

Language Documentation and Dictionaries

Ken Manson

Payap University

&

SIL International



Language Documentation

- “lasting multipurpose record of a language”
 - Himmelmann (2006:1)
- In practice this means a video/audio collection
 - Often not seen a relevant for language communities
 - Language communities like dictionaries and grammars

Text Collections

- Not the only repository of cultural knowledge
- Low frequency words may not be present

– BNC (100,000,000 words)

- 'liver' 1,633x
- 'tail' 2,630x
- 'burn' 1,683x
- 'yellow' 4,454x
- 'intestine(s)' 474x

Lg Doc **AND** Dictionaries

- Not an either-or
- Do-able in a short time with low resources
 - Mosel's (2004:42) claim that AUD10,000 was not enough to produce a dictionary is wrong
- Community involvement is essential

History of dictionary development

- 1st dictionary for linguist – lg learning, interlinearising
 - Limited to text collection and wordlists for phonological analysis
- Forms the basis for a community dictionary
- Revised and expanded by interested community members &/or linguist
 - Methodology often not motivated
 - Linguist is a bottle-neck
 - Technology complicated

Community lexicography

- Empower the community to develop the dictionary
 - Easy-of-access software
 - Task oriented
 - Frees linguist to focus on technical issues
 - Creates community ownership

Dictionary Development Program

- Key features
 - Semantic domain-based
 - Many different “dictionary” products
 - wordlist, thesaurus, finderlist, dictionary...
 - Task oriented
 - Collect words, gloss words, add part of speech, add example sentences...
- www.sil.org/computing/DDP/index.htm

