

The PanLex Project

International Computer Science Institute, Berkeley, California, 27 January 02015

The goal of open-domain automatic translation of text among all human languages is ambitious, but what about the more limited goal of translating any lexeme into any language? Even this involves trillions of mappings, with quality loss when performed inferentially through intermediate languages. From 2006 to 2010, researchers at the University of Washington's Turing Center headed by Oren Etzioni combined digital dictionaries into a "translation graph" and achieved more efficient automatic inference of translations than with existing methods. Their graph, built mainly from Wiktionaries and Freelang and Freedict dictionaries, contained about 3 million lexemes. They discovered computationally practical sampling algorithms for inferring new translations without loss of precision.

Now sponsored by The Long Now Foundation in San Francisco and partnering with The Rosetta Project and the Internet Archive, the PanLex project is building and deploying this resource for research and development communities via an API, monthly database snapshots, and online applications. PanLex's database has now grown to contain 21 million lexemes, in about 10,000 languages and dialects, written in 60 scripts, with 1.2 billion attested pairwise translations. No longer a proof-of-concept project, it aspires to integrate all known lexical translations into a consistent data structure, supporting theoretical and applied research in language typology, semantic universals, machine translation, web search, text summarization, human-computer interaction, controlled languages, endangered-language revitalization, biolinguistic diversity, and other fields.

PanLex is open-source and employs Unicode, PostgreSQL, GNU/Linux, and other open-source standards and software. Its content curation is currently a small-team effort, but crowdsourced contributions are planned.

People

David Kamholz, lexical data specialist, received his linguistics Ph.D. from UC Berkeley in 2014. He has worked on computational lexicography and the historical linguistics of Austronesian languages. **Jonathan Pool**, project director, has a Ph.D. in political science from The University of Chicago. His research and teaching have focused on language politics and policy, and on language choice as a decision-theoretic problem. The PanLex team also includes research associate **Susan Colowick**, Rosetta Project director **Laura Welcher**, local volunteers, and occasional interns. The project's steering committee includes **Emily Bender** at the University of Washington and **Steven Bird** at the University of Melbourne. The project has a 19-member advisory committee.



The PanLex Project

<http://panlex.org>

(A project of The Long Now Foundation, San Francisco)

International Computer Science Institute

Berkeley, California

27 January 02015

The PanLex Project: Outline

- Problem
- Solution concept
- Concept evaluation (2006–2010)
- Open resource development (2010–)
- People, partners, publications

Problem

- Enable panlingual open-domain automatic translation
- Prioritize panlinguality

Solution concept

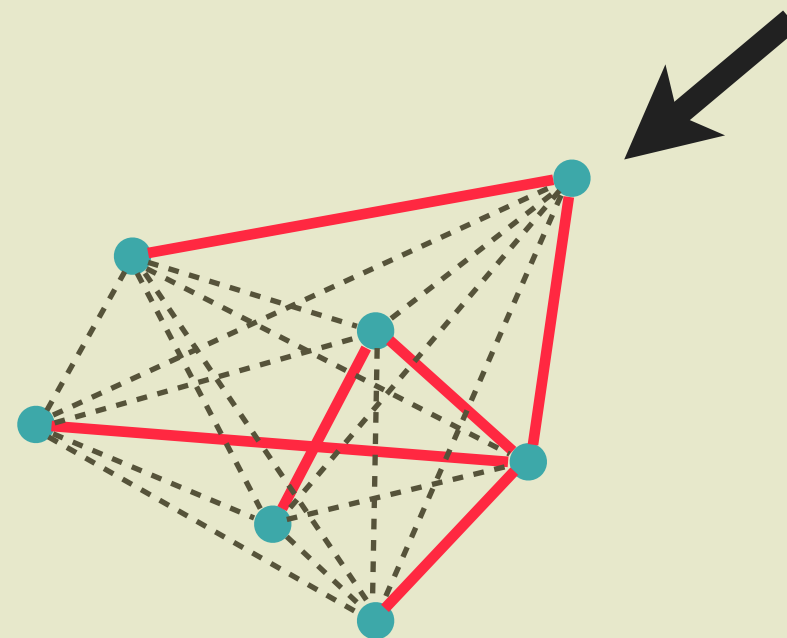
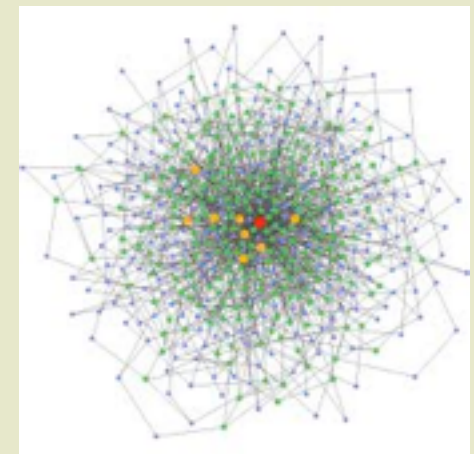
Reiter/Etzioni “Lexical Hypothesis”:

- Lexicons vs. grammars:
 - Lexeme data more:
 - Panlingually abundant
 - Tractable
 - Lexemic-only communication more useful
- Therefore, begin with a lexemic solution

