



A tool for sharing interlinearized and lexical data in diverse formats

Daniel Kaufman Raphael Finkel
(ELA & Queens College) (University of Kentucky)

Quick demo

- ▶ Creating an account
- ▶ Searching for a word in Kratylos
- ▶ Searching across files/projects
- ▶ Exporting an example

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons
- ▶ Typically, there is electronic data from multiple scholars on a single language or a group of closely related languages

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons
- ▶ Typically, there is electronic data from multiple scholars on a single language or a group of closely related languages
- ▶ How can this data be tamed? How can other scholars and community members make use of it?

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons
- ▶ Typically, there is electronic data from multiple scholars on a single language or a group of closely related languages
- ▶ How can this data be tamed? How can other scholars and community members make use of it?
- ▶ Archives have not yet provided a solution. (They have enough problems of their own.)

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons
- ▶ Typically, there is electronic data from multiple scholars on a single language or a group of closely related languages
- ▶ How can this data be tamed? How can other scholars and community members make use of it?
- ▶ Archives have not yet provided a solution. (They have enough problems of their own.)
- ▶ There is no widely available system for eliciting feedback and corrections on linguistic data from a community of speakers

The problem

- ▶ Long-term language documentation projects accrue data in different electronic formats.
 - ▶ Audio files and video files
 - ▶ various current and legacy formats for media annotation, interlinearized glossed text and lexicons
- ▶ Typically, there is electronic data from multiple scholars on a single language or a group of closely related languages
- ▶ How can this data be tamed? How can other scholars and community members make use of it?
- ▶ Archives have not yet provided a solution. (They have enough problems of their own.)
- ▶ There is no widely available system for eliciting feedback and corrections on linguistic data from a community of speakers
- ▶ Existing electronic lexicons are still very much based on *print models* despite having none of the constraints of traditional print dictionaries.

Our solution in progress

- ▶ **KRATYLOS**: a website that can store and search over FLEx databases, Shoebox, Toolbox, Praat and ELAN files including associated audio/video media

Our solution in progress

- ▶ KRATYLOS: a website that can store and search over FLEx databases, Shoebox, Toolbox, Praat and ELAN files including associated audio/video media
- ▶ While it is not an archiving facility, it can salvage data trapped in formats that are otherwise difficult to share.

Our solution in progress

- ▶ KRATYLOS: a website that can store and search over FLEx databases, Shoebox, Toolbox, Praat and ELAN files including associated audio/video media
- ▶ While it is not an archiving facility, it can salvage data trapped in formats that are otherwise difficult to share.
- ▶ We demonstrate this with the Wakhi language documentation project.

FLEX (<http://fieldworks.sil.org/flex/>)

- ▶ We are not aiming to replace popular existing software like FLEX. We are aiming to make data from such programs more easily sharable.
- ▶ A bit more about FLEX:
 - ▶ FLEX builds on Shoebox and Toolbox, two older and far simpler pieces of software for creating linguistic databases.
 - ▶ It has wide-ranging features for building sophisticated lexicons + a very detailed approach to interlinearized glossed text.

Wakhi - FieldWorks Language Explorer

English Icelandic (Iceland) Help

File Send/Receive Edit View Data Insert Format Tools Parser Window Help

English Default Paragraph

Texts & Words

- Interlinear Texts
- Concordance
- Complex Concordance
- Word List Concordance
- Word Analyses
- Bulk Edit Wordforms
- Statistics

Texts

Title

Show All

Shaw (Wakhi tale)

Shaw (Forbes Persian stories translated to Wakhi)

Reinhardt Text 2

PKH tsibir vrit

PKH Sayings

PKH satkəkba iwət buj

PKH ji xəðorgvung tu

PKH ji næxtjir tu

PKH ji boj

PKH irək-bə miðir.

