



The LEGO Project

Lexicon Enhancement via the GOLD Ontology

Brent Miller, The LINGUIST List



Overview

- Introduction
- Doing LEGO
- Current Status
- Future of LEGO



Introduction

LEGO and the Need for Interoperability



A Variety of Data

- Standards
 - LIFT
 - LMF
 - TEI
- File Formats
 - PDF
 - Excel/Access
 - MDF (Toolbox)
 - .doc/.odt (Word/OpenOffice)



Why Interoperate?

- Greater access to language data
- More intelligent searches
- Ease of comparison between lexicons



What is LEGO?

- Three-year project sponsored by the NSF
- Participants: LINGUIST List, University at Buffalo
- Goal: Create a datanet of interoperable lexicons
 - Map grammatical information to GOLD
 - Map structure to a common schema (LL-LIFT)
 - Output in XML where lexicon contributor allows
 - Preserve source's integrity



LEGO's Purpose

- Not intended to develop a lexicon creation or display tool
- Will support multi-lexicon searches and comparisons
- Will demonstrate the value of digital standards in linguistic research



Doing LEGO

Team Structure and Workflow

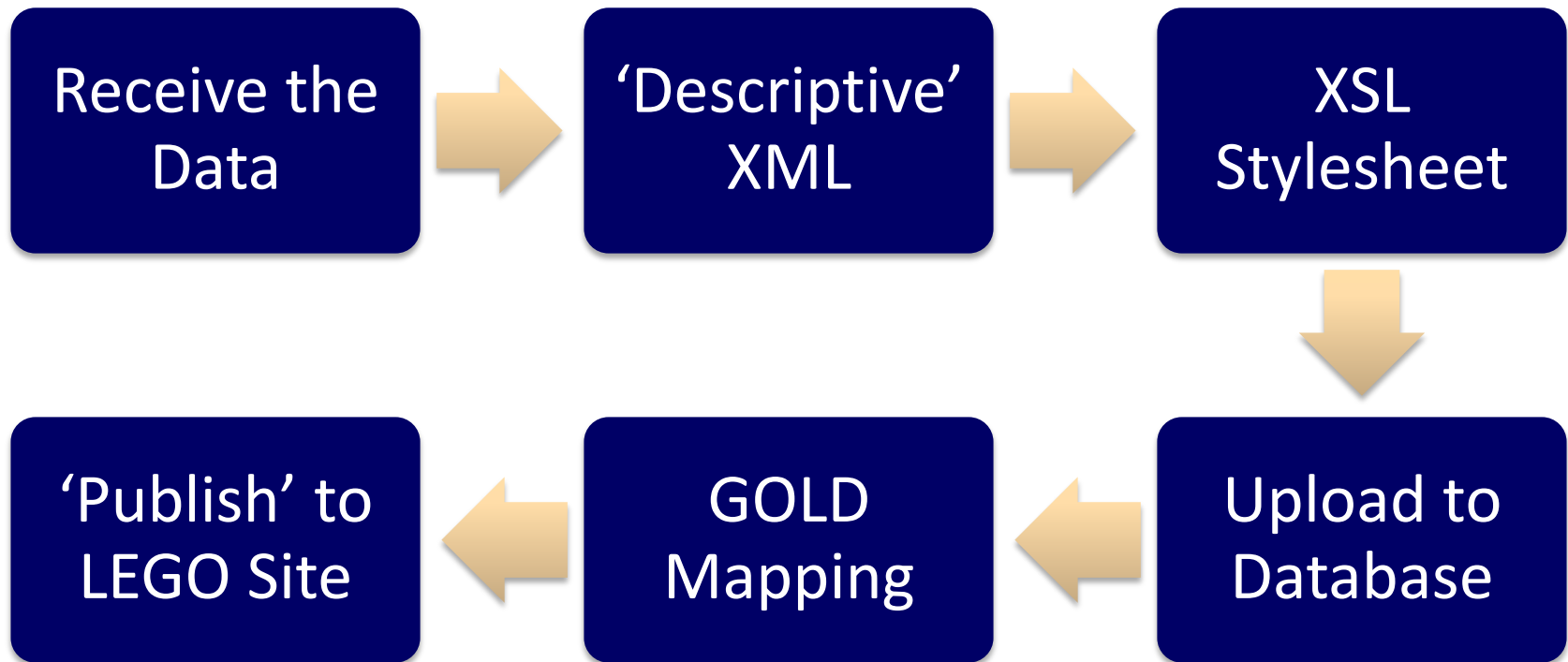


Team Structure

- Three principle investigators
 - Jeff Good, University at Buffalo
 - Helen Aristar-Dry and Anthony Aristar, Eastern Michigan University
- Three graduate students
 - Brent Miller, Justin Petro, Erica Wicks
- One undergraduate, Lili Xia
- One programmer, Lily Zheng



