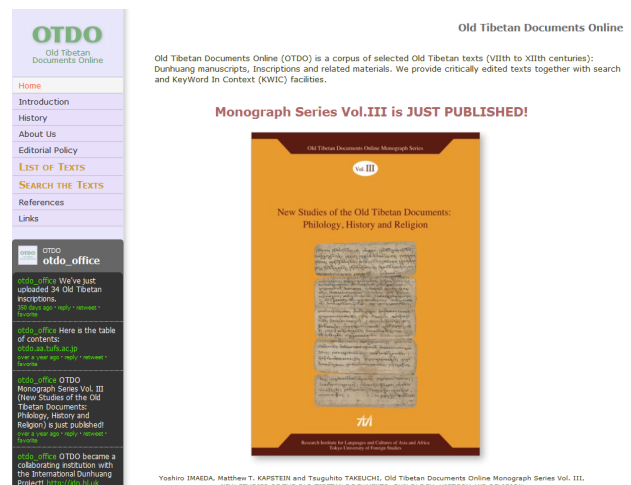


図 1 <http://otdo.aa.tufs.ac.jp/>
Fig. 1 <http://otdo.aa.tufs.ac.jp/>

¹ 東京外国語大学アジア・アフリカ言語文化研究所
ILCAA, Tokyo University of Foreign Studies

仕事は 1990 年代後半になって AA 研の星泉に託され、同計算機を使った仕事を継続しつつ、2001 年に [2] が出版された。その後、時代の変遷に伴い、大型計算機のデータをすべてテキストデータに変換し、それをもとに試行錯誤を繰り返しながら、web 上でも公開するようになった。当初はそのテキストをそのまま掲載するという形のものであったが、2003 年に最初の検索システムもつくられた。これが OTDO の始まりであり、2001 年から 2006 年まで AA 研に置かれた COE プロジェクト GICAS (「アジア書字コーパスに基づく文字情報学の創成」、Grammatological Informatics based on Corpora of Asian Scripts) のサブプロジェクトとしても位置づけられていた。

2004 年ごろになり、OTDO に一つの大きな転機が訪れた。wiki への移行である。これを契機にオンラインでの共同編集が開始され、現行システムの基礎となる新しい検索システムも作られた。同時にその成果は AA 研から OTDO モノグラフシリーズとして出版が継続されている ([3]-[5])。また、International Dunhuang Project (IDP, <http://idp.bl.uk>) との連携も行われ、その果たす役割を大きくしている。

OTDO では以上のような経緯により、その資料の収集・集積にあたって、これまで wiki を利用したコーパスおよび関連資料の編集作業が行われてきたが、対象となるテキストの増加や編集者の入れ替わりに伴いいくつかの問題点・不足点が見出されてきた。そこでそれに対応すべく、昨年度より新インターフェースの構築を進めることとなった。

本稿では、共同編集作業が開始されて以来の wiki ベースのシステムの紹介とその問題点、それに対して今回新たに取り組んでいる新インターフェースの構築作業について報告する。

2. OTDO の既存オンラインサービス*2

2.1 公開コンテンツ

現在公開されている古チベット語文書は 100 点あまりに及んでおり、ウェブサイト上では KWIC コンコーダンスとして利用することができ、IDP との相互リンクも行われている。一例として、図 2 は “bod yul” (チベット) の検索例であり、テキストの順序に応じて出力されているが、前後の文字をチベット語の語順でソートして表示することも可能である。個別のテキストに関しては、図 3 のような段落形式、あるいは行単位形式で参照可能であり、上部には目録的情報が併記されている。

2.2 編集用 wiki

他方、テキストおよび関連情報の構築に際しては、先述の通り wiki が用いられてきたが、過去の経緯により、二

つのシステムが併存している。一つは PukiWiki*3 ベースのものである。これは主として 2006 年頃までの編集の場であり、現在公開中のデータの中では、敦煌文書の管理に使用されている。もう一つは MediaWiki*4 ベースのもの

