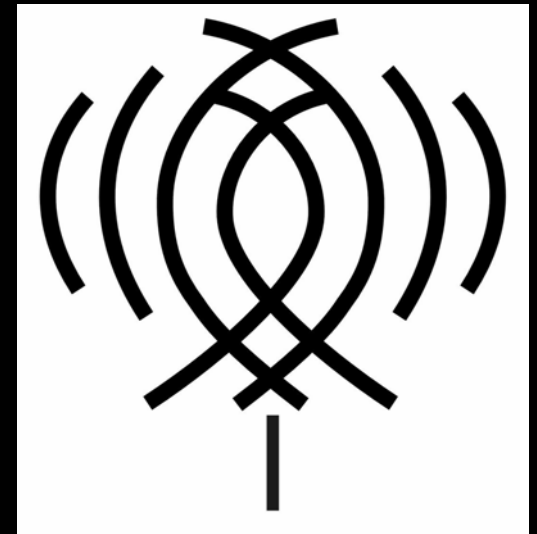


# Paper Dictionary & Its Virtual Version

**Making a Traditional  
Dictionary into an Electronic  
Lexicon**

Seminar on  
Localization and  
Human Technologies  
21-23 Nov, 2001 Bangalore



**Udaya Narayana Singh**  
CIIL, Mysore

# 1. Paper Dictionaries

- 'Easy to use' theory
  - ◆ **Easy to browse** – with a flip, pages will open, and alphabetically, entries will flow on
  - ◆ **Easy to procure** – in terms of distribution & availability
  - ◆ **Easy to buy** – cheaper than usual reference books, and more the popularity, the greater are chances of a cheaper and localized (say, Asian?) edition or a paper back version
  - ◆ **Easy to replicate** – In different sizes (royal/demy/crown, etc) and word-volumes (10,000 most frequent words, 30,000 word dic, or large comprehensive dictionaries,etc)
  - ◆ **Easy to use** – all you require is the power of vision, a definite search requirement, a knowledge of order of alphabets, and an idea of arrangements of sub-entries and information under a lemma
  - ◆ **Easy to apply** – does not make any assumptions about users' knowledge, and hence makes everything (including most obvious grammatical descriptions) explicit

# Creating Paper Lexicons

## Oxford English Dictionary

**dygogram** ('dægəʊgrəm). [See quot. 1862.] A diagram consisting of a curve showing the variation of the horizontal component of the force of magnetism exerted upon the ship's compass-needle by the iron in the ship's composition while making a circuit or curve.  
1862 A. SMITH *Deviat. Ship's Compass* App. ii. 151 Graphic representations of the amount and direction of the forces which act on the magnetic needle. These representations are designated by the general name of 'Dygo-gram', a contraction for 'Dynamo-gonio-gram' or 'Force and angle diagram'. 1881 MAXWELL *Electr. & Magn.* II. 73 Such a curve, by means of which the direction and magnitude of the force on the compass is given in terms of the magnetic course of the ship, is called a Dygogram.

From print to online entry

### 'Easy to make' theory

- **Established tradition** in most knowledge societies
- **Strong**, and by-now, Standard lexicographical training
- Once made, revisions are **less cumbersome**
- Availability of specialized manpower, and **division of labour**
- **Begins with field data** and working glossaries
- **Expanded incrementally**
- **A focussed approach** possible, and hence different kinds of lexicons

# More advantages

- 'Easy to sell' theory
- **Good business** proposition for publishers world over
- **Everyone** – every educated person **needs** a dictionary
- One feels like having one **in addition** - even when one has an electronic dictionary
- Once made, **revision costs are minimum**
- **Initial development cost** is often **unpaid** by publishers as they come from funding bodies and societies
- Easy to skirt the **IPR issue** as it becomes a branded product

## Oxford English Dictionary

wimple, *n.*

SECOND EDITION  
1989

[PRONUNCIATION](#)

[SPELLINGS](#)

[ETYMOLOGY](#)

[QUOTATIONS](#)

[DATE CHART](#)

**I. 1.** A garment of linen or silk formerly worn by women, so folded as to envelop the head, chin, sides of the face, and neck; now retained in the dress of nuns. Also *gen.* a veil.

Used loosely in early glossaries as a rendering of L. *anabola*, *cyclas*, *peplum*, *ricinum*.