Workflow





Current Status

Our Data, Website, and Faceted Search



Lexical Data

- Completed
 - 11 wordlists (10 Qiang dialects, Sáliba)
 - 6 lexicons (Western Sisaala, Potawatomi, Ibibio, Wichita, Tuva, Shoshone)
- 11 nearing completion (Fulfulde, Archi, Udi, Mocoví, Jarawara, Nhirrpi, Titan, Maa, Mbodomo, Western Pantar, Mocho', Buffalo Wordlists)



The LEGO Site

- Homepage (in development)
 - <http://lego.linguistlist.org>
- Browse lexicons
 - Each lexicon has a homepage
- Browse entries
 - Each entry has its own page
- Faceted search
 - Allows for fine-grained GOLD-aware searches of morphological information across lexicons

Lexicon Enhancement via the GOLD Ontology

Lexicon Enhancement via the GOLD Ontology (LEGO) is a project funded by the National Science Foundation (BCS-0753321) to establish tools and standards to facilitate the sharing and interoperation of lexical data. It is implemented jointly by The LINGUIST List at Eastern Michigan University and The University at Buffalo. LEGO activities include conversion of legacy lexical resources, evaluation of existing standards for encoding lexical data and, where needed, proposing new standards, and providing interfaces through which individuals can search the LEGO datanet of lexical resources and upload their own lexical data.

Disclaimer: As an infrastructure project, we are primarily interested in allowing existing lexical data to be included in our datanet and promoting standards to allow others to construct comparable datanets, if they so wish. In order to develop this infrastructure, we have converted a number of legacy resources which have allowed us to test our approach, but we have not engaged in collecting new lexical data, nor have we tried augment the information found in these materials beyond what has been needed to achieve interoperation.

[Read more...](#)

All data on the LEGO site is, by default, covered under the following Creative Commons license: Attribution-NonCommercial-ShareAlike (BY-NC-SA). If a lexicon contributor prefers a different license, their license will take precedence.



Language (ISO 639-3 code)

Lexicon (source)

Xuecheng (cng)		View LEGO Lexicon
Goukou (cng)		View LEGO Lexicon
Mawoma (cng)		View LEGO Lexicon
Luoxiang (cng)		View LEGO Lexicon
Sáliba (sbe)		View LEGO Lexicon
Western Sisaala (ssl)	Source: Moran, Steven. 2008. A Grammatical Sketch of Western Sisaala. Saarbrücken, Germany: Verlag Dr. Müller (VDM).	View LEGO Lexicon
Shoshone (ssh)		View LEGO Lexicon
Potawatomi (pot)		View LEGO Lexicon
Huilong (cng)		View LEGO Lexicon
Udi (udi)		View LEGO Lexicon
Luhua (cng)		View LEGO Lexicon
Ibibio (ibb)		View LEGO Lexicon
Wichita (wic)		View LEGO Lexicon
Tuva (tyv)		View LEGO Lexicon
Taoping (qxs)		View LEGO Lexicon
Wabo (cng)		View LEGO Lexicon
Weicheng (cng)		View LEGO Lexicon
Weigu (cng)		View LEGO Lexicon

Ibibio

This lexicon was created in August 2010 by Amy Brunett while working with Mfon Udoinyang. The lexicon focuses on two semantic fields: food items and parts of the body.

Enter

This lexicon was created in August 2010 by Amy Brunett while working with Mfon Udoinyang. The methodology of the conducted fieldwork consisted of first doing a literature review of the language status and general linguistic features. The primary sources used included A Grammar of the Ibibio Language (1990) by Okon Essien and Ibibio Phonetics and Phonology by Eno-Abasi E. Urua. The second step was to create an elicitation plan. In correspondence with Dr. Dafydd Gibbon, Dr. Gibbon suggested "concentrate on a small, coherent, practical, concrete, relatively technical semantic field." Thus, the author decided to focus on two semantic fields: food items and parts of the body.

