

## オンライン版ハウサ語電子辞書<sup>1)</sup>

中 村 博 一\*

### The Online Version of Electronic Hausa Dictionary

Hirokazu NAKAMURA

One of the key issues on which contemporary anthropological studies have focused is the ethnography of media where anthropologist himself or herself would be also involved as an audience or tossed back and forth by the big IT wave. This article is concerned with a small tool that floats in this ocean: the online version of electronic Hausa dictionary, compiled by an anthropologist in 2002. It could hopefully be a case to find a solution for digital divide in the IT globalism and even in the Hausa localism.

#### 概要

モダンティに関する近年の人類学的な論考において移動とともにキーワードとなるのはメディアである(例えばAppadurai 1996)。そこにはメディアに翻弄されながら活動するわれわれ自身の姿やその道具自体も当然映し出されねばならないだろう。特に電子的媒体により資料の使い方や研究の様式は大きく変わってきたのだから。90年代後半から10年に満たない間の変貌を考えてみればよい。図書をはじめ基本情報を手にするためのネットワークは現在欠かすことができないし、YahooやGoogleの検索は全く日常的な手順になっている。

本稿が関わるオンライン辞書にしてもブロードバンドが普及したこの数年でようやく実用的になったと言えるだろう。いくらプログラムが秀逸でもネットワーク整備なしにはそ

の特性を生かせないからだ。こうした動向はサハラ以南のアフリカにおいても基本的に変わらない。2年あまりの間に北部ナイジェリア各大学のインターネット利用は急速に一般化し、スタッフは研究室で無線LANを使っている。光ファイバーの敷設も進行中である。頻発する停電もあらかじめ対策を立てた上でネットワーク構築が行われているため、あまりストレスを感じない。例えば北西部のウスマン・ダン・フォディヨ大学と本学と大きな差があるとすれば速度と台数のみということになるだろうか。この大きな潮流によってパソコンとネットワークがあれば特別な投資や環境を要しないオンライン辞書の有用性はますます高まっていくと思われる。

1999年より筆者はサハラ以南のアフリカの言語のうち西アフリカで広く通用するハウサ語電子辞書の普及に力を入れてきた。残念なことにオンライン辞書の有用性が高まる一方で言語間の情報格差は深刻だと言わざるをえない。サハラ以南のアフリカの諸言語におい

\* なかむら ひろかず 文教大学人間科学部人間科学科

で電子化ないしはオンライン化された辞書はほとんどないからである。しかし特殊なフォントがそれほど必要でなければ、技術的にも経済的にも比較的容易にオンライン辞書開発が可能なのは意外に知られていない。

本稿では、2002年に運用を開始したオンライン版ハウサ語辞書（約4万語収録）について概観し、開発の要件やこれまでの利用状況についても見ていく。今後の展望も考える。

#### オンライン版ハウサ語辞書について

2004年9月現在、公開されているハウサ語関連オンライン辞書は2種類ある。ウィーン大学のF. StoiberとJ. Ahmerが開発したハウサ・英・ドイツ語辞書と本稿が扱う辞書である。ウィーン大学の辞書は2001年より公開されている。それぞれの言語に対応する単語が表示される語彙集的な辞書である。本稿が扱うのは、ハウサ語研究・教育界で広辞苑的な位置にあるG. P. Bargery 1934 A Hausa-English Dictionary and English-Hausa Vocabulary. Oxford University Press. のオンライン版である（以下Bargeryとする）。古くなったとはいえ各地域の方言形など収録語彙の多さでは右に出るものがない。UCLAのハウサ語コースでは中級以上の学習者に対しBargeryを参考書としてあげている。母語としてのハウサ語話者が多い北部ナイジェリアの大学のハウサ語研究室では今も高い評価を受けており、若干の語彙集を付加して1993年にナイジェリア国内で復刊されている。

#### オンライン版以前のBargeryの電子化について

紙媒体のBargeryは第2刷（1951年）の後、永らく絶版状態が続いた。その評価の高さと入手の困難さから幻の辞書と言われていた1980年代半ば、東大AA研の松下は大型電算機を使い電子版のBargeryを開発した。他言語と比べてもかなり早い時期に実用化したものだけに、サハラ以南の言語の電子化の先駆的事例としてだけでなくこの業績はもっと注