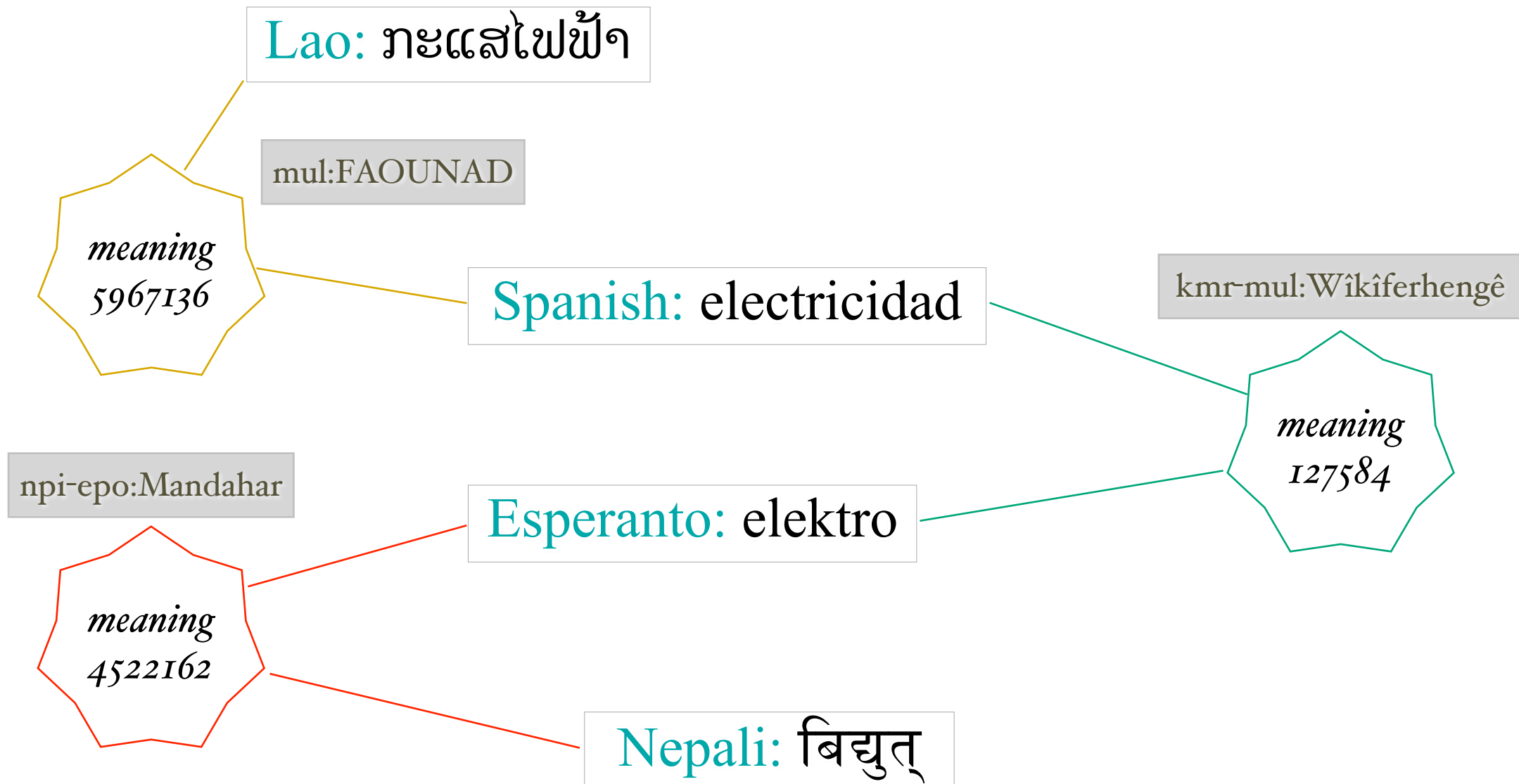
Solution concept

Reiter/Etzioni lexemic solution:

- Acquire lexeme translations from multiple sources
- Combine them into a *translation graph*
- Infer missing translations



Solution concept

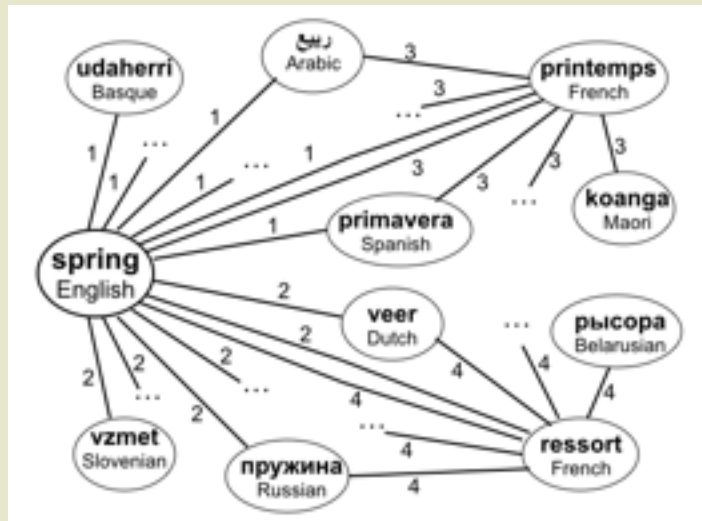


Concept evaluation

Prototype translation graph (University of Washington Turing Center):