When beginning elicitation, the first step was to elicit for all three persons and determine whether the language made an inalienable/alienable distinction. Fieldworks Language Explorer (FLEX) was the original program used to input data after each elicitation session. The author includes tones as accents over the vowels since they are phonemic in Ibibio.

[More on the LEGO version](#) [Browse the LEGO version](#)

Browse Western Sisaala Lexicon (LEGO Version)

Source: Moran, Steven. 2008. A Grammatical Sketch of Western Sisaala. Saarbürcken, Germany: Verlag Dr. Müller (VDM).

a b d e é f g h i j k l m n o p s t u v w y z ɔ ɛ ɪ ɲ ɔ ʒ ɖ ʒ ɛ

Form (alphabetic)	Grammatical Info	Definition
a	PRO, 1, PL	we
aba	CONJ	and
abrobe	N	pineapple
ahyi	ADV	no
akutu	SG, N	orange (fruit)
akutuwa	PL, N	oranges (fruit)
alonperri	N	airplane
amo	SG, PL, N	tomato; tomatoes
ana	N	father
anawa	N	fathers
anbiie	N, CMPD	paternal uncle
anballe	CMPD, N	paternal uncles

Headword: admobinkéwgemek

Author Token(s): Noun, Singular, Inanimate

Gold Concept(s): [Noun](#), [Singular Number](#), [Inanimate Gender](#)

Definition:

English: car dealer, repair garage

Paradigmatic Variant : admobinkégemgwén

Grammatical Information: Noun, Singular, Plural, Inanimate

Paradigmatic Variant : admobinkégemgok

Grammatical Information: Noun, Singular, SpatialCase, Inanimate



Faceted Search

- Choose lexicons
- Text search
 - Search across forms, variants, glosses, definitions, etymology, examples, notes
 - Displays keyword in context
- Filters
 - Easily added/removed
 - Narrow search in real time

Choose which lexicons you'd like to search across by clicking the boxes next to the lexicon name, then clicking "Next". If you click "Select All", you will search across all the lexicons. "Next" will bring you to the next page of the search where you can enter text and filter your search results.

Select All

Clear All

Next

Language Name**Type**

<input type="checkbox"/> Potawatomi	Lexicon
<input type="checkbox"/> Shoshone	Lexicon
<input type="checkbox"/> Tuva	Lexicon
<input type="checkbox"/> Taoping	Wordlist
<input type="checkbox"/> Xuecheng	Wordlist
<input type="checkbox"/> Weicheng	Wordlist
<input type="checkbox"/> Weigu	Wordlist
<input type="checkbox"/> Luoxiang	Wordlist
<input type="checkbox"/> Luhua	Wordlist
<input type="checkbox"/> Wabo	Wordlist
<input type="checkbox"/> Mawoma	Wordlist
<input type="checkbox"/> Goukou	Wordlist
<input type="checkbox"/> Huilong	Wordlist
<input type="checkbox"/> Wichita	Lexicon
<input type="checkbox"/> Western Sisaala	Lexicon
<input type="checkbox"/> Sáliba	Wordlist
<input type="checkbox"/> Ibibio	Lexicon
<input type="checkbox"/> Udi	Lexicon

Select All

Clear All

Next

Choose which lexicons you'd like to search across by clicking the boxes next to the lexicon name, then clicking "Next". If you click "Select All", you will search across all the lexicons. "Next" will bring you to the next page of the search where you can enter text and filter your search results.