目されてよいように思われる。1991年にはNECPC98シリーズ互換機のパソコンで動く電子辞書としてリニューアルされた。この時点で普及の条件はある程度整ったと言えるのだが、IBM互換機ではなかったこと、日本国内のハウサ語教育の場が皆無だったことでその後10年、利用者は数人にすぎなかった。2001年になってIBM PC / AT機で動くエミュレータで作動を確認してから、各地のハウサ語・歴史・文化研究者に呼びかけようやくある程度の普及を見るようになった（松下1991、塩田・中村2001、中村2002）。国内唯一の大阪外語大のハウサ語講座でも利用されている。

#### なぜオンライン版か

本稿の目的に沿うと電子辞書は2つに分類できる。単体で利用するものとオンライン版である。単体利用のものはパソコン・携帯端末にインストールしたソフトウェアや、電子辞書として販売されているハードウェアに分かれる。これらは一種の完結した時空間とも言え、情報の修正・更新には手間がかかる。Bargeryのオンライン化にあたって考えたのは、まずはこの点である。前節で述べたエミュレータ・タイプのBargeryにはかなりのバグが見られる。サポート体制が標準的な市販の電子辞書と異なり修正にはかなりの労力と時間が予想された。一度オンライン化してしまえば修正後直ちに最新版を反映することができる。第二にコストパフォーマンスのよさをあげることができる。開発と運営に必要なソフトウェアに投資がいらぬ。ハードウェアについても辞書とウェブ用のサーバを用意し、自宅や職場で使用するブロードバンドの一部を流用すればその日から電子辞書を運用することができる。電気代は1ヶ月数百円。商業的に成功しうるコンテンツと異なり、利用者もそれほど多くはないため、3、4年前の中古パソコンの処理能力でエミュレータ・タイプのBargeryの数十倍のパフォーマンスを發揮できる。4万語の辞書の全文検索もたった数秒で終わる。

このように辞書を開発する側にとってオンライン化の効用は大きい。しかしながら、単体と比べ問題となる点もある。ネットが使えない場所では利用できない、サーバが停止すると利用できないなどである。これらは、現在北部ナイジェリアで見られるようなネットの加速的な普及や、サーバの分散化でオンライン版のメリットに見合うだけの小さなリスクに抑えられるであろう<sup>2)</sup>。

#### オンライン版の考え方

オンライン版辞書とはアクティブなホームページと考えればよい。基本的な検索プロセスは以下のである。1. ブラウザで辞書サイトへアクセスする。2. 検索文字列をタイプして送信する。3. 送信された文字列をウェブ・サーバが受け、検索プログラム(CGIなど)を動かす。4. 検索プログラムが辞書のデータベースを呼び出し、当該文字列を検索する。5. 検索プログラムが検索結果をブラウザに表示する、ないしは表示できる形式(htmlなど)で返す。このように文字列の入力に対応し、検索プログラムとデータベースが連携して働くプロセスがオンライン辞書の要となる。

そこで開発者の実際の作業とは、開発と運用に使うパソコンの準備と関連ソフトのインストール、当該辞書のサイト(ホームページ)と検索プログラムの作成、辞書テキストおよびデータベースの作成の3点に集約されることとなる。

#### パソコンの準備とソフトのインストール

紙媒体で1000頁、4万語程度の辞書でもオンライン版の実用にはPentiumIIのパソコンで十分である。数千円程度で容易に入手できるものだ。メモリーは256M、ハードディスクは10G程度が目安となる。ここにソフトをインストールする。必要なのはOS、エディタ、プログラム言語、ウェブ・サーバ、データベース・サーバなどだ。オンライン版BargeryのOSはRedhat系Linux(Redhat Linux7.3以降)

を順次利用している。Fedora Core 1でも安定して動く。その他のソフトも雑誌付録の無料パッケージに付属しており、手間も費用も全くかからない。これまで各種エディタ、Perl、Apache、PostgreSQLを利用してきた。Bargery程度のデータサイズ(約20M)ではよほど複雑な検索プログラムを設計しない限り検索速度についての不満はほとんど生じないと思われる<sup>3)</sup>。