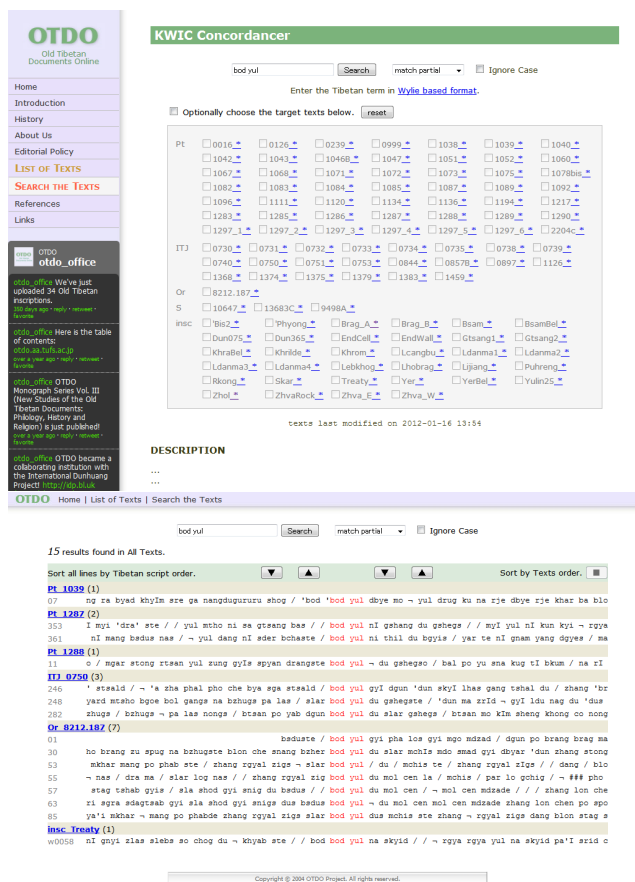


図 2 KWIC サーチ例
Fig. 2 sample of KWIC search



図 3 テキスト表示例
Fig. 3 sample of text display

*2 本システムの構築は有限会社ツクラカンの千葉文博氏による。

*3 <http://pukiwiki.sourceforge.jp/>

*4 <http://www.mediawiki.org/wiki/MediaWiki/ja>

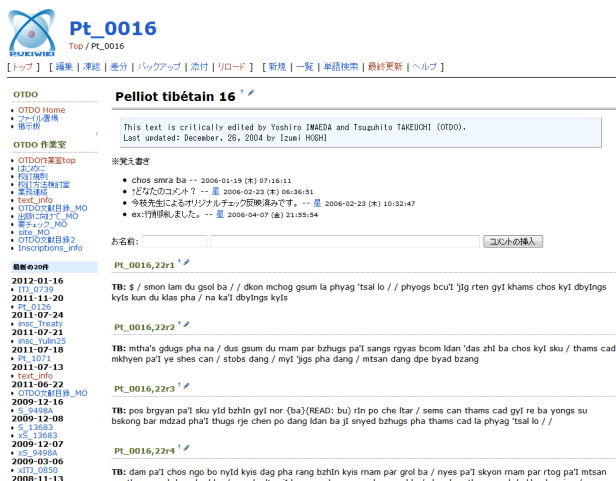


図 4 PukiWiki でのテキスト編集例

Fig. 4 sample of editing on PukiWiki (Text)



図 5 PukiWiki でのテキスト情報編集例

Fig. 5 sample of editing on PukiWiki (TextInfo)

であり、移行期間を含むその後の活動の場となっている。現在は主として碑文データが含まれている。

仕様としては共通する部分が多くを占めているので、以下では主として PukiWiki ベースの方を例にあげて紹介したい。例えば図 3 に対応する wiki の編集画面は図 4 のようになっている。

このような wiki のページは大別すると三種に分けられる。

- テキスト転写・校訂 (文献ごと)
- テキストの一覧・目録情報 (まとめて一ページ)
- 参考文献 (まとめて一ページ)

実際の検索サービスでは、前二者が組み合わされて表示される形となっている。いずれも基本的構成要素としてはテンプレートと各項目の羅列という形をもつものである。

PukiWiki と MediaWiki とで参考文献の項目など些細な部分の差異はあるが、少し大きなものとしてユーザ管理



図 6 PukiWiki での参考文献編集例

Fig. 6 sample of editing on PukiWiki (Bibliography)

の方法にちがいがあ。前者の場合は個別のユーザ管理は行われておらず、編集は事前の合意にしたがって行われており、一方、後者では個別のユーザ管理が行われているという点で編集管理上進歩しているといえる。