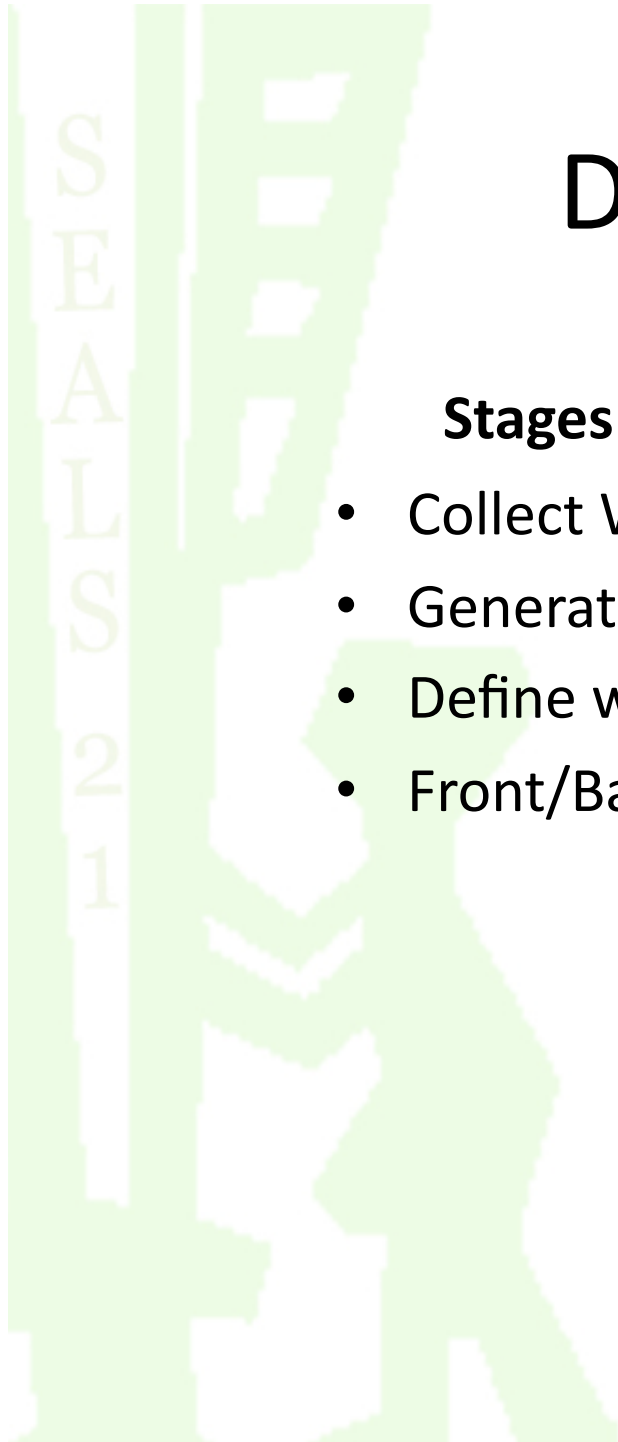
Dictionary Stages

Stages

- Collect Words
- Generate additional fields
- Define words
- Front/Back matter

Outputs

- Thesaurus
- Basic dictionary
- Definitions dictionary
- Dictionary
 - Expanded
 - Spelling
 - School
 - Specific domain
 - Etc.



Collecting words



- 2-week workshop
 - \$200 – materials & food for 20 people
 - \$100 – my travel
- Semantic domains
 - ~60 folders with prompt questions
 - Translated into many national languages
- Typically collect around 10,000 to 15,000 entries

Semantic domain examples

5.4.3.6	Shave	2.6.4.0	Grow, get bigger
5.4.4	Care for the teeth	2.6.4.7	Initiation
5.4.5	Anoint the body	2.6.4.8	Peer group
5.4.6	Ritual scar	2.6.5	Male, female
5.4.6.1	Circumcision	2.6.5.1	Man
5.4.7	Care for the fingernails	2.6.5.2	Woman
5.5	Fire	2.6.6	Die
5.5.1	Light a fire	2.6.6.1	Kill
5.5.2	Tend a fire	2.6.6.2	Corpse
5.5.3	Extinguish a fire	2.6.6.3	Funeral
5.5.4	Burn	2.6.6.4	Mourn
5.5.5	What fires produce	2.6.6.5	Bury
5.5.6	Fuel	2.6.6.6	Grave
5.5.7	Fireplace	2.6.6.7	Inherit
5.6	Cleaning	2.6.6.8	Life after death
5.6.1	Clean, dirty	3	Language and thought
5.6.2	Bathe	3.1	Soul, spirit
5.6.3	Wash dishes	3.1.1	Personality
5.6.4	Wash clothes	3.1.2	Mental state

Prompt Examples

5.5 Fire

Use this domain for general words that refer to fire and for words referring to types of fire. These words may be specific for what is being burned (forest fire 'a fire burning a forest'), the size of the fire (inferno 'a very large hot fire'), or the place where the fire burns (hellfire 'the fire in hell').

- (1) What general words refer to a fire? *fire, combustion*
- (2) What types of fire are there? *fireball, torch, firebrand, fireworks, firecracker, burn a field before planting*
- (3) What words refer to what is being burned? *housefire, forest fire, grass fire, wood fire, oil fire*

5.5 မီး: Fire

မီးနှင့် မီးအမျိုးအစားများကို ရည်ညွှန်းသော ရေဘုယျစကားလုံးများ အတွက် အထက်ပါ ခေါင်းစဉ်ဝေါဟာရကို အသုံးပြုပါ။ ဤစကားလုံးများသည် လောင်ကျွမ်းနေသောအရာ(တောမီး-တောကိုလောင်ကျွမ်းစေသောမီး)၊ လောင်ကျွမ်းသောပမာဏ(တဟုန်းဟုန်းတောက်သောမီး-အလွန်ကြီးသောမီး)၊ သို့မဟုတ် မီးလောင်သောနေရာ(ငရဲမီး-ငရဲပြည်တွင်လောင်သောမီး)တို့ကို တိတိကျကျဖော်ပြချက်များ ဖြစ်ကောင်းဖြစ်မည်။