#### 辞書のサイトと検索プログラムの作成

Bargeryのウェブ・サイトはADSLでつながっている。2カ所にサーバを置いているがいずれもウェブ・サーバとデータベース・サーバを兼ねている。ダイナミックDNSを使い、IPアドレスの変更はcrondに置いたスクリプトで監視している。利用者がサイトへアクセスし、バーをクリックするとフレームで上下に分かれた検索窓が開く。上部のブランクに検索文字列をタイプ、検索方式を指定、ボタンをクリックすると下部に結果が表示される。詳しくはサイトのソースで確認することができる(末尾のドメインを参照)。

検索プログラム作成にあたっては、以前のBargeryの検索システムをもとに、東外大の松下とウィーン大のAhmerの意見を参考にした<sup>4)</sup>。Bargeryはハウサ・英語辞書部(3万7千語)と英・ハウサ語彙集部(4千6百語)からなる。大部分を占めるハウサ・英語辞書の基本的な検索方式として以下のように計画した。1. 紙媒体のように見出し(項目)で引けること、これには検索文字列を含む検索(語頭・語中・語尾)と検索文字列と正確に一致する見出し検索を含む。2. 辞書テキスト全文検索ができること、ヒットした文字列の色と背景色を変えて見出し毎に表示する。3. 英・ハウサ語彙集は例文等がなく語彙も少ないため、検索文字列を含む検索(語頭)のみとする。

こうした計画に沿って文献リストの各書籍掲載のPerlスクリプトを参照しながらCGIを作成した(特に江上2002、末尾のAppendix参照のこと)。

### 辞書テキストの整形とデータベースの作成

辞書テキスト自体は以前のBargeryの順編成テキストファイルをベースとした。仮にファイルを最初から作成しなければならない場合、延べ800時間前後の入力作業が必要になろう<sup>5)</sup>。現在のハウサ語出版物では、いわゆる“glottalized consonants”用の特殊なアルファベットもかなり見られるが、Bargeryはこうした文字が使われる以前に出版されており、フォントに関する制約をほとんど受けない。ただ特殊な発音記号が使われているので通常の英文タイプの記号に置き換えた。こちらの技術・時間・予算的制約のためである。

この作業プロセスでは、テキストファイルを前節で述べた検索計画用に整形し、データベースに読み込ませる。Bargeryではデータベース・サーバとしてPostgreSQLを使っている。順編成テキストファイルはAからZまで紙媒体の辞書全文をファイル化したものと考えてよい。そのままでは読み込めない。検索単位のレコードに切り分け、各レコード内に各種検索専用のインデックスをつくり足し、さらにブラウザ表示のため見出しと本文にhtmlタグを埋め込む必要がある。この一連の作業を整形と言う。検索は見出しとその本文単位で行うが、これを1レコードと呼ぶ。ある見出しとその本文のセットのことを言う。PostgreSQLでは、Bargeryのどの見出しと本文のセットをとっても1レコードに収まるので、以前のようなレコードを小分けする作業を考えなくてよい。まずは、見出しと本文のまとまり毎に改行マークを付加する(切り分けていく)。ミスタイプをチェックしながらの手作業でもよいが4万回ほど改行キーを押す必要があるので、プログラムを組むか、エディタの機能を利用するのが速い。

次に各レコード内に検索専用のインデックス部分をつくる。例えば見出し検索には見出し検索用のインデックスを、全文検索には全文検索用のインデックスをつくり足す。逆引き検索をしたければ専用の逆引き見出しイン

デックスをつくり足せばよい。実際の検索はデータベース上で、入力された文字列をこれらインデックスと照らし合わせることで実行される。そのためインデックス部分から余分な記号やスペースを削除し、インデックスは大文字小文字の区別をなくした大文字だけの連続体となっている。利用者が入力した検索文字列も同様、プログラム上で余分な記号やスペースを削除、大文字に変換される。そして全レコードの当該インデックスと比較され、一致すれば答えを返す。つまり、実際の辞書検索の担い手となるのはインデックスなので、ブラウザに表示される検索結果としての見出しや本文は、紙芝居の絵のようにまさに表示されるためだけにhtmlタグをつけられ、整形された見出しと本文ということになる。なお各レコード内に見出し、本文、見出し検索用インデックス、全文検索用インデックスなどは、タブキーで区切って区別する(表参照)。