3. 既存システムの長所と短所

3.1 既存システムの長所

前節で紹介した既存の wiki システムの概要は、一見してわかるように、データの構築という点で非常にシンプルなものである。各々のページに関して言えば、おおよそ、編集上のコメント類、テンプレート、実際の内容という三つの要素から構成されており、編集者は既存の内容を踏まえ、修正を施し、テンプレートをコピーして新規項目を追加し、必要に応じてコメントを残すという作業を行っている。ここで作り上げられたデータが、一方で公開検索用システムに随時インポートされ、他方で出版用に組版されて、一般に提供される。

この方式の最大の利点は、システム面でのコストは最低限に抑えられるということではないかと思われる。wiki そのものもつ編集機能や履歴管理、その他豊富な利用実績をもつプラグイン等、少なくとも共同編集という点に関しては、その出発点として非常にふさわしいものであったと考えられる。

3.2 既存システムの短所

結果として十年を越える研究活動がとりたてて大きな問題もなく進められてきたわけであるが、これには OTDO 編集者たちに関する事情も関係しているといえる。つまり、これまでの活動では、十人未満という比較的少人数かつほぼ固定されたメンバーによる作業であったことである。そうした状況では、事前に一定の取り決め・編集方針・書式・確認事項等を共通認識として保持していくことが比較的容易である。しかしながら、収集する情報あるいはそのタイ

プの変化や増加, 人員の異動の可能性とともに, いくつかの問題^{*5} が現れてきたことは当然予想されるべきことであろう。ここではそのいくつかの問題点を列挙しよう。

個別化されたシステム

まず第一に, システムが大きく分けて三つに分かれている点である。検索表示システムと編集システムとしての PukiWiki および MediaWiki である。編集者側から見れば, 表記の規則の不統一になってしまっていたり^{*6}, 一定の区切りがついた段階で編集結果が検索システムに反映されるまで, 編集者とその間の編集作業を検索システムを使用して利用することができない, といったことなどが問題となっている (a)。また, 新たな編集者が加入した際も, 過去の経緯を知らないため, 混乱を招きかねない状況にあるといえる。

データ構造

第二点として, 先に紹介したように, 既存システム上の基本作業は, テンプレートのコピーとその編集である。そこには十分なクロスリファレンスが考慮されておらず^{*7}, 羅列に過ぎないものといえる。この点で, データの蓄積と共に整合性のチェックに手間のかかるものとなっている。と同時に基本的には細かなメタデータをもたない単なるテキストである点も, データの応用性という意味で不十分であるといえる (b)。

編集管理

次に, 編集管理上の問題としては, 編集作業の進み具合を容易に視覚的に確認できる手段をもっていないという点がある。つまり, 一つのテキストの編集あるいはチェック作業に複数人が携わっているにもかかわらず, 互いにそれを確認するすべが基本的には各テキストページ上にあるコメント以外に用意されていなかった。その結果として, 個別のテキストの進行度合いあるいは全体の進み具合の把握にもいづらか不具合を生じてしまっているといえる (c)。

その他

また, 検索面での問題では, 時代やジャンル別の検索ができないこと (d), 行をまたいだ検索, また転写テキストにみられる検索上の問題 (n/ng の区別^{*8} など) 等の改善も期待される。

*5 ただしここでいう問題とは, 利用者にとっていささか不便な点, あるいは発生することが将来懸念される問題なども含む。

*6 これは使用時期が異なるため徐々に改変が行われた結果でもある。

*7 もちろん項目ごとにページをわけると, クロスリファレンスとして機能させることも可能であろうが, 少なくとも既存システムでは編集の便宜上そのような形はとられていなかった。

*8 チベット語などの転写においてはひとつの文字に対して複数のアルファベット文字の組み合わせとして転写されることがしばしばある。そのため単純に転写テキストを検索するだけでは両者を区別することはできない。

4. 新しいインターフェース

4.1 基本的立場

前節で述べたような問題が自覚される中, 昨年度より本システムに関する見直しが試みられた。これらの問題, とりわけ編集上の問題の多くは, 既存の wiki ベースのシステムのページ構造を変え, フォームの設置等の措置をとることで解決できる部分であると思われるが, 検索システムとの統合という点では全く新しいインターフェースを準備する方が容易であると判断するにいたった^{*9}。この新インターフェースの作成にあたって留意した点は主として以下の点である。

- 既存データの負担の少ない引き継ぎ
- 既存のインターフェースが提供する機能は可能な限りカバーすること (ユーザ管理・履歴管理・シンプルさ)
- 先に挙げた問題に対する一定の改善策の提供