PKH i potʃo

PKH hib aqlimər potso tu

PKH bozərgon

GSK tru nasiat

GSK Səyira

GSK mirpritʃ

GSK maj zman

GSK L'Oisillon Enchanté

GSK Le Preux Ziemtchi

Lexicon

Texts & Words

Grammar

Notebook

Lists

25/Jan/2016 Queue: (-/-) No Parser Loaded

Text

Title Wak Shaw (Wakhi tale)

Eng

Info Baseline Gloss Analyze Tagging Print View Text Chart

Free If he says "Why dost thou not come forth?" say "Thou dost not love me; if thou lovest me give me (thy) ning."

5.5 Word jaw ðaj wəzdəj , təm prit

Morphemes jaw ðaj wəz -d -əj ta -m prit

Lex. Entries ja₂ ðaj wəz -t₂ -i₁ ta -m prit

Lex. Gloss DEF/3SG man come PST PST LOC.UP PROX front

Lex. Gram. Info. det n v v:Past1 v:(Pst2) Loc/Dir det:(Proximity) n

Word Gloss *** **

Word Cat. pro n *** Loc/Dir n

Free Her husband came, she went not forth into his presence.

5.6 Word çatəj ki " tʃizərat zɨ

Morphemes çə -t -əj ki tʃiz -ər =ət zɨ

Lex. Entries çan -t₂ -i₁ ki tʃiz -r =ət₂ zɨ

Lex. Gloss say PST PST COMP what DAT 2SG 1SG.GEN

Lex. Gram. Info. v v:Past1 v:(Pst2) comp interrog n:(Outer case) TAM poss

Word Gloss said *** **

Word Cat. v comp *** **

Free He said: "Wherefore camest thou not into my presence?"

5.7 Word ðəjd çatəj

Morphemes ðəjd çə -t -əj

Lex. Entries ðəjd çan -t₂ -i₁

Lex. Gloss daughter say PST PST

Lex. Gram. Info. n v v:Past1 v:(Pst2)

Word Gloss said

5.5 Word	jaw	ðaj	wəzdəj			, təm		prɪt	nə	njə
Morphemes	jaw	ðaj	wəz -d -əj			ta -m		prɪt	nə =	njə
Lex. Entries	ja ₂	ðaj	wəz -t ₂ -i ₁			ta -m		prɪt	nə =	nɪw
Lex. Gloss	DEF/3SG	man	come PST PST			LOC.UP PROX		front	NEG	exit
Lex. Gram. Info.	det	n	v v:Past1 v:(Pst2)			Loc/Dir det:(Proximity)		n	<Not Sure>	v
Word Gloss	***		***			***			***	***
Word Cat.	pro	n	***			Loc/Dir		n	***	v

Free Her husband came, she went not forth into his presence.

5.5	Word	jaw	ðaj	wəzdəj		,	təm		prɪt	nə	njə
	Morphemes	jaw	ðaj	wəz	-d	-əj	ta	-m	prɪt	nə =	njə
	Lex. Entries	ja ₂	ðaj	wəz	-t ₂	-i ₁	ta	-m	prɪt	nə =	nɪw
	Lex. Gloss	DEF/3SG	man	come	PST	PST	LOC.UP	PROX	front	NEG	exit
	Lex. Gram. Info.	det	n	v	v:Past1	v:(Pst2)	Loc/Dir	det:(Proximity)	n	<Not Sure>	v
	Word Gloss	***		***			***		***		***
	Word Cat.	pro	n	***			Loc/Dir		n	***	v

Free Her husband came, she went not forth into his presence.