整形が終了したら、このテキストファイルをPostgreSQLのデータベースに読み込む。検索計画に従って読み込み用のテーブル(枠のようなもの)をあらかじめ作成しておく。テキストファイルにおけるレコード間の区切り(¥n)とレコード内部の区切り(¥t)と、テーブルの構成とが矛盾しなければ特に問題は起こらない。読み込みにはcopyコマンドを使う(石井2001参照)。

### オンライン版の利用状況

2004年9月現在、オンライン版Bargeryの運用開始から2年近くが過ぎた。当初はハウサ・英語辞書のみであり、6ヶ月ほどをプログラムの改変やバグ対策、英・ハウサ語彙集の作成に費やした。サーバを2カ所に分散し安定的に動くようになって1年少しである。オンライン版運用開始については、ハウサ語関連の研究者が集まるH - hausuと国内のアフリカ言語研究者の組織AFLANGに情報を流した。

商業的に成功しない言語のオンライン辞書という前提があり、ほぼ予想通りアクセスは

表： 見出し検索用インデックスのみを付加したファイルの構造（上）と実際のファイルの一部（下）。

レコード 1=見出し検索用インデックス	見出し	本文	
レコード 2=見出し検索用インデックス	見出し	本文	
レコード 3=見出し検索用インデックス	見出し	本文	
レコード 4=見出し検索用インデックス	見出し	本文	
レコード n=見出し検索用インデックス	見出し	本文	

※正規表現で  $\$n$  は改行、  $\$t$  はタブマークを示す。

```

ABIDO¥t      <b>abido</b>¥t [a"bi"do+] (n.m.). (Yor.) <i>A kind of kola-nut of the</i> hannunruwa <i>variety.</i> ¥n
ABIN¥t<b>abin</b>¥t [a"bi+n]. <i>A thingn of</i> i.e. abu with genitival 'n'. ¥n
ABINCI¥t<b>abinci</b>¥t [a"bi+nci+] (n.m.). <i>Food</i> (lit. a thing of eating). Cf. [a"bi+n ci+] (n.m.). <i>Eating utensil.</i> ¥n
ABINKANIYAKA¥t<b>abin kan iyaka</b>¥t (n.m.). (East Hausa.) <i>Red sorrel</i> (= <b>yakuwa</b> q.v.), so called because at one
time their chief was called Yakuwa, so it was not polite to use his name in reference to the plant. ¥n
ABIYA¥t<b>abiya</b>¥t (S. & Kats.) (n.f.; pl. a"bi"ya+i) = abuya q.v. ¥n
ABKA¥t<b>abka, abko, abku,</b>¥t &c. Vide auka, &c. ¥n
ABKAWA¥t<b>abkawa</b>¥t = aukawa q.v. ¥n
ABKE¥t<b>abke</b>¥t (Kats.) = afe q.v. ¥n
ABKI¥t<b>abki</b> ¥t = auki q.v. ¥n
    
```

多くて1日10カ所程度からである。1カ所からのアクセス回数は多様で、数時間、数日にわたる場合もある。日本国内からのアクセスは、ほぼ皆無と言ってよい。99%が海外からで、中南米を除くすべての地域からアクセスがある。大学や研究機関からのアクセスが多いのも特徴であろう。ハウサ語プログラムをもつBBCからもしばしばアクセスがあり、ハウサ語学習・研究用の辞書としてオンライン版Bargeryが利用されていることがうかがわれる。この1年の間にロンドン大学SOAS（東洋アフリカ研究学院）図書館、UCLAのハウサ語学科、ヨーテボリ大学（スウェーデン）のアフリカ関連ホームページにリンクされた。これらは筆者の全く知らないうちにリンクされたもので、オンライン版Bargeryは世界のいくつかの場所でそれなりに評価されているようである<sup>6)</sup>。

今後の展望 持続可能なオンライン辞書へ向かって

オンライン版Bargeryは2002年1月から開発にとりかかり、7月に初期のプログラムが動いた。以前のBargeryに関する若干の知識と辞書テキストが利用できたため、比較的短期間で運用が可能となったと考える。時間の大半は、辞書テキストの整形とPerlスクリプトに費やした。80年代から残るミスタイプの修正