**a1100** *Aldhelm Gloss.* l. 4296 (Napier 112) *Cyclade*, *i. ueste*, wimple.  
**a1100** *Gloss.* in Wr.-Wülcker 107/37 *Ricinum*, winpel uel orl. *Ibid.* 125/8 *Anabola*, winpel. **c1200** *Trin. Coll. Hom.* 163 Hire winpel wit oðer maked zeleu mid saffran. **c1240** *Angr. R.* 420 (MS. C), Sum seið þæt hit limpeð to

# 2. Basic Disadvantages

- Slim ones are not comprehensive
- Comprehensive ones are bulky
- Bulky ones tear off easily
- Weight makes them difficult to handle, even if binding lasts long
- Difficult to make it more than bilingual – almost impossible beyond trilingual
- Complicated entries become invariably longer – consequently difficult to use
- Addition of contexts is ideal but expensive – to create/buy/refer to
- Spelling variation usually not taken care of
- Further, one must know the exact spelling to refer to
- One must have also mastered the pattern to know if a search is inside a head-entry as a sub- or sub-sub-entry
- Once published, a paper lexicon becomes dated whereas the concerned language, as we know, is ever-evolving
- Size restrictions hamper their coverage

### 3. Problems with Paper Lexicons in an Electronic Age



- Working on two modes at a time difficult
- Working bi-modally is also time consuming & clumsy
- For translators, and users of bilingual resources, reading orientation is not easy – traveling from one system to another
- Big difference in terms of space, color, shade, light, background & type-styles
- For each purpose, a different paper lexicon is needed
- Often several are used as same words are described differently in diff. lexicons
- Since paper lexicons are not necessarily culled out of texts/corpora/data/records, co-referencing is limited

# 4. Converting to E-lexicon: The Claim of being user-friendly

The screenshot displays the Oxford English Dictionary interface for the word 'wimple'. Annotations with blue lines point to various elements:

- home page link:** Points to the 'Oxford English Dictionary' logo.
- back to the results list:** Points to the '< RESULTS' button.
- entry name:** Points to the word 'wimple, n.'.
- entry display buttons:** Points to the navigation tabs: PRONUNCIATION, SPELLINGS, ETYMOLOGY, QUOTATIONS, and DATE CHART.
- find word:** Points to the search input field containing 'wimple' and the 'FIND WORD' button.
- entry version:** Points to the text 'SECOND EDITION 1989'.
- list by entry:** Points to the 'LIST BY ENTRY' button in the bottom navigation bar.
- feature buttons:** Points to the bottom navigation bar containing 'SEARCH', 'LIST BY ENTRY', 'LIST BY DATE', and 'ENTRY MAP'.
- entry text:** Points to the main definition and etymology text of the entry.
- support buttons:** Points to the 'PRINT', 'MAIL', 'HELP', and 'SIGN OUT' buttons.

The main content of the entry includes:

**wimple, n.** SECOND EDITION 1989

PRONUNCIATION | SPELLINGS | ETYMOLOGY | QUOTATIONS | DATE CHART

**I. 1.** A garment of linen or silk formerly worn by women, so folded as to envelop the head, chin, sides of the face, and neck: now retained in the dress of nuns. Also *gen.* a veil.

Used loosely in early glossaries as a rendering of L. *anabola*, *cyclas*, *peplum*, *ricinum*.

*a*1100 *Aldhelm Gloss.* 1. 4296 (Napier 112) *Cyclade*, .i. *ueste*, wimple. *a*1100 *Gloss.* in *Wr.-Wülcker* 107/37 *Ricinum*, *winpel uel* orl. *Ibid.* 125/8 *Anabola*, *winpel*. *c*1200 *Trin. Coll. Hom.* 163 Hire *winpel* wit oðer maked 3eleu mid saffran. *c*1240 *Ancr. R.* 420 (MS. C), Sum seið þæt hit limpeð to