5.5 Word	jaw	ðaj	wæzdəj	,	təm	prɪt	nə	njə		
Morphemes	jaw	ðaj	wæz	-d	-əj	ta	-m	prɪt	nə =	njə
Lex. Entries	ja ₂	ðaj	wæz	-t ₂	-i ₁	ta	-m	prɪt	nə =	nɪw
Lex. Gloss	DEF/3SG	man	come	PST	PST	LOC.UP	PROX	front	NEG	exit
Lex. Gram. Info.	det	n	v	v:Past1	v:(Pst2)	Loc/Dir	det:(Proximity)	n	<Not Sure>	v
Word Gloss	***	***	***	***	***	***	***	***	***	***
Word Cat.	pro	n	***	***	***	Loc/Dir	n	***	***	v

Free Her husband came, she went not forth into his presence.

wəzdəj

wəz

-d

-əj

wəz

-t₂

-i₁

come

PST

PST

v

v:Past1

v:(Pst2)

The big lacuna (recap)

- ▶ FLEx projects are largely stuck in FLEx - no means of sharing database functions.
- ▶ More generally, little means for complex searches over interlinearized text on the web.
- ▶ Multiple researchers may have data sets on the same language that are difficult to unify.
- ▶ For existing lexicon web apps (e.g. Webonary, Lexique Pro), no possibility for regular expression searches or for searching over multiple languages simultaneously.

An attempt at a solution: Kratylos

- ▶ Supported by NSF DEL Grant #1500753, which includes database work and fieldwork on Wakhi (Iranic; Pakistan, Tajikistan and Afghanistan), Purhepecha (language isolate; Mexico) and Koda (Munda; Bangladesh).
- ▶ Programming work: Raphael Finkel and RA Jiho Noh at the University of Kentucky.
- ▶ Linguistic work: Daniel Kaufman and RAs Husniya Khujamyorova, Daniel Barry, Shamim Ahmed, Lluvia Camacho-Cervantes.

Behind the scenes

- ▶ The web server, Apache2, invokes Perl scripts using the Common Gateway Interface (CGI). The scripts use several standard modules:
 - ▶ CGI (and submodules Carp, Session, and cookie), HTML::Template, Digest(submodules SHA and MD5), JSON, and Unicode::Normalize.
- ▶ Kratylos converts uploaded data, if necessary, into a new XML format. For example, the EAF format, although in XML, is not divided into entries, so Kratylos reformats it into entries, each of which contains all the relevant tiers (such as headword, part of speech, and gloss) and a reference to the media file.

Behind the scenes

- ▶ Kratylos then applies a template to convert the XML into a Qddb (Quick and dirty database) representation.
- ▶ The template is format-specific and coordinates:
 1. the XML fields, described as XPath expressions
 2. the Qddb representation of those fields, which is hierarchical
 3. the formatting that the linear display should employ for those fields, which involves Cascading Style Sheets (CSS).
- ▶ The web pages that Kratylos presents to the user use the Bootstrap and JQuery libraries to format pages. The query results page also contains JavaScript code that converts entries on the fly text.
- ▶ Kratylos maintains a MySQL database coordinating projects with their owners and other information.

Sharing versus archiving

- ▶ Optimally, such a tool should exist within an archive so you have complex searches over a stable set of data.

Sharing versus archiving

- ▶ Optimally, such a tool should exist within an archive so you have complex searches over a stable set of data.
- ▶ But we are approaching the problem from the linguist's perspective rather than the archivist's perspective.

Sharing versus archiving

- ▶ Optimally, such a tool should exist within an archive so you have complex searches over a stable set of data.
- ▶ But we are approaching the problem from the linguist's perspective rather than the archivist's perspective.
- ▶ And hoping to help bridge the gap between archiving and publication.
 - ▶ Citations are important but absolute permanence and completely standard metadata for each piece of data is beyond our reach.

Kratylos (<http://www.kratylos.org>)

- ▶ Table of public files which can be searched over. Logging in reveals your private files.