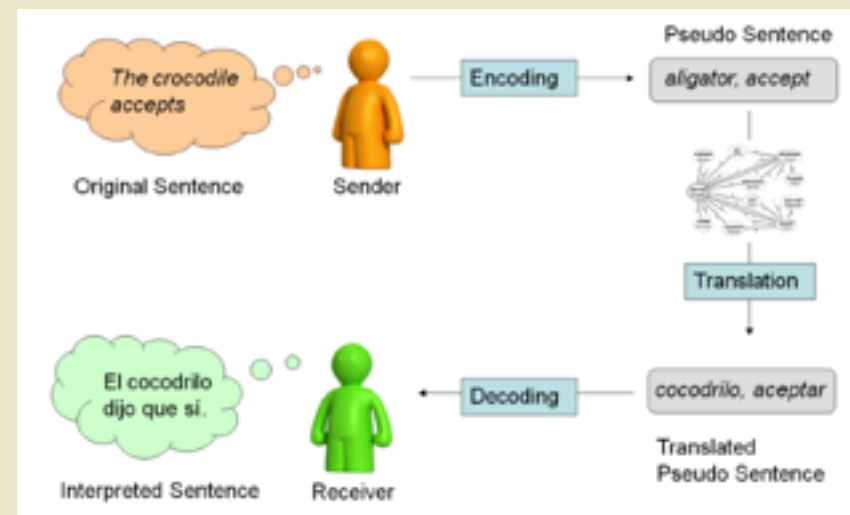
- 300–600 sources from Wiktionary, Freelang, Freedict, ...
- 1–10 million lexemes (vertices)
- 2–60 million undirected pairwise translations (edges)

Concept evaluation



Lexical translation inference

Multilingual web search



Lexeme-only communication

Concept evaluation

Demonstration applications



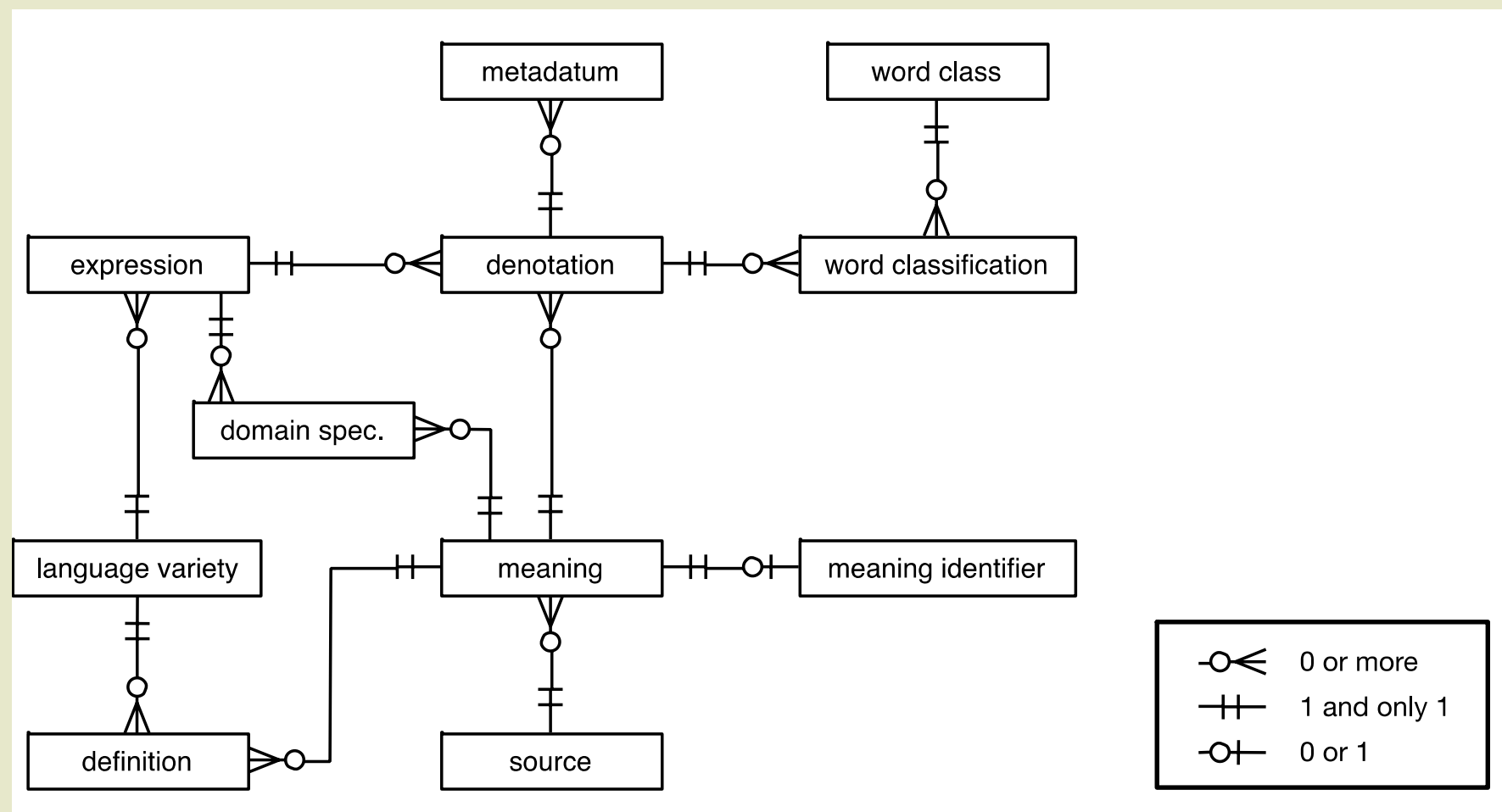
The image displays two web applications side-by-side. The left application is the 'Panlingual Translator', which features a green header with the text 'Panlingual Translator BETA Translate into any language.' Below the header, there are language selection dropdowns for 'English [English]' and 'Romanian [românește]'. The main interface includes an input field for English text (e.g., 'computer eat cookie'), a 'Translate' button, and a 'Reset' button. It also shows a 'Backtranslation' section with the translated text in Romanian ('calculator', 'mancare', 'biscuit') and a 'Panlingual Mail' section with a 'Compose Message' button and a list of email folders (Inbox, Sent Mail, Drafts, Trash). The right application is 'PanImages', which has a blue header with the text 'PanImages Cross-Lingual Image Search'. It features a search input field, a 'Search' button, and a 'Show Images' button. The page also includes a feedback link to 'panimages@cs.washington.edu' and a footer note: 'Panlingual Mail is provided by the Turing Center'.