SEARCH | LIST BY ENTRY | LIST BY DATE | ENTRY MAP | PRINT | MAIL | HELP | SIGN OUT

SUBSCRIBER: Guest

# 5. Distinct advantages

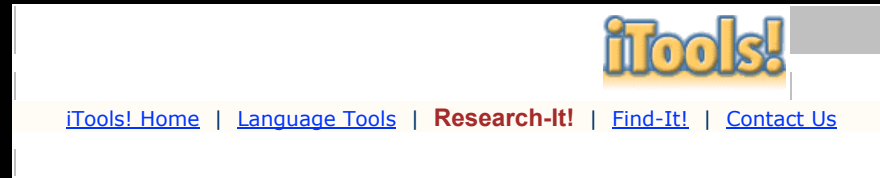
- Many volumes with add-on parts get compressed into one e-lexicon
- Time taken to create truly large dictionaries, such as OED (=44 years, 1884-1928) with its supplementaries (1933 & 1972-86) get converted into e-mode in 10 years (1990-2000)
- 34 million pound sterling spent on e-conversion. But once it is created, updating is hassle-free.
- Storage and retrieval become quick and easy

# 6. More plus-points

- ⑩ **Watch out the volume, too: Search zillions of information with the click of a mouse**
- ⑩ **Use it as a word-finder (= meaning known but the word forgotten)**
- ⑩ **Find all foreign words & expressions**
- ⑩ **Find dialectal variations**
- ⑩ **Find all instances of classical borrowing**
- ⑩ **Find collocations and contexts**
- ⑩ **Find out the etymon**
- ⑩ **Search for quotations from a large bank**
- ⑩ **Search for contemporary use from on-line texts**
- ⑩ **Search for all usages -- author-wise, genre-wise and age-wise**
- ⑩ **Use it as a companion volume to social & literary history**
- ⑩ **Display entries according to your needs – in detail or in brief**
- ⑩ **Turn on and off pronunciation and variant spellings**
- ⑩ **Gain access to new and revised words every month/quarter**
- ⑩ **One need not wait for the next edition for updation**

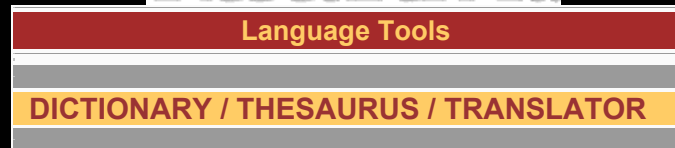
# 7. Versatility Of E-lexicons:

A help,  
Or a hindrance?



iTools!  
[iTools! Home](#) | [Language Tools](#) | **Research-It!** | [Find-It!](#) | [Contact Us](#)

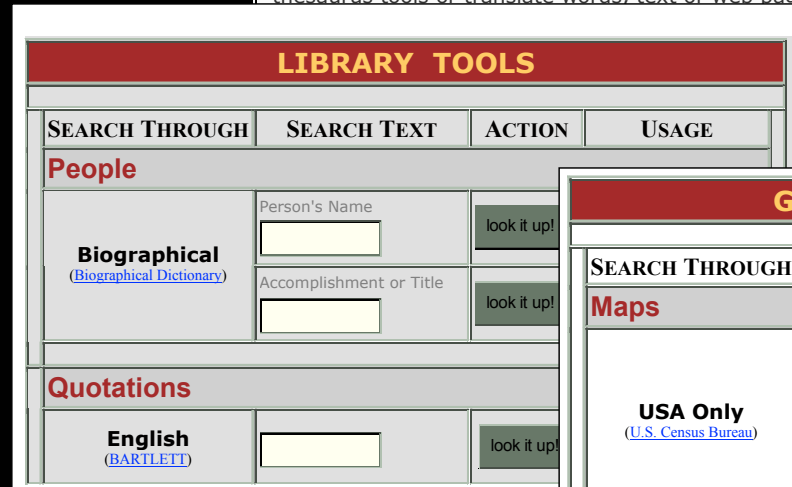
## Research-It!



Language Tools  
DICTIONARY / THESAURUS / TRANSLATOR

### [Language Tools](#)

Look up words with general and specialized dictionary and thesaurus tools or translate words, text or web pages into