は現在も継続中である。検索システムの改変についても、他のハウサ語辞書のオンライン化を念頭に置きながら検討中である。著作権に関する課題も当然含まれる。

なお、こうしたオンライン化の試みについては先のウィーン大の辞書以外にもいくつか情報がある。数年来CNRSの研究者らが既存の辞書の電子化ではなくハウサ語の新たな語彙集の構築を模索しており、言語学をはじめとする多様な領域からの相互協力によって新オンライン辞書が生まれる可能性がある。ナイジェリア国内においてもハウサ・スワヒリ・アラビア・英語・フランス語辞書を編纂しようとする動きがかなり以前からあり、経済的な問題のために長年中座状態であるが、オンライン版のコストパフォーマンスとネットワークの加速的な普及が何らかの進展をもたらすかもしれない。

最後にひとつだけ提案しておきたい。今後のオンライン版の充実・普及を考えると、フォントの問題は避けて通れないと思われるが、紙媒体の特殊なアルファベットをそのまま流用する前にオンライン版のカラー表示の有効利用も考えられてよい。柳田國男は色彩によるコントロールについてかつて述べていたが、われわれが目にする世界の多様な文字のユニークさの背後に色彩表現の実質的な制限があったと考えることもできるわけで、カ

ラーディスプレイという極めて今時のメディア表現をオンライン辞書の一部として積極的に使わない手はないと思われる。具体的には、特殊フォントの恩恵に浴ずるのは主にわれわれを含めたハウサ語を母語としない話者であろうから、あえてフォント開発に投資をせずとも(してもよいが)豊富な文字色や背景色によって特殊な子音だけでなく音調をも表現することが十分可能であろう。その場合には色彩に関する心理学的な知見も必要となろう。

オンライン版 Bargery が IT 世界における言語間の情報格差について、特にサハラ以南のアフリカ言語の情報化へ一石を投じることになれば幸いである。

#### 注

- 1) 本稿は、2002年10月開催のAFLANGにおける口頭発表(英語)と2004年6月開催の日本文化人類学会第38回研究大会における口頭発表原稿に加筆修正したものである。
- 2) この一年ほどサーバを2カ所に分散している。現在ナイジェリア国内でこの原稿を書いているが、UPSの能力を超える停電が先日埼玉県内であったらしく、メインサーバの辞書プログラムを再開できないでいる。しかし、利用者はメインサーバのサイトから別のサーバにアクセスすることができる。ナイジェリア北西部からも快適に検索ができています。
- 3) Perlはインタプリタなので処理速度が遅いといしばしば言われる。このため当初は高速化モジュールを利用したが、データサイズゆえか何もしくなくてもストレスを感じないので、現在は停止している。
- 4) 率直で的確なコメントをいただいた松下周二教授とJulia Ahmer氏に感謝申し上げます。
- 5) それでも紙媒体の出版より安価ですむと思われる。
- 6) 辞書をはじめとする出版物の多くはカノ方言を扱い、カノ方言で書かれている。Bargeryのありがたみをしみじみ感じるのはハウサ語の標準語とされたカノ方言の通用地域を離れてからである。例えば西へ500キロ離れたソコト州で話を聞いているとよくわからない言葉と遭遇する。これが本当のハウサ語だと話者は断言するが、英語に堪能な翻訳者が隣にいる機会はありませんのでお互いに不自由を感じながらの会話となる。1万語程度の辞書ではもうどうにもならない。こういう時、

耳慣れない言葉をノートに書きとめておくと後でBargeryが助けてくれる。とするとはからずも、ハウサ語メディア内部における情報格差にもオンライン版Bargeryは何らかの影響を与える可能性があるのではなからうか。これについては別に考えてみようと思う。

#### 文献(一般)

Appadurai, A. 1996 *Modernity at Large: Cultural Dimension of Globalization*. Minneapolis: University of Minnesota Press.

Ginsburg, F. D. et al. (eds.) 2002 *Media Worlds: Anthropology on New Terrain*. Berkeley: University of California Press.