したがって以下ではこの点を踏まえつつ新インターフェースを紹介していきたい。なお, 本システムはバックエンドとして PostgreSQL を使用した Ajax ベースのインターフェースを採用している。

4.2 構成

基本的な構成要素としては, 大枠として従来同様に三つ (+管理画面) とした。

- テキスト転写・校訂 (文献ごと)
- テキストの一覧・目録情報 (一覧・個別)
- 参考文献 (一覧・個別)

アカウントの種別としては, admin (管理者) と member (編集者) を設け, それぞれの作業権限を区別している。

つまり, 本インターフェース上では, 非ログインユーザは各種検索のみを行うことができ, member は適宜その場で編集画面に切り替えることで編集作業を行い, さらに admin はユーザ管理を含むすべての作業を行うことができるようになっていく。このようにすることで, 同一のシステムの中で利用と編集の両者をリアルタイムかつ一連の手順で実行できるようになり, 編集時の負担・混乱の軽減が期待される (a)。

テキスト編集とその管理

図 7 はテキストの編集画面である。基本的には行単位の編集であり, 各行ともに行ラベル・行の内容・それに対するコメントからなる。ここでは, 行ごとの編集をサポートするとともに, 主として wiki での編集になれた編集者のための文獻単位でのテキストファイルとしてのインポート・エクスポートを可能にしている。なお, 編集機能に関しては, 現時点では wiki 時の書式に加えていくつかの入力補

*9 これに関しては近年比較的近い分野で永崎研宣氏らによって行われている web コラボレーションシステムの試み (文献 [6]) も踏まえた結果である。

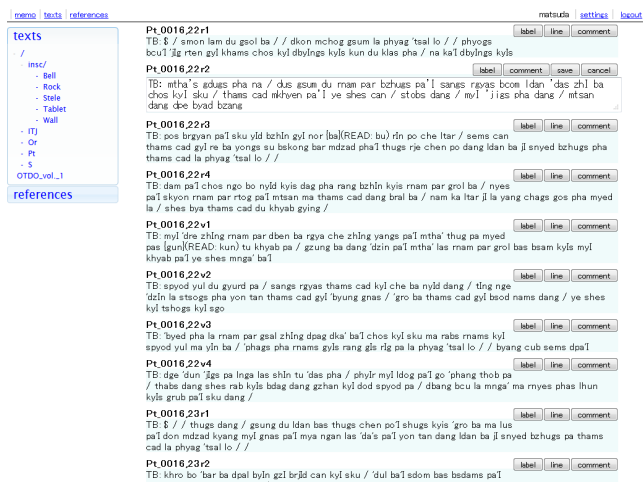


図 7 テキスト編集例

Fig. 7 sample of text editing

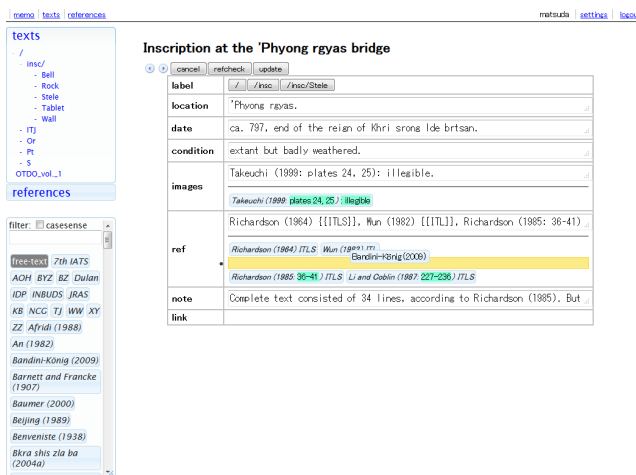


図 9 テキスト情報編集例

Fig. 9 sample of text information editing



図 8 テキスト表示例

Fig. 8 sample of text display

助機能(校訂記号の挿入など)を追加しているのみである。

図 8 はテキストの一覧・検索用の画面である。ここでは Gmail などを用いらている「階層つきラベル」機能を導入することによって、時代別・ジャンル別などの多様なテキストの種別からの効率的な検索を提供する (d)。なお、ラベル用と文字列用の検索窓を用意し^{*10}、階層つきラベルについては別途左段にツリー表示を設けて、目次機能の代替としても一覧性を高めている。同時にラベル機能によって、未校正・校正済み等の判別やその一覧を容易に視覚化することも可能であり、一般公開・未公開の制限を付加することもできるようにしている (c)。