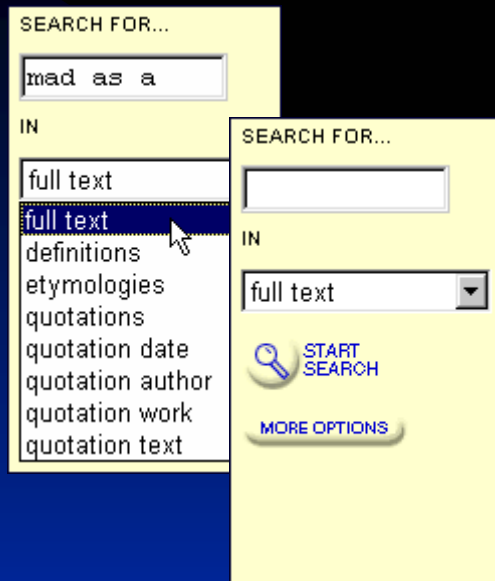


LIBRARY TOOLS			
SEARCH THROUGH	SEARCH TEXT	ACTION	USAGE
<b>People</b>			
<b>Biographical</b> <a href="#">(Biographical Dictionary)</a>	Person's Name <input type="text"/>	look it up!	
	Accomplishment or Title <input type="text"/>	look it up!	
<b>Quotations</b>			
<b>English</b> <a href="#">(BARTLETT)</a>	<input type="text"/>	look it up!	



GEOGRAPHICAL TOOLS		
SEARCH THROUGH	SEARCH TEXT	ACTION
<b>Maps</b>		
<b>USA Only</b> <a href="#">(U.S. Census Bureau)</a>	City <input type="text"/> State <input type="text"/>	
	OR ZipCode <input type="text"/>	
	Street Address <input type="text"/> <small>(e.g., 600 Mountain Ave, or Mountain Ave &amp; South St, or just Mountain Ave)</small>	

# 8. Other qualities – The plural search options



The highlighted matches can  
Take us straight to their  
occurrence in an entry  
or to a Phrase anywhere.

Operated by clicking on **Search...** at the bottom of the screen, and this lets one search the full text of the Dictionary.

In addition, it gives one the choice of searching for

- definitions,
- ★ • etymologies,
- ★ • proximate expressions
- ★ • word associations
- ★ • phrasal combinations
- ★ • collocations and compounds
- ★ • synonyms
- ★ \* antonymous expressions
- ★ • primary, secondary and associative meanings
- ★ • quotations, or to the
- ★ • default option of 'full text' to any of these text
- areas.

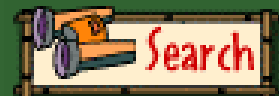
# 9. The multiple display options

SEARCH / LIST BY ENTRY / LIST BY DATE / ENTRY MAP

## Manipulable display

- List entries by date
- List by most frequent occurrences
- List by providing many help buttons, such as
  - \* Pronunciation
  - \* Etymology
  - \* Spelling options
  - \* Variants
  - \* Textual Occurrences
  - \* Date of appearance

- **Combine several dictionaries in one**
  - # by intelligently creating displays for enlarged or abridged versions
  - # by merging a dictionary & a thesaurus
  - # by merging it with a pronunciation book
  - # by merging with a grammar manual
  - # by appending it to a style manual
  - # by making it available along with a word processor
  - # by tagging it along with a language accessor or a translation tool



## Virtual Bangladesh : English to Bengali Dictionary

Enter an English Word *[all lower case]*:

Trans\_late

Clear\_

[Display all the entries](#)

amazon.com.

Trilingual  
Dictionary  
Bidhu B.  
Dasgupta

[Privacy Information](#)



[Back to Survival Bengali](#)

### 10. The Disadvantages –

**Limited Purpose**

**Over-simplified**

**Not robust enough**

**Lack of innovativeness**

**Extension of touristic designs**

# Apte Sanskrit Dictionary Search

This is a Web Sanskrit Dictionary based on "The Practical Sanskrit-English Dictionary" of Vaman Shivaram Apte. And it contains only the first word (or phrase in some case) of each numbered meaning.

Input transliteration scheme is shown in the table below.

A Verb should be searched by its root form. A noun should be searched by its stem form, i.e. `deva', `lakSmii', `aatman' etc. but feminin stem `aa' derived from `a' should be searched by dropping last `a' i.e. `kaniSThaa' appears in the item `kaniSTha', but `kathaa' appears under itself. If feminin form takes `ii' stem, it comes under different head word, i.e. `kaniina' and `kaniinii' are two items. Compound words (-Comp.) are not taken.