- (1) မီးကို ရေဘုယျအားဖြင့် မည်သည့် စကားလုံးများဖြင့်ရည်ညွှန်းသနည်း။
ဥပမာ ။ ။ မီး၊ လောင်ကျွမ်းခြင်း၊
- (2) မီး အမျိုးအစားများကိုဖော်ပြပါ။
ဥပမာ ။ ။ ပေါက်ကွဲမှုကြောင့်ဖြစ်ပေါ်သည့်မီးလုံး၊ မီးတုတ်၊ မီးရှူးမီးတိုင်၊ ဖြောက်အိုး၊ မီးပန်း၊ ရှိမီး၊
- (3) လောင်ကျွမ်းနေသော အရာကို မည်သည့် စကားလုံးများဖြင့်ဖော်ပြသနည်း။
ဥပမာ ။ ။ တောမီး၊ ကောက်ရိုးမီး၊ မီးသွေးမီး၊ ထင်းမီး၊ ဆီမီး၊

Dictionary Software

- LIFT dictionary format

- WeSay

– www.WeSay.org



- Fieldworks Language Explorer

– fieldworks.sil.org

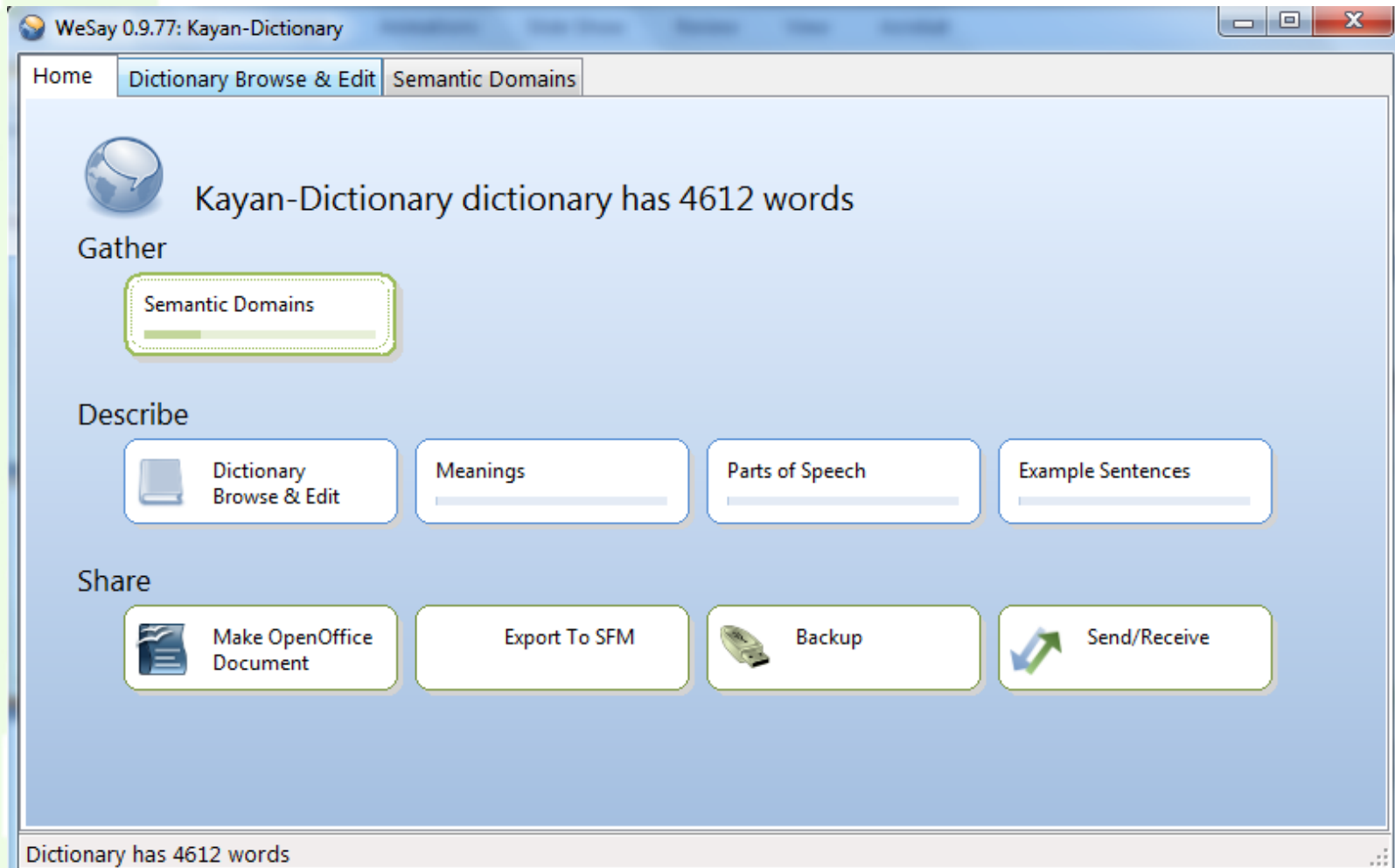


- LexiquePro

– www.LexiquePro.org

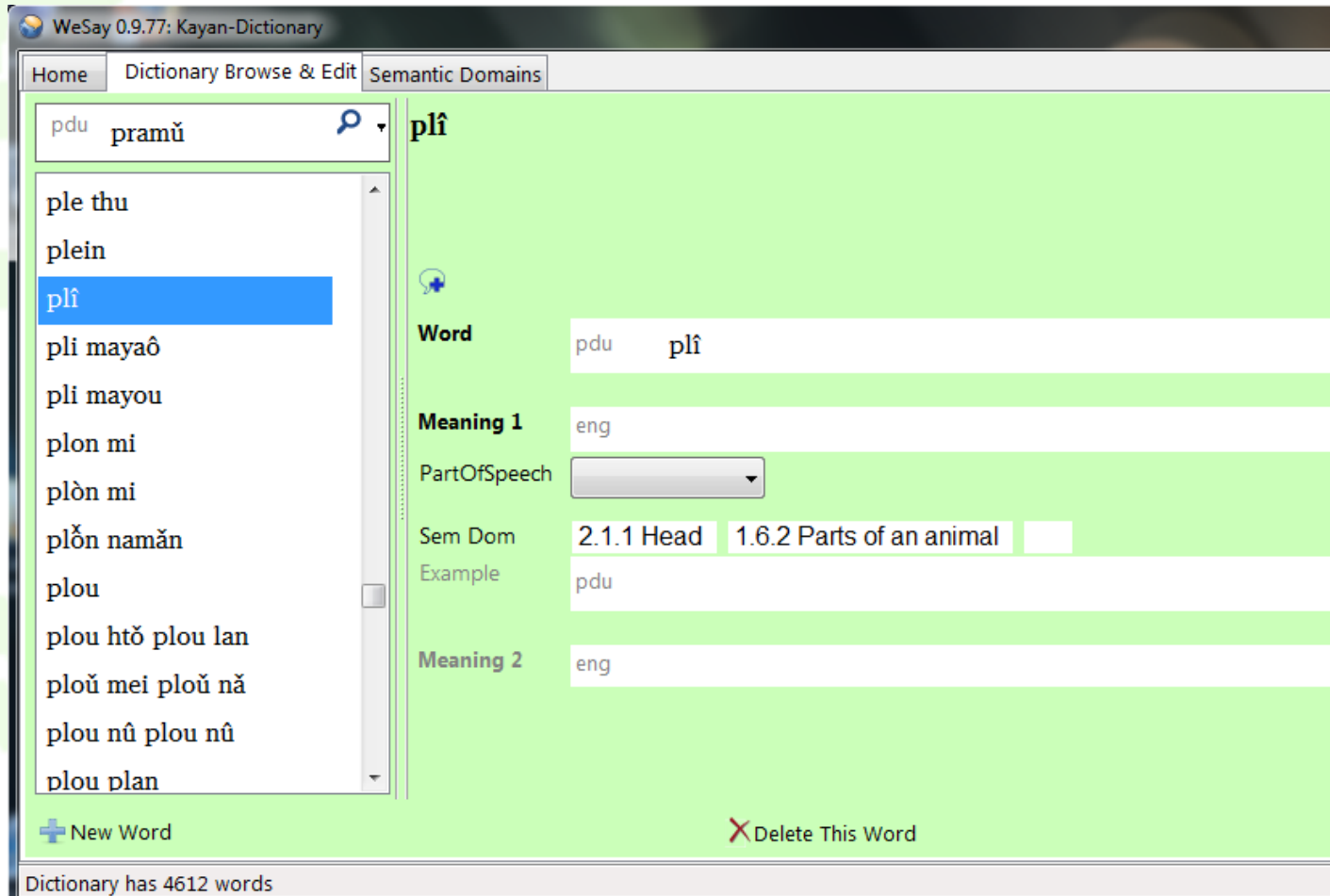


WeSay



www.wesay.org

WeSay



www.wesay.org

FLEX

Kayan DDP - FieldWorks Language Explorer

File Edit View Data Insert Format Tools Parser Window Help

Lexicon

Entries

Headword	Lexeme F...	Glosses	Grammatic...
ple hoa	ple hoa		
ple ka so	ple ka so		
ple kada	ple kada		
ple kan bon	ple kan bon		
ple kon	ple kon		
ple pûn lai	ple pûn lai		
ple phaphon	ple phaphon		
ple shen	ple shen		
ple ta	ple ta		
ple htwai soa	ple htwai soa		
ple thu	ple thu		
plein	plein		
pli mayaô	pli mayaô		