目録情報・書誌情報

次に具体的なテキストの目録情報の表示に関しては、図 9 のような形で用意している。基本的には従来の項目を踏襲しつつ、主にクロスリファレンスの問題を解消している (b)。具体的にはツールボックスとして左段にリファレンスの一覧(フィルタリング付)をボタン状に配置し、それを表中にドラッグアンドドロップすることにより容易にクロスリファレンスを登録・編集することができるようにしている。また、既存の wiki 方式で記述されたテキストを一定程度までパースすることも可能なようにしており、適宜

*10 検索については実装中であり、今後仕様の変更があらう。

エラーを表示することでクロスリファレンスの不整合を解消する手助けをするとともに必要なメタデータの抽出も試みている。参考文献類についても同様の処理を施している (b)。

5. 新インターフェースの課題

前節では現在構築中の新しいインターフェースの概要について説明した。

そこで残された課題としてまず挙げられることは、現時点で不足している機能を実装し、実作業に携わる編集者とのやり取りを通して充実を図ることは言うまでもないであろうが、これに関してはおよそ時間の問題であると考えている。

一方で、現在の wiki ベースのデータを基本的に引き継いだ形でのデータには十分にメタデータ化されていない情報が多数残されてしまっている。例えばテキストの異読情報等である。現状ではこうしたデータは検索側プログラムで大まかな処理を施すことによって対応しているが、決して望ましい状況とは言えないであろう。これに関しては、一定のパース処理を介入させることによってメタデータ化を促すようにしていく必要があると考えている。またそれに対して広く普及している規格との関連は不可欠であり、その点についても改善の余地が残されている^{*11}。

6. 結び

以上、これまでの OTDO の取り組みと既存のシステム、それに対して新たなインターフェースのあり方を報告してきた。本プロジェクトが 1980 年代という比較的早い段階からコンピューターを利用した研究を行い、構成員としては小規模とはいえ、オンラインでの学術的共同作業を通してその成果を公開してきたこと、そしてその経験を踏まえた上で新たなインターフェースを構築しつつあることは

*11 現時点では TEI/XML の適用が好ましいであろうか。

非常に有意であると考えている。今後も実作業に携わる編集者の意見を整理し、ユーザビリティの向上に努めると同時に、共同作業における問題点とそれに対する緩和策について検討していきたいと考えている。

謝辞 本稿で報告した新インターフェースの構築と本稿の執筆にあたっては、従来のシステムを構築した有限会社ツクラカン千葉文博氏、および人文情報学研究所永崎研宣氏に多大なご教授をいただいた。ここに記して感謝の意を表す次第である。

参考文献

- [1] Imaeda, Y., Takeuchi, T.: *Choix de documents tibétains conservés à la Bibliothèque nationale*, Tome III, Corpus syllabique, Bibliothèque nationale, Paris (1990).
- [2] Imaeda, Y., Takeuchi, T., Hoshi, I., Ohara, Y. and Ishikawa, I.: *Choix de documents tibétains conservés à la Bibliothèque nationale*, Tome IV, Corpus syllabique, ILCAA, Université des Langues Étrangères de Tokyo, Tokyo (2001).
- [3] Imaeda, Y., Takeuchi, T., Hoshi, I., Ohara, Y., Ishikawa, I., Iwao, K., Nishida, A. and Dotson, B.: *Tibetan Documents from Dunhuang, kept at the Bibliothèque Nationale de France and the British Library*, Old Tibetan Documents Online Monograph Series Vol. I, ILCAA, Tokyo University of Foreign Studies, Tokyo (2007).
- [4] Iwao, K., Hill, N. and Takeuchi, T.: *Old Tibetan Inscriptions*, Old Tibetan Documents Online Monograph Series Vol. II, ILCAA, Tokyo University of Foreign Studies, Tokyo (2009).
- [5] Imaeda, Y., Kapstein, M. T. and Takeuchi, T.: *New Studies of the Old Tibetan Documents: Philology, History and Religion*, Old Tibetan Documents Online Monograph Series Vol. III, ILCAA, Tokyo University of Foreign Studies, Tokyo (2011).
- [6] 永崎 研宣, 苔米地 等流, Wangchuk, D., Almogi O., 下田 正弘: 人文学のためのコラボレーション—ITLR コラボレーションシステムの開発を中心的事例として—, じんもんこん 2011 論文集, 155-160 (2011).