Concept evaluation

- Marcus Sammer, Kobi Reiter, Stephen Soderland, Katrin Kirchhoff, Oren Etzioni, [“Ambiguity Reduction for Machine Translation: Human-Computer Collaboration”](#), AMTA 2006.
- Marcus Sammer, Stephen Soderland, [“Building a Sense-Distinguished Multilingual Lexicon from Monolingual Corpora and Bilingual Lexicons”](#), MT Summit 2007.
- Oren Etzioni, Kobi Reiter, Stephen Soderland, Marcus Sammer, [“Lexical Translation with Application to Image Search on the Web”](#), MT Summit 2007.
- Susan Colowick, [“Multilingual Search with PanImages”](#), *Multilingual* 2008.
- Stephen Soderland, Christopher Lim, Mausam, Bo Qin, Oren Etzioni, Jonathan Pool, [“Lematic Machine Translation”](#), MT Summit 2009.
- Mausam, Stephen Soderland, Oren Etzioni, Daniel S. Weld, Michael Skinner, Jeff Bilmes, [“Compiling a Massive, Multilingual Dictionary via Probabilistic Inference”](#), ACL-IJCNLP 2009.
- Janara Christensen, Mausam, Oren Etzioni, [“A Rose is a Roos is a Ruusu: Querying Translations for Web Image Search”](#), ACL-IJCNLP 2009.
- Katherine Everitt, Christopher Lim, Oren Etzioni, Jonathan Pool, Susan Colowick, Stephen Soderland, [“Evaluating Lematic Communication”](#), *trans-kom* 2010.
- Mausam, Stephen Soderland, Oren Etzioni, Daniel S. Weld, Kobi Reiter, Michael Skinner, Jeff Bilmes, [“Panlingual Lexical Translation via Probabilistic Inference”](#), *Artificial Intelligence* 2010.

Open resource development

- Refined database schema (language *varieties*, domains, ...)
- Open-source DB (PostgreSQL) and OS (GNU/Linux)



Open resource development

Language varieties:

- Literary standards
- Dialects
- Script-specific varieties
- Orthographic standards
- Controlled languages
- Standard codes for countries, chemical elements, etc.

ajg-000	Aja
ajg-001	Hwe
aja-002	Dogbo
ajg-003	Sikpi
ajg-004	Tohoun
ajg-005	Tado
ajg-006	Tala
aji-000	Ajie
ajp-000	اللهجة الجنوبية

uzn-000	o'zbek
uzn-001	Ўзбекча
uzn-002	أوزبېك تىلى

ISO 639-2, 3, 5

Open resource development

Language-variety codes: example of Western Armenian

- [PanLex](#): hye-001
 - [BCP 47](#): hy-arevmnda
 - [Glottolog](#) glottocodes: west2348
 - [MultiTree](#): hye-war
 - [WALS Codes](#): arw
 - [Linguasphere](#): 57-AAA-ac
 - [Lexvo](#): <http://lexvo.org/id/iso639-3/hye>
 - [Wikidata](#): Q180945
 - [Freebase](#): <http://www.freebase.com/m/01tzg9>
 - [ISO 639-1](#): hy
 - [ISO 639-2 \(B\)](#): arm
 - [ISO 639-2 \(T\)](#): hye
 - [ISO 639-3](#): hye
- } Armenian generally

Open resource development

More sources, digital and printed

5500 acquired, 1700 consulted

- Monolingual dictionaries
- Bilingual dictionaries
- Multilingual dictionaries
- Wiktionaries
- Glossaries
- Vocabularies
- Wordlists
- Terminologies
- Wordnets
- Thesauri
- Standards
- Vocabulary databases
- Locale databases