pli mayou	pli mayou		
plî	plî		
plon mi	plon mi		
plou	plou		
plou nû plou	plou nû plou nû		

Entry

plî (sem. domains: 2.1.1 - Head, 1.6.2 - Parts of an animal.)

Lexeme Form Key **plî**

Morph Type stem

Citation Form Key

Components

Note Eng

Sense 1

Gloss Eng

Definition Eng

Grammatical Info. <Not Sure>

Example

Semantic Domains 2.1.1 - Head 1.6.2 - Parts of an animal

Lexical Relations

Variants

Allomorphs

Grammatical Info. Details

Category Info. Stem/root of unknown category; takes any affix

14/Mar/2011 17/Mar/2011 Queue: (-/-) No Parser Loaded 2950/4609

fieldworks.sil.org

LexiquePro

Kayan - Lexique Pro

File Edit View Tools Help

Back Forward Home Pronounce Edit More Info

Kayan Categories Search

a a
a aò ran
a aò htan lǔ
a ao can
a bwei kei ta ma khao
a cǒ cǎn cǒ plan
a cǒ kran
a cǒ phi khu
a kei
a khei ba andyǎn cǎ
a khei ba aò htaô
a khû a nǎ
a khû tha
a ma cǒ dai cǒ yan
a ma htan ta thei ngǎ atǎn b
a mu cu
a môn kau
a mún

Lexique Pro

Kayan

• a • ai • ao • aò • aô • au • aú • b • c • d • e • ei • è • g • gh • h • i • î • k • kh • l • m