Select All Clear All Next

Language Name

Type

<input checked="" type="checkbox"/> Potawatomi	Lexicon
<input checked="" type="checkbox"/> Shoshone	Lexicon
<input type="checkbox"/> Tuva	Lexicon
<input type="checkbox"/> Taoping	Wordlist
<input type="checkbox"/> Xuecheng	Wordlist
<input type="checkbox"/> Weicheng	Wordlist
<input type="checkbox"/> Weigu	Wordlist
<input type="checkbox"/> Luoxiang	Wordlist
<input type="checkbox"/> Luhua	Wordlist
<input type="checkbox"/> Wabo	Wordlist
<input type="checkbox"/> Mawoma	Wordlist
<input type="checkbox"/> Goukou	Wordlist
<input type="checkbox"/> Huilong	Wordlist
<input type="checkbox"/> Wichita	Lexicon
<input checked="" type="checkbox"/> Western Sisaala	Lexicon
<input type="checkbox"/> Sálíba	Wordlist
<input type="checkbox"/> Ibibio	Lexicon
<input type="checkbox"/> Udi	Lexicon

Select All Clear All Next

Lexicon Lists:

Lexicon: Potawatomi

Lexicon: Shoshone

Lexicon: Western Sisaala

[Reselect lexicons](#)**Currently Used Filters:**

None

Narrow Results By:

Lexicon Types:	Entries
Lexicon	22247

GOLD Labels:

verbal	View in GOLD	7262
noun	View in GOLD	5685
animate gender	View in GOLD	5384
singular number	View in GOLD	5225
View more		

Author Labels:

verb	6621
animate	5384
singular	5071
noun	4677
View more	

Code of Subject Language:

pot	About...	11768
ssh	About...	9375
ssl	About...	1104

Code of Gloss Language:

eng	About...	22247
---------------------	--------------------------	-------

Note types:

bibliographic	9375
dialect	4898

Full Text Search: IN [All Fields](#) [Submit](#)

Use an asterisk (*) for partial searches, e.g. ba* for bat, baby, bamboozle, etc.

Click items to the left to filter results.

Here you can search across the lexicons contained in the LEGO database. There are two ways to search:

1. Using the search box above allows you to perform a full-text search over the entry headwords, variants, glosses/definitions, examples, and notes. Use an asterisk (*) to return all entries.
2. You can filter the set of entries returned by the full-text search by selecting filters on the left. All the filters you add will appear in the "Currently Used Filters" box, and can be removed by clicking the "Remove Filter" link.

Note: You can use the filters alone to filter all entries in the database. However, if you wish to make a text search, you must do so before you use any filters. Text searches override the filters, but not vice versa.

Lexicon Lists:

Lexicon: Potawatomi

Lexicon: Shoshone

Lexicon: Western Sisaala

[Reselect lexicons](#)

Currently Used Filters:

None

Narrow Results By:

Lexicon Types:	Entries
Lexicon	22247
GOLD Labels:	
verbal	View in GOLD 7262
noun	View in GOLD 5685
animate gender	View in GOLD 5384
singular number	View in GOLD 5225
View more	
Author Labels:	
verb	6621
animate	5384
singular	5071
noun	4677
View more	
Code of Subject Language:	
pot	About... 11768
ssh	About... 9375
ssl	About... 1104
Code of Gloss Language:	
eng	About... 22247
Note types:	
bibliographic	9375
dialect	4898

Full Text Search:

IN

Use an asterisk (*) for partial searches, e.g. ba* for ba
Click items to the left to filter results.

All Fields
Form
Variants
Gloss/Definitions
Etymology
Examples
Example Translations
Notes

Here you can search across the lexicons contained in the database. There are two ways to search:

1. Using the search box above allows you to perform a full-text search over the entry headwords, variants, glosses/definitions, examples, and notes. Use an asterisk (*) to return all entries.

2. You can filter the set of entries returned by the full-text search by selecting filters on the left. All the filters you add will appear in the "Currently Used Filters" box, and can be removed by clicking the "Remove Filter" link.

Note: You can use the filters alone to filter all entries in the database. However, if you wish to make a text search, you must do so before you use any filters. Text searches override the filters, but not vice versa.

Lexicon Lists:

Lexicon: Potawatomi
 Lexicon: Shoshone
 Lexicon: Western Sisaala

[Reselect lexicons](#)
Currently Used Filters:

Query: translation:water* [Remove](#)

Narrow Results By:

Lexicon Types:	Entries
Lexicon	149

GOLD Labels:

common noun	View in GOLD	35
verbal	View in GOLD	29
noun	View in GOLD	22
inanimate gender	View in GOLD	19
View more		

Author Labels:

n	46
verb	26
inanimate	19
animate	18
View more	

Code of Subject Language:

ssh	About...	95
pot	About...	42
ssl	About...	12

Code of Gloss Language:

eng	About...	149
---------------------	--------------------------	-----

Note types:

bibliographic	95
dialect	52

Full Text Search: IN

Use an asterisk (*) for partial searches, e.g. ba* for bat, baby, bamboozle, etc.
 Click items to the left to filter results.