Salata سَلَاةٌ ج سَلَاةٌ
Hüküm, nüfuz, سُلْطَانٌ / سُلْطَاتٌ
otorite, güç, kuvvet
Güç, kuvvet, otorite, سُلْطَانٌ ج سَلَاةٌ
—yönetim, idare, —manevi etki, tesir, —
delil, —sultan, hükümdar
Kadın sul

γλώσσα για ειδικούς σκοπούς
MT (70.20)
Da: fagsprog
De: Fachsprache
En: language for special purposes
Es: lenguaje especializado
Fi: kieli tiettyihin tarkoituksiin
Fr: langage spécialisé
He: שפה למטרות מיוחדות
Hu: szaknyelv
It: lingua speciale
Nl: vaktaal
Sv: fackspråk
BT γλώσσες

Arrest: ሰጋጋሪ ምጽኑ፡ **Tigujauniq:** Arrestation

The act of placing a person in custody, according to law. The powers of ordinary citizens and peace officers to arrest a person are set out in the 1996, Part XVI.

Ikitittiniq: Crime

berately setting fire to property. 1996, sections 433-436.

SE (English: Sweden)

An tSualainn	·ga·
isveç	·az·
Ísveç	·tr·
Iswidhan	·so·
Rootsi	·et·
Ruotsi	·fi·
Ruotta	·se·
Schweden	·de·
Schweede	·gsw·

Open resource development

More data. Currently:

- 21 million *expressions* (\approx lemmas of lexemes)
- Expressions written in 60 scripts (Latin, 한글, Հայոց գրեր, தமிழ் அரிச்சுவடி, ...)
- 10,000 language varieties
- 1.2 billion undirected pairwise translations

Open resource development

Access to raw data via:

- [Internet Archive](#) Open Library (printed originals)
- [Links to original sources](#)

Open resource development

The screenshot shows a web browser window displaying a bilingual dictionary. The address bar shows the URL <https://archive.org/stream/praktikavortaro00murj#page/19/mode/2up>. The page title is "Nova praktika vortaro hebra-Esperanto". The page content is a table of words in three columns: Esperanto, Hebrew, and Esperanto. The words are arranged in rows, with some words having multiple Hebrew equivalents. The page is titled "Nova praktika vortaro hebra-Esperanto" and includes a search bar and navigation controls.

abio	אִשׁוּן (ב)	incidento	
konfirmo, aprobo	אִשׁוּר	malbenita	אָרוּר
fago	אִשׁוּר (ב)	paki	אָרְז
fundamento	אִשְׁדָּה	cedro	אָרְז
testiko	אִשְׁךָ (א)	rizo	אָרְז (ב)
grapolo; traŭbo	אִשְׁכּוּל	gastigi	אָרַח
grapfrukto	אִשְׁכּוּלִית	leono	אָרִי, אָרִיָּה (ז)
kalio (K)	אִשְׁלָג (כ)	pak/ado, ~umo	אָרִיָּזָה
iluzio	אִשְׁלָה	kahelo	אָרִיָּח
kulpa	אִשָּׁם	longeco	אָרְךָ
kulpulo	אִשְׁמָאי	longitudo	גֵּאוֹגְרָפִי
diablo; satano	אִשְׁמָנִי	vivlong(ec)o	חַיִּים
diurno-parto	אִשְׁמוּרָה, אִשְׁמָרָה	longa	אָרְךָ
giĉeto; embrazuro	אִשְׁנָב	prolongigo	אָרְכָּה
balaaĵo; rub(aĵ)o	אִשְׁפָּה	arkivo	אָרְכִּיוֹן

Open resource development

Access to processed data via:

- Monthly CSV, JSON, and XML [dumps](#)
- [API](#) (read-only)
- API apps for humans and search engines
- [RDF interface](#) linking (earlier) PanLex to DBpedia
- Expert web UI (“[PanLem](#)”)
- Team only: CLI (psql)

Open resource development

Applications

Snorql: Exploring <http://ld.panlex.org/sparql>

SPARQL: Browse:

PREFIX rdf: <<http://www.w3.org/1999/02/22-rdf-syntax-ns#>>

WHERE {

Go!

PanLex Tattoos

Choose a word:

- bliss
- bond
- bravery
- Buddha
- Buddhism
- butterfly
- calligraphy
- camouflage
- carp
- cat
- century
- challenge
- chance
- change
- chaos
- charm
- choice
- Christmas Eve
- civilization
- clarity

Or type one:

PanLinx

русский rus-000
ирригация

Afrikaans afr-000 [besproeiing](#)

العربية arb-000 [سقي](#)

беларуская bel-000 [ірыгання](#)

বাংলা ben-000 [সেচ](#)

brezhoneg bre-000 [dourañ](#)

български bul-000 [Напоояване](#)

български bul-000 [напоояване](#)

català cat-000 [reg](#)

català cat-000 [regadiu](#)

català cat-000 [regatge](#)

čeština ces-000 [irigace](#)

čeština ces-000 [vyplachování](#)

čeština ces-000 [výplach](#)

čeština ces-000 [zavlažování](#)


čeština ces-000 [závlaha](#)

简体字 cmn-000 [水利](#)

简体字 cmn-000 [水利化](#)

简体字 cmn-000 [灌注](#)

简体字 cmn-000 [灌溉](#)



Gaeilge

Focal Cuardach »

Ó

eo af

Go

TümSöz

Burada Türkçe sözleri yüzlerce dile çevirebilirsiniz!