• n • gn • ng • o • ò • ô • p • ph • r • s • t • ht • th • u • û • w • y •

Kayan: 4608 words

a
ai
ao
aò
aô
au
aú
b
c
d
e
ei
è
g
gh
h
i
î
k
kh
l
m
n
gn
ng
o
ò
ô
p

www.LexiquePro.org

LexiquePro

The screenshot displays the LexiquePro application window for the Kayan language. The window title is "Kayan - Lexique Pro". The menu bar includes "File", "Edit", "View", "Tools", and "Help". Below the menu bar is a toolbar with icons for "Back", "Forward", "Home", "Pronounce", "Edit", and "More Info". The main interface is divided into three sections: "Kayan", "Categories", and "Search". The "Categories" section on the left shows a hierarchical tree structure with "2.1.1.3. Nose" selected. The main content area displays the title "2.1.1.3. Nose" and a list of words: **daũ niphũ**, **kasai**, **niphũ**, **niphũ lanyi**, **niphũthi**, **niphũyi**, **niphũ-ĩ**, **phĩ niphũ**, and **tharu niphũ**. Below the list, it indicates the "Parent category: 2.1.1. Head". On the right side, there is a vertical alphabetical index from "a" to "p". The status bar at the bottom shows "Kayan: 4608 words" and "Category: 9 words".