अ <sub>a</sub>	आ <sub>aa</sub>	इ <sub>i</sub>	ई <sub>ii</sub>	उ <sub>u</sub>	ऊ <sub>uu</sub>	ऋ <sub>R</sub>	ॠ <sub>RR</sub>
ऌ <sub>L</sub>	ॡ <sub>LL</sub>	ए <sub>e</sub>	ऐ <sub>ai</sub>	ओ <sub>o</sub>	औ <sub>au</sub>	अं <sub>aM</sub>	अः <sub>aH</sub>

VOWELS

क् <sub>k</sub>	ख <sub>kh</sub>	ग <sub>g</sub>	घ <sub>gh</sub>	ङ <sub>G</sub>
च् <sub>c</sub>	छ <sub>ch</sub>	ज <sub>j</sub>	झ <sub>jh</sub>	ञ <sub>J</sub>
ट <sub>T</sub>	ठ <sub>Th</sub>	ड <sub>D</sub>	ढ <sub>Dh</sub>	ण <sub>N</sub>
त् <sub>t</sub>	थ <sub>th</sub>	द <sub>d</sub>	ध <sub>dh</sub>	न <sub>n</sub>
प् <sub>p</sub>	फ <sub>ph</sub>	ब <sub>b</sub>	भ <sub>bh</sub>	म् <sub>m</sub>
य <sub>v</sub>	र <sub>r</sub>	ल् <sub>l</sub>	व <sub>v</sub>	

CONSONANTS

[tjun@aa.tufs.ac.jp](mailto:tjun@aa.tufs.ac.jp)

# Capeller's Sanskrit-English Dictionary

**Sanskrit :**

**English :**

**Maximum-Output:**

At present the digital version of Cappeller's Sanskrit Dictionary (1891) contains approx. 50.000 main entries. You can search for one Sanskrit main entry in the dictionary under Sanskrit or for a translation into Sanskrit under English . The transliteration is based on the Harvard-Kyoto (HK) convention as follows:

a A i l u U R RR IR IRR e ai o au M H  
k kh g gh G c ch j jh J  
T Th D Dh N t th d dh n  
p ph b bh m y r l v z S s h

Note: WAIS search is not case sensitive.

See: [Report on the Cologne Sanskrit Dictionary Project](#)

Suggestions and comments to: [IITS-lexicon@uni-koeln.de](mailto:IITS-lexicon@uni-koeln.de)



**Webmasters:**  
[A. Zeini and](#)  
[T. Grote-Beverborg.](#)

# 11. Other problems

- **Some agencies using e-lexicons as potential money-churners**
- **Some spend much less on R & D, and more on promo**
- **Some are quite nascent in design as well as application**
- **Often not compatible with MAT tools**
- **Relating a system to auto-taggers or bigger devices which work on daily on-line dumps usually not done**
- **Often not available in public domain or ones that are available aren't large enough to be of good use**
- **For Indian languages, script support is still a problem for localization of any robust methodology**
- **Even if scriptal problems are resolved on CD-versions of such e-lexicons, the web-versions are invariably in problems – at least in some browsers; let's look at Gerhard Huet's e-dictionary of Sanskrit-Francaise:**

Search

Search for an entry matching an initial pattern:  
(Transliterate for ā, for ñ, for ñ, for ṭ, for ś, for  
Or go to the section starting with an initial letter:

a ā i ī u ū r ṛ l ḷ e ai o au

k k g g ñ  
h h h

c c j jh ñ  
h h

t th d d n  
h h

t th d d n  
h h

p p b b m  
h h

y r l v ś ṣ s h

## Dictionnaire sanscrit-français

Gérard Huet

Version 170 (2001-10-19)

Offline printable version may be downloaded  
from [here](#)

© Gérard Huet 2001

# 12. Where do we go from here?

- **Priority 1:** Solve the Indian languages script problem on both UNIX and Windows environments. With Linux localization done by NCST, let's tackle the other.
- **Priority 2:** Let's convert all available lexical resources into E-lexicons, with of course, necessary editing and add-on features.
- **Priority 3:** Enhance Indian languages corpora – add voice corpora as well as efficient tagging devices.
- **Priority 4:** Quick look-ups for translators could be made ready even before fashionable products are launched, purely to enable translators work on-line.
- **Priority 5:** Set up large-scale longer-term institutions or instruments for lexicographical work which would be authoritative, comprehensive, multi-utility, and also constantly up-dating.
- **Priority 6:** Design things in a manner so as to make the lexical resources useful for more difficult products such as MAT systems, etc.
- **Priority 7:** Link up as many Indian languages pairs as possible, and if possible, use trilingual formats, by involving both English and Hindi.