Söz veya deyim:

Open resource development

API

- Provides live access to the database
- Public and free (within rate limits)
- Takes JSON queries and returns JSON responses for a variety of database objects
- Example query requesting translations of the English (eng-000) expression “church” into Western Armenian (hye-001), sorted by quality:

```
/ex { "uid": ["hye-001"], "trtt":  
  ["church"], "truid": ["eng-000"],  
  "include": "trq", "sort": "trq desc" }
```

Open resource development

Source acquisition

- Prioritization of sources that are tractable (amenable to analysis) and which cover poorly documented languages
- Already ingested sources are mostly in born-digital formats (text, HTML, PDF, etc.)
- Increasing acquisition of print sources, purchased and scanned in partnership with the Internet Archive

Open resource development

Source workflow

- Identification: choice of a source from the source archive
- Consultation: determination of the logical structure of the source and what language varieties it contains
- Analysis: semi-automated computational extraction of usable data
- Submission: ingestion of the final source file into the PanLex database

Open resource development

Source consultation

Some of the Cebuano Words Recorded by Pigafetta in 1521

by [Jessie Grace U. Rubrico](#)

ENGLISH	OLD CEBUANO	MODERN CEBUANO	ENGLISH	OLD CEBUANO	MODERN CEBUANO
man	lac	lalaki	balances	tigban	timbangan
woman	paranpaon	babaye	weight	tahil	timbang
young woman	beni beni	dalaga	pearl	mutiara	mutya
married woman	babay	babayeng minyo	mother of pearl	tipay	tipay
buttocks	samput	balat-ang	pipe (musical instrument)	subing	plawta
thigh	paha	paa	disease of St. Job	alupalan	hubag
knee	tuhud	tuhod	bring me	palatin comorica	ambi
shin	bassag-bassag	bitiis	certain rice cakes	tinapai	puto/bibingka
ankle	bolbol	buulbuul	good	main	maayo
heel	tiochid	kiting	no	tifale	dili
sole (foot)	lapa lapa	lapalapa	knife	capol/sundan	kutsilyo
gold	balaoan	bulawan	scissors	catle	gunting
silver	pilla	plata/pilak	to shave	chunthinch	mamalbas
brass	concach	tumbaga/bronse	well adorned man	pixao	pustorawo

Open resource development

Source analysis

- Tabularization: transformation of source data into well-defined tables with one entry per row
- Serialization: normalization of tabularized data and conversion into a format that can be validated and ingested into the PanLex database
- Suite of tools written in Perl developed to streamline the process (<http://github.com/longnow/panlex-tools>)

Open resource development

Tabularization: methods

- Custom scripts (usually Perl) using regular expressions, HTML/XML parsers, etc.
- PDF-to-text conversion (with or without layout analysis)
- PDF-to-XML dump (preserves style information)
- Manual transcription (for scanned print sources)

Open resource development

Tabularization: HTML example

```
<tr>
<td>man</td>

<td>lac</td>

<td>lalaki</td>

<td>balances</td>

<td>tigban</td>

<td>timbangan</td>
</tr>

<tr>
<td>woman</td>

<td>paranpaon</td>

<td>babaye</td>

<td>weight</td>

<td>tahil</td>

<td>timbang</td>
</tr>
```

```
manΔlacΔlalaki-
balancesΔ  tigbanΔ timbangan-
womanΔ  paranpaonΔ  babaye-
weightΔ tahilΔ  timbang-
young womanΔbeni beniΔ  dalaga-
pearlΔ  mutiaraΔmutya-
married womanΔ  babayΔ  babayeng minyo-
mother of pearlΔtipayΔ  tipay-
buttocksΔ  samputΔ  balat-ang-
pipe (musical instrument)Δ  subingΔ  plawta-
thighΔ  pahaΔ  paa-
disease of St. JobΔalupalanΔ  hubag-
kneeΔ  tuhudΔ  tuhod-
bring meΔ  palatin comoricaΔ  ambi-
shinΔ  bassag- bassagΔ  bitiis-
certain rice cakesΔ  tinapaiΔputo/  bibingka-
ankleΔ  bolbolΔ  buulbuul-
goodΔ  mainΔ  maayo-
heelΔ  tiochidΔkiting-
noΔ  tifaleΔ  dili-
sole (foot)Δlapa lapaΔ  lapalapa-
knifeΔ  capol/sundanΔ  kutsilyo-
goldΔ  balaoanΔbulawan-
scissorsΔ  catleΔ  gunting-
silverΔ  pillaΔ  plata/pilak-
to shaveΔ  chunthinchΔ  mamalbas-
brassΔ  concachΔtumbaga/  bronse-
```

Open resource development

Tabularization: PDF example with text styles coding important information

revivir *vi* chédlane', chela' š'i'i
revocar *vt* checa'a
revolcar 1. *vt* chchix̣ chtole
2. *v prnl* chas chata', chbix̣ chtole
revolver *vt* 1. chchixe, chyitje
2. chta (*líquido*)
rey *m* rey
rezar *vi* chon rsar