www.LexiquePro.org

LexiquePro

Flow.

sǔ htan *Category: 1.3.1.4. Spring, well, 1.3.2.1.*

Flow, 1.3.2.6. Tide.

sǔ kroa han *Category: 1.3.2.1. Flow.*

sǔ lan *Category: 1.3.2.1. Flow, 1.3.2.6. Tide.*

sǔ lan pla *Category: 1.3.2.1. Flow.*

sǔ myûn lan *Category: 1.3.2.1. Flow.*

sǔ thwai kyon *Category: 1.3.2.1. Flow.*

sǔ thwai lan phon *Category: 1.3.2.1. Flow.*

sǔ zoa swai *Category: 1.3.2.1. Flow.*

1.3.2.2. Pour

lan zoa sǔ *Category: 1.3.2.2. Pour, 1.3.4. Be in water.*

lwai sǔ *Category: 1.3.2.2. Pour.*

pryûn sǔ *Category: 1.3.2.2. Pour.*

sǔ han pla *Category: 1.3.2.2. Pour.*

su thwai sǔa *Category: 1.3.2.3. Cam, rough.*

1.3.2.6. Tide

sǔ htan *Category: 1.3.1.4. Spring, well, 1.3.2.1.*

Flow, 1.3.2.6. Tide.

sǔ htan lan *Category: 1.3.2.6. Tide.*

sǔ lan *Category: 1.3.2.1. Flow, 1.3.2.6. Tide.*

1.3.3. Wet

caò *Category: 1.3.3. Wet.*

caò bu *Category: 1.3.3. Wet.*

swi sǔ *Category: 1.3.3. Wet.*

1.3.3.1. Dry

han pwan dyûn *Category: 1.3.3.1. Dry.*

kan khai *Category: 1.3.3.1. Dry.*

kan mi *Category: 1.3.3.1. Dry.*

lon mon *Category: 1.3.3.1. Dry.*

lon shei *Category: 1.3.3.1. Dry.*

LexiquePro

A - a

- a-₁** *Pfx.* 3rd person pronoun. **atai** he says **pra athabă tapra** he who prays
- a-₂** *Pfx.* pronoun expressing that the thing/quality of the following word is possessed by another. **atai** he says **pra athabă tapra** he who prays
- abakyădò** *N.* concerning.
- abayai** *N.* menstration.
- abá abwai** *ADJ.* complete.
- abá abwi** *ADJ.* complete.
- abaŭ akhă** *ADJ.* white and bright.
- abli abla** *ADJ.* deceitful.
- ablĭ** *ADJ.* clean.
- ablĭ ablăn** *ADJ.* clean.
- abònphu** *N.* change from one state to another.
- abũ ahtõn** *unknown.* ? **abũ ahtõn** ??
- abwan abai** ??
- acă** *CONJ.* **acă na khakrăd că acă na că krăd ka Acă**
- akhauacă** *CONJ.* since.
- akhaŭ ahaĭ** *ADJ.* brave.
- akhein aklan** *N.* time.
- akhein akhađ** *N.* fragment.
- akheinaklan** *N.* season.
- akhraŭ akhraò** *ADJ.* sickly. **ŭ lacŭ akhraŭ akhraò khĭ** ??
- akhu akha** *N.* benefit. **phŭ phăpră akhu akha cõ** **aò pè** ??
- akhŭ acŭ** *N.* lineage.
- alămaòngă** *Pp.* in front of.
- alĭ akhlă** *ADJ.* settled. **tamaŭ alĭ akhlă lă mahalŭ** **sadĭ**
- alon amă** *ADJ.* wrong.
- alou amă** *ADJ.* wrong.
- alu atăn** *N.* merit.
- alŭ anba** *ADI* taste

Thank you

- Questions?
- Palaso (www.palaso.org)
- WeSay (www.WeSay.org)
- Fieldworks (fieldworks.sil.org)
- LexiquePro (www.LexiquePro.org)

A dictionary is never finished, just abandoned