<input checked="" type="checkbox"/>	Language	Version	Data Sources	Access	Maintainer
<input type="checkbox"/>	Aceh-Chamic	base	Parallel Bible Corpus (paralleltext.info)	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Amuzgoan	base	Amy Bauernschmidt (2013)	public	Jiho Noh (University of Kentucky)
<input type="checkbox"/>	Aramaic	aramaic2	Daniel Kaufman, 28 December 2015	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Aramaic	aramaic1	Daniel Kaufman, 28 December 2015	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Archi	archi2	http://www.philol.msu.ru/~languedoc/eng/archi/corpus.php	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Archi	archi1	Marina Chumakina	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Bakola	base	http://dobes.mpi.nl/projects/bakola/	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Balantak	base	Parallel Bible Corpus (paralleltext.info)	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Bambam	base	Parallel Bible Corpus (paralleltext.info)	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Batak	base	Parallel Bible Corpus (paralleltext.info)	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Bugis	base	Parallel Bible Corpus (paralleltext.info)	public	Raphael Finkel (University of Kentuc
<input type="checkbox"/>	Bukvar	base	Elena Perekhvalskaya, 2/2016	public	Raphael Finkel (University of Kentuc

Showing 1 to 63 of 63 entries
[show project list](#)

Selected Languages:

Select languages above then search a word or a phrase...

Find!

Query Type: String Word Pattern

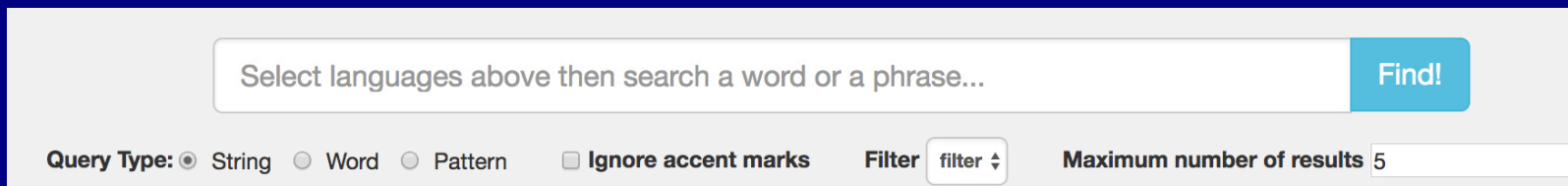
Ignore accent marks

Filter

Maximum number of results

Kratylos (<http://www.kratylos.org>)

- ▶ A search bar which can be set to query full words, strings or regular expressions (“patterns”).
- ▶ Any data field can be targeted by a query. The options are populated automatically by the categories in the data.



Select languages above then search a word or a phrase... **Find!**

Query Type: String Word Pattern Ignore accent marks Filter filter Maximum number of results

Kratylos (<http://www.kratylos.org>)

- ▶ Displaying particular fields can be toggled on and off for different audiences.

Linear display uses formatting to distinguish fields.

gurung (lift) Data source: Recorded from Narayan Gurung of Siklis village in the
Endangered Language Initiative at CUNY Graduate Center. Researchers: Peter Graif,

2014-03-24T14:41:03Z

2014-03-24T14:41:03Z

gvr

tilí k^hora:

stem

en

hog house

Glover, Warren W. 1972. A Vocabulary of the Gurung Language. 6.A.10

Outline display uses text to distinguish fields.

DateCreated 2014-03-24T14:41:03Z

DateModified 2014-03-24T14:41:03Z

Headword

HLanguage *gvr*

HText tilí k^hora:

MorphologicalType *stem*

Sense

Gloss

GLanguage *en*

GlossData hog house

SourceNote Glover, Warren W. 1972. A Vocabulary of the Gurung Language. 6.A.10

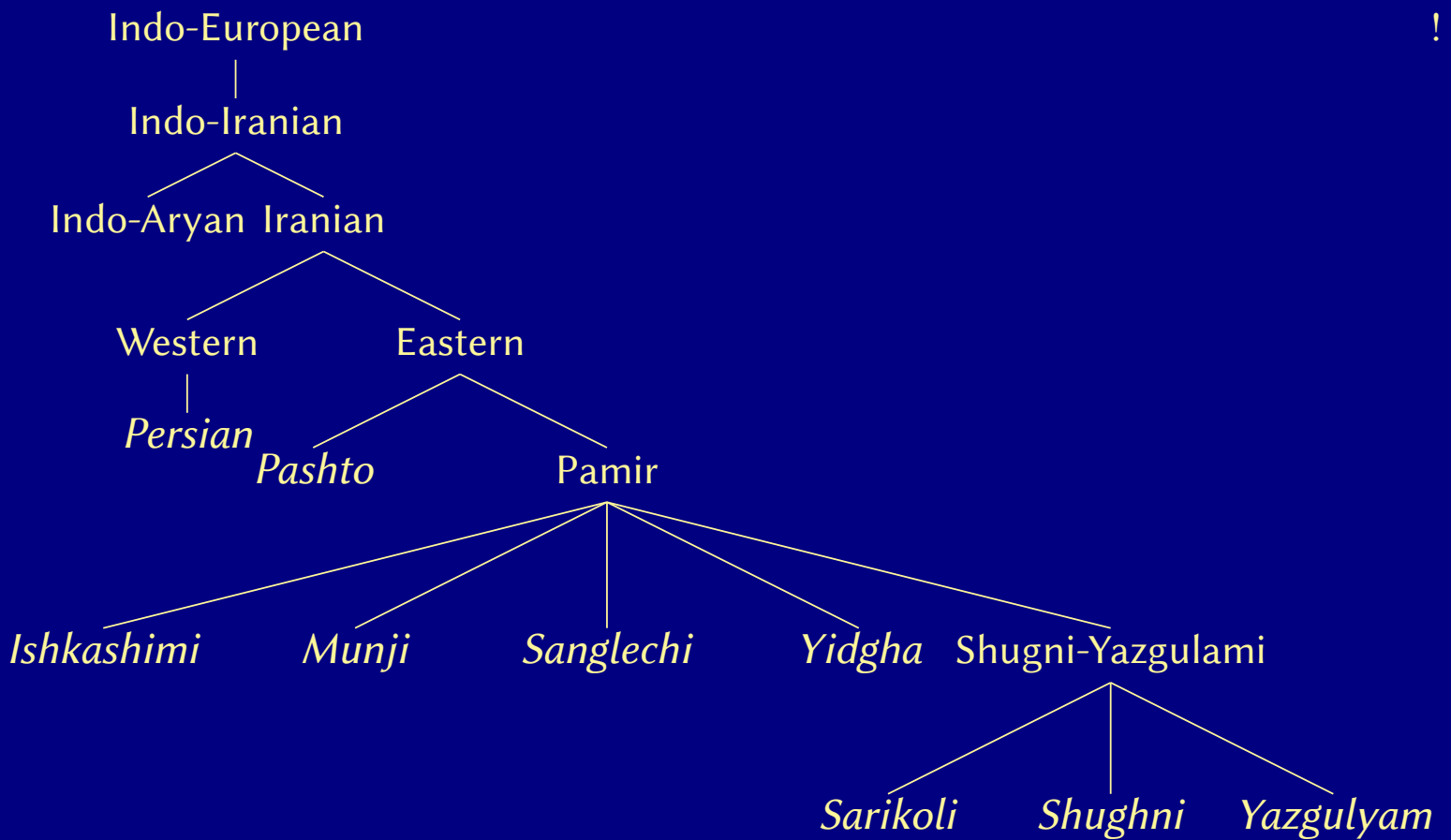
Wakhi project

- ▶ An Iranian language spoken by a small transnational community around the intersection of Afghanistan, Tajikistan, Pakistan and China.

Behind the scenes



- ▶ The relation between Wakhi and the other Iranian languages is still unclear but it is often grouped together with the Pamiri languages (Shughni, Roshani, Bartangi, Oroshori, Sarikoli, Munji, Ishkashimi). Recent work suggests the Pamiri languages constitute a sprachbund rather than a phylogenetic group.