Results 1 - 100 of 149 :

Previous 1 [2](#) [Next](#)

Lexical Item**Lexicon****Context**
[bashkjézhék](#)

Potawatomi (Welcher, 2003)

 cut s.t. open (e.g. a **watermelon**)

[bijbyék](#)

Potawatomi (Welcher, 2003)

 in the **water**
[bmojgen](#)

Potawatomi (Welcher, 2003)

 yoke for **water**, packstrap

[dékaboyak](#)

Potawatomi (Welcher, 2003)

 cold **water**, cold drinks

[ban'disip](#)

Shoshone (Di Paolo, 2010)

 bug, giant, **water**
[dkabo-dgojgen](#)

Potawatomi (Welcher, 2003)

 baking powder (cold **water** ingredient)

[dkemit](#)

Potawatomi (Welcher, 2003)

 go across **water**
[dkep](#)

Potawatomi (Welcher, 2003)

 spring (**water** source)

[dobie: dobie](#)

Western Sisaala (Moran, 2003)

 drops of **water**
[gonabo](#)

Potawatomi (Welcher, 2003)

 melt-**water** from snow

[Gété-gteganek](#)

Potawatomi (Welcher, 2003)

Watersmeet, Michigan

[gwabtót](#)

Potawatomi (Welcher, 2003)

 run out of the **water**
[gwa'ek](#)

Potawatomi (Welcher, 2003)

 take s.t. out of fire or **water** (with somet

[gwashek](#)

Potawatomi (Welcher, 2003)

 come out of the **water**
[gwatot](#)

Potawatomi (Welcher, 2003)

 take s.t. out of fire or **water**
[gwidék](#)

Potawatomi (Welcher, 2003)

 float, be in the **water**, be submerged

[gwidot](#)

Potawatomi (Welcher, 2003)

 soak s.t., put s.t. in the **water**
[gwijek](#)

Potawatomi (Welcher, 2003)

 float, be in the **water**
[gwinat](#)

Potawatomi (Welcher, 2003)

 soak s.o., put s.o. in the **water**
[gwpiit](#)

Potawatomi (Welcher, 2003)

 go inland from **water**, come ashore

[hiikkuhunux](#)

Shoshone (Di Paolo, 2010)

 kind of large **water** bird (crane?)

[jigbyék](#)

Potawatomi (Welcher, 2003)

 by the **water**, on shore

[kaka](#)

Western Sisaala (Moran, 2003)

water snake

[kakse](#)

Western Sisaala (Moran, 2003)

water snakes

[kwedbyék](#)

Potawatomi (Welcher, 2003)

 on the **water**
[kwejbyék](#)

Potawatomi (Welcher, 2003)

 on top of the **water**
[mbikadek](#)

Potawatomi (Welcher, 2003)

water s.t. down



Filters

- GOLD concepts
- Author grammatical information tokens
- Language codes
- Note types
- Entry relation types

Lexicon Lists:

Lexicon: Potawatomi
 Lexicon: Shoshone
 Lexicon: Western Sisaala

[Reselect lexicons](#)
Currently Used Filters:

None

Narrow Results By:

Lexicon Types:	Entries
Lexicon	22247

GOLD Labels:

verbal	View in GOLD	7262
noun	View in GOLD	5685
animate gender	View in GOLD	5384
singular number	View in GOLD	5225
View more		

Author Labels:

verb	6621
animate	5384
singular	5071
noun	4677
View more	

Code of Subject Language:

pot	About...	11768
ssh	About...	9375
ssl	About...	1104

Code of Gloss Language:

eng	About...	22247
---------------------	--------------------------	-------

Note types:

bibliographic	9375
-------------------------------	------

Full Text Search: IN

Use an asterisk (*) for partial searches, e.g. ba* for bat, baby, bamboozle, etc.
 Click items to the left to filter results.