Open resource development

Tabularization output of previous example

```
revolcarΔ vtΔ chchix_ chtol_e¬  
revolcarΔ v prnlΔ chas chata'►chbix_ chtol_e¬  
revolverΔ vtΔ chc_hix_e►chyitje¬  
revolver (líquido)Δ vtΔ chta¬  
reyΔmΔ rey¬  
rezarΔ viΔ chon rsar¬
```

Open resource development

Serialization: methods

- Tagging of columns as expressions, definitions, word classes, metadata, etc.
- Recategorization of non-lemmatic portions of expressions as definitions (e.g. parenthesized text) or word classes (e.g. question marks)
- Normalization of strings as expressions or definitions on the basis of length, Unicode characters, or attestation in PanLex

Open resource development

Serialization: tagging and normalization example

manΔlacΔlalaki-
balancesΔ tigbanΔ timbangan-
womanΔ paranpaonΔ babaye-
weightΔ tahlΔ timbang-
young womanΔbeni beniΔ dalaga-
pearlΔ mutiaraΔmutya-
married womanΔ babayΔ babayeng minyo-
mother of pearlΔtipayΔ tipay-
buttocksΔ samputΔ balat-ang-
pipe (musical instrument)Δ subingΔ plawta-
thighΔ pahaΔ paa-
disease of St. JobΔalupalanΔ hubag-
kneeΔ tuhudΔ tuhod-
bring meΔ palatin comoricaΔ ambi-
shinΔ bassag- bassagΔ bitiis-
certain rice cakesΔ tinapaiΔputo/ bibingka-
ankleΔ bolbolΔ buulbuul-
goodΔ mainΔ maayo-
heelΔ tiochidΔkiting-
noΔ tifaleΔ dili-
sole (foot)Δlapa lapaΔ lapalapa-
knifeΔ capol/sundanΔ kutsilyo-
goldΔ balaoanΔbulawan-
scissorsΔ catleΔ gunting-
silverΔ pillaΔ plata/pilak-
to shaveΔ chunthinchΔ mamalbas-
brassΔ concachΔtumbaga/ bronse-



⟨ex⟩manΔ ⟨ex⟩lacΔ ⟨ex⟩lalaki-
⟨ex⟩balancesΔ ⟨ex⟩tigbanΔ⟨ex⟩timbangan-
⟨ex⟩womanΔ ⟨ex⟩paranpaonΔ ⟨ex⟩babaye-
⟨ex⟩weightΔ⟨ex⟩tahlΔ ⟨ex⟩timbang-
⟨ex⟩young womanΔ ⟨ex⟩beni beniΔ ⟨ex⟩dalaga-
⟨ex⟩pearlΔ ⟨ex⟩mutiaraΔ ⟨ex⟩mutya-
⟨ex⟩married womanΔ ⟨ex⟩babayΔ ⟨ex⟩babayeng minyo-
⟨ex⟩mother of pearlΔ ⟨ex⟩tipayΔ ⟨ex⟩tipay-
⟨ex⟩buttocksΔ ⟨ex⟩samputΔ⟨ex⟩balat-ang-
⟨df⟩pipe (musical instrument)⟨ex⟩pipeΔ⟨ex⟩subingΔ⟨ex⟩plawta-
⟨ex⟩thighΔ ⟨ex⟩pahaΔ ⟨ex⟩paa-
⟨df⟩disease of St. JobΔ⟨ex⟩alupalanΔ ⟨ex⟩hubag-
⟨ex⟩kneeΔ ⟨ex⟩tuhudΔ ⟨ex⟩tuhod-
⟨df⟩bring meΔ ⟨ex⟩palatin comoricaΔ ⟨ex⟩ambi-
⟨ex⟩shinΔ ⟨ex⟩bassag-bassagΔ ⟨ex⟩bitiis-
⟨df⟩certain rice cakesΔ⟨ex⟩tinapaiΔ ⟨ex⟩puto⟨ex⟩bibingka-
⟨ex⟩ankleΔ ⟨ex⟩bolbolΔ⟨ex⟩buulbuul-
⟨ex⟩goodΔ ⟨ex⟩mainΔ ⟨ex⟩maayo-
⟨ex⟩heelΔ ⟨ex⟩tiochidΔ ⟨ex⟩kiting-
⟨ex⟩noΔ⟨ex⟩tifaleΔ⟨ex⟩dili-
⟨df⟩sole (foot)⟨ex⟩soleΔ ⟨ex⟩lapa lapaΔ ⟨ex⟩lapalapa-
⟨ex⟩knifeΔ ⟨ex⟩capol⟨ex⟩sundanΔ ⟨ex⟩kutsilyo-
⟨ex⟩goldΔ ⟨ex⟩balaoanΔ ⟨ex⟩bulawan-
⟨ex⟩scissorsΔ ⟨ex⟩catleΔ ⟨ex⟩gunting-
⟨ex⟩silverΔ⟨ex⟩pillaΔ ⟨ex⟩plata⟨ex⟩pilak-
⟨df⟩(to) shave⟨ex⟩shaveΔ ⟨ex⟩chunthinchΔ⟨ex⟩mamalbas-
⟨ex⟩brassΔ ⟨ex⟩concachΔ ⟨ex⟩tumbaga⟨ex⟩bronse-