#### (開発関連)

石井達夫 2001 『PostgreSQL完全攻略ガイド 改訂第3版』技術評論社

稲葉香・斉藤浩 2002 「データベースにさわってみよう」『WEB + DBPRESS 8』61 - 74頁

江上秀樹 2002 「習うより慣れよう!今すぐできるPerl」『WEB + DBPRESS 8』75 - 87頁

おもてじゅんいち 2002 「Perlスクリプト入門」『Linux magazine 1月号』169 - 76頁

紙谷歌寿彦 2001 『はじめての人のためのかんたんPerl / CGI入門』秀和システム

Stones, R & Matthew, N. 2002 『PostgreSQL活用テクニック』インプレス

羽生章洋・阪上徹 2001 『はじめてのPostgreSQLプログラミング』ソフトバンク

宮坂雅輝 2002 『CGI / Perl Handbook 2nd edition』ソフトバンク

#### (Bargery電子辞書関連)

塩田勝彦・中村博一 2001 「ハウサ語辞書電子化覚書」『生活科学研究 23』105 - 11頁

中村博一 2002 「PC / AT機におけるハウサ語電子辞書使用試論」『言語と文化 14』65 - 79頁

松下周二 1991 『電子辞書 DICSEARCH』東外大AA研電算室

オンライン版 Bargery サイト :

<http://maguzawa.dyndns.ws>

上記サイトへのリンク :

ロンドン大学 SOAS 図書館

<http://www.soas.ac.uk/library/index.cfm?navid=233>

UCLA ハウサ語学科

<http://www.humnet.ucla.edu/humnet/aflang/Hausa/References/hausalinks.html>

ヨーテボリ大学

<http://www.african.gu.se/webresources/afroasiatic.html>

Appendix: 検索用スクリプト

```
#!/usr/bin/perl

use strict;
use CGI;
use DBI;

print "Content-type: text/html\r\n\r\n";
print "<html><head><title>result</title></head><body bgcolor=#F0E68C>";

(my $item, my $input) = input();
my $key = check($input);
(my $dbh, my $sth) = db_search($item, $key);
result($item, $input, $sth);
$sth->finish(); $dbh->disconnect();

print "</body></html>";

### get info ###
sub input {
    my $query = new CGI;
    my $item = $query->param('item');
    my $input = $query->param('entry');

    return $item, $input;
}

### input check ###
sub check {
    my $input = shift;
    $input =~ s/[^a-zA-Z]//g;
    $input = uc $input;

    if ($input =~ /^$/) {print "<font size=4 color=#FF0000>Type some,
please!</font><br>";
    exit;
}
```

```

    } else {return $input;
    }
}
### word search ###
sub db_search {
my $item = shift;
my $key = shift;
my $sth;
my $dbh = DBI->connect("dbi:Pg:dbname=bunk"
                      , "jinka", "jinka") or die "unable to connect to DB.
$DBI::errstr";

if ($item eq "word") {
    $sth = $dbh->prepare("select * from kamu3 where entry like ? order by entry");
    $sth->execute("$key");
}
} elsif ($item eq "start") {
    $sth = $dbh->prepare("select * from kamu3 where entry like ? order by entry");
    $sth->execute("$key%");
}
} elsif ($item eq "part") {
    $sth = $dbh->prepare("select * from kamu3 where entry like ? order by entry");
    $sth->execute("%$key%");
}
} elsif ($item eq "context") {
    $sth = $dbh->prepare("select * from kamu3 where con like ? order by entry");
    $sth->execute("%$key%");
}
} elsif ($item eq "eng") {
    $sth = $dbh->prepare("select * from eng where entry like ? order by entry");
    $sth->execute("$key%");
}
}

return $dbh, $sth;
}

```

```

### result disp ###
sub result {

    my $item = shift;
    my $input = shift;
    my $sth = shift;

    if ($sth->rows == 0) {
        print "<b><font size=4 color=#FF0000>Sorry, no matching key
found!</font></b>";
    } else {
        while(my @row = $sth->fetchrow_array()) { if ($item eq "context") {

$row[2]          =          s($input)(<font          color="#CC0000"
style="background-color:#FFCCCC">$input</font>)gi;

            print ($row[1], "<br>", $row[2], "<br><br>"), "\n";
                } else {
                    print ($row[1], "<br>", $row[2], "<br><br>"), "\n";
                }
            }
        }
    }
}

```