Here you can search across the lexicons contained in the LEGO database. There are two ways to search:

1. Using the search box above allows you to perform a full-text search over the entry headwords, variants, glosses/definitions, examples, and notes. Use an asterisk (*) to return all entries.
2. You can filter the set of entries returned by the full-text search by selecting filters on the left. All the filters you add will appear in the "Currently Used Filters" box, and can be removed by clicking the "Remove Filter" link.

Note: You can use the filters alone to filter all entries in the database. However, if you wish to make a text search, you must do so before you use any filters. Text searches override the filters, but not vice versa.

Lexicon: Potawatomi
Lexicon: Shoshone
Lexicon: Western Sisaala

.....
Currently Used Filters:

Remove

Lexicon Types:	Entries
Lexicon	7262

animate gender	View in GOLD	3373
intransitive verb	View in GOLD	3213
singular number	View in GOLD	2406
transitive verb	View in GOLD	1709
View more ...		

verb	6621
animate	3373
intransitive	3213
singular	2406
View more ...	

pot	About...	6621
ssh	About...	631
ssl	About...	10

eng About... 7262

bibliographic 631

Submit

Previous 1 2 3 4 5 6 7 ...73 Next

[admobinkét](#)
[admobinkét](#)
[agmesét](#)
[agmesét](#)
[ahkuh](#)
[asakmegké](#)
[asakmekkét](#)
[aikutoowenne](#)
[babgoshkak](#)
[babgoshkak](#)
[babgoshké](#)
[babne'at](#)
[babne'at](#)
[babnénmat](#)
[babnénmek](#)
[babwi'at](#)
[babwi'at](#)
[babwichgét](#)
[babwichgét](#)
[babwitol](#)
[babwitol](#)
[bachbedit](#)
[bachdasét](#)
[badgedot](#)
[badgenat](#)
[badgébtot](#)
[badgécojiek](#)

[illegible]

Lexicon: Potawatomi
Lexicon: Shoshone
Lexicon: Western Sisaala

Currently Used Filters:

goldlabel: animate gender Remove

Lexicon Types:	Entries
Lexicon	3373

intransitive verb	View in GOLD	2407
singular number	View in GOLD	1630
first person	View in GOLD	1285
transitive verb	View in GOLD	950
View more		

animate	3373
verb	3373
intransitive	2407
singular	1630
View more ...	

pot About... 3373

eng About... 3373

semantic 2
































Use an asterisk (*) for partial searches, e.g. ba* for bat, baby, bamboozle, etc.
Click items to the left to filter results.

[Previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) ... [34](#) [Next](#)

[admobinkét](#)
[agmesét](#)
[agmesét](#)
[asakmegké](#)
[asakmekkét](#)
[babnénmat](#)
[babwi'at](#)
[babwichgét](#)
[babwichgét](#)
[bachbedit](#)
[bachdasét](#)
[baknan!](#)
[baknat](#)
[baknegazot](#)
[baknemwat](#)
[bakshek](#)
[bakshek](#)
[bakshen](#)
[bannagwzet](#)
[bannejat](#)
[bapashkwét](#)
[bapiwnewit](#)
[bapiwshemat](#)
[bapnezet](#)
[bapnoigé](#)
[bapnoigéshke](#)
[bapnoigéshke](#)

[illegible]

Headword: admobinkét

Author Token(s): Animate, Verb, Intransitive

Gold Concept(s): [Animate Gender](#), [Verbal](#), [Intransitive Verb](#)

Definition:

English: make cars, fix cars

Paradigmatic Variant : ndadmobinké

Grammatical Information: Animate, Singular, Verb, Intransitive, FirstPerson

Paradigmatic Variant : admobinkét

Grammatical Information: Verb, Participle



Future of LEGO

Immediate and Long-Term Plans



2011-2012

- Create a lexicon creator log-in
 - Allow users to edit and add to their data
 - User-tagging of GOLD concepts
 - Upload of our users' original lexicon documents
 - Enhance publically-available datanet of lexicons
 - Facilitate open participation of linguists
- Solicit a large number of new lexicons
- Refine the import/export facility
- Publicize the site



2012 and Beyond

- Continue to solicit new data and refine the interface
- The more data that's present on the site, the more useful it will become to semanticists, typologists, lexicographers, translators, and other researchers