Open resource development

Serialization: generation of final source file

```
◀ex▶manΔ  ◀ex▶lacΔ  ◀ex▶lalaki~  
◀ex▶balancesΔ  ◀ex▶tigbanΔ◀ex▶timbangan~  
◀ex▶womanΔ  ◀ex▶paranpaonΔ  ◀ex▶babaye~  
◀ex▶weightΔ◀ex▶tahilΔ  ◀ex▶timbang~  
◀ex▶young womanΔ  ◀ex▶beni beniΔ  ◀ex▶dalaga~  
◀ex▶pearlΔ  ◀ex▶mutiaraΔ  ◀ex▶mutya~  
◀ex▶married womanΔ  ◀ex▶babayΔ  ◀ex▶babayeng minyo~  
◀ex▶mother of pearlΔ  ◀ex▶tipayΔ  ◀ex▶tipay~  
◀ex▶buttocksΔ  ◀ex▶samputΔ◀ex▶balat~ang~  
◀df▶pipe (musical instrument)◀ex▶pipeΔ◀ex▶subingΔ◀ex▶plawta~  
◀ex▶thighΔ  ◀ex▶pahaΔ  ◀ex▶paa~  
◀df▶disease of St. JobΔ◀ex▶alupalanΔ  ◀ex▶hubag~  
◀ex▶kneeΔ  ◀ex▶tuhudΔ  ◀ex▶tuhod~  
◀df▶bring meΔ  ◀ex▶palatin comoricaΔ  ◀ex▶ambi~  
◀ex▶shinΔ  ◀ex▶bassag~bassagΔ  ◀ex▶bitiis~  
◀df▶certain rice cakesΔ◀ex▶tinapaiΔ  ◀ex▶puto◀ex▶bibingka~  
◀ex▶ankleΔ  ◀ex▶bolbolΔ◀ex▶buulbuul~  
◀ex▶goodΔ  ◀ex▶mainΔ  ◀ex▶maayo~  
◀ex▶heelΔ  ◀ex▶tiochidΔ  ◀ex▶kiting~  
◀ex▶noΔ◀ex▶tifaleΔ◀ex▶dili~  
◀df▶sole (foot)◀ex▶soleΔ  ◀ex▶lapa lapaΔ  ◀ex▶lapalapa~  
◀ex▶knifeΔ  ◀ex▶capol◀ex▶sundanΔ  ◀ex▶kutsilyo~  
◀ex▶goldΔ  ◀ex▶balaoanΔ  ◀ex▶bulawan~  
◀ex▶scissorsΔ  ◀ex▶catleΔ  ◀ex▶gunting~  
◀ex▶silverΔ◀ex▶pillaΔ  ◀ex▶plata◀ex▶pilak~  
◀df▶(to) shave◀ex▶shaveΔ  ◀ex▶chunthinchΔ◀ex▶mamalbas~  
◀ex▶brassΔ  ◀ex▶concachΔ  ◀ex▶tumbaga◀ex▶bronse~
```

```
[:  
0  
  
ex  
eng-000  
man  
ex  
ceb-002  
lac  
ex  
ceb-000  
lalaki  
  
ex  
eng-000  
balances  
ex  
ceb-002  
tigban  
ex  
ceb-000  
timbangan  
  
ex  
eng-000  
woman  
ex  
ceb-002  
paranpaon  
ex  
ceb-000  
babaye
```


People

Project team:

- David Kamholz, Lexical Data Specialist
- Jonathan Pool, Project Director
- Susan Colowick, Research Associate
- Laura Welcher + Emily Bender + Steven Bird, Steering Committee

Original team, University of Washington Turing Center:

Oren Etzioni

Stephen Soderland

Mausam

Marcus Sammer

Kobi Reiter

Michael Skinner

Christopher Lim

Janara Christensen

Michael Schmitz

Partners

- University of Washington Turing Center (initiator): <http://turing.cs.washington.edu>
- The Long Now Foundation (sponsor): <http://longnow.org>
- The Rosetta Project: <http://rosettaproject.org>
- Internet Archive: <http://archive.org>
- Unicode Consortium: <http://unicode.org>
- Global Glossary: <http://globalglossary.org>
- Open Knowledge Foundation: <http://okfn.org>

Publications

Papers:

- Timothy Baldwin, Jonathan Pool, Susan M. Colowick, [“PanLex and LEXTRACT: Translating all Words of all Languages of the World”](#), Coling 2010.
- Jonathan Pool, [“Panlingual Globalization”](#), *Handbook of Language and Globalization* 2010.
- Francesca Gola, [“An analysis of translation divergence patterns using PanLex translation pairs”](#), MS thesis, Univ of Wash, 2012.
- Patrick Westphal, Claus Stadler, Jonathan Pool, [“Countering language attrition with PanLex and the Web of Data”](#), *Semantic Web Journal* 2014.
- David Kamholz, Jonathan Pool, Susan M. Colowick, [“PanLex: Building a Resource for Panlingual Lexical Translation”](#), LREC 2014.

Presentations:

DELPH-IN Summit 2011

IUC (Internationalization and Unicode Conference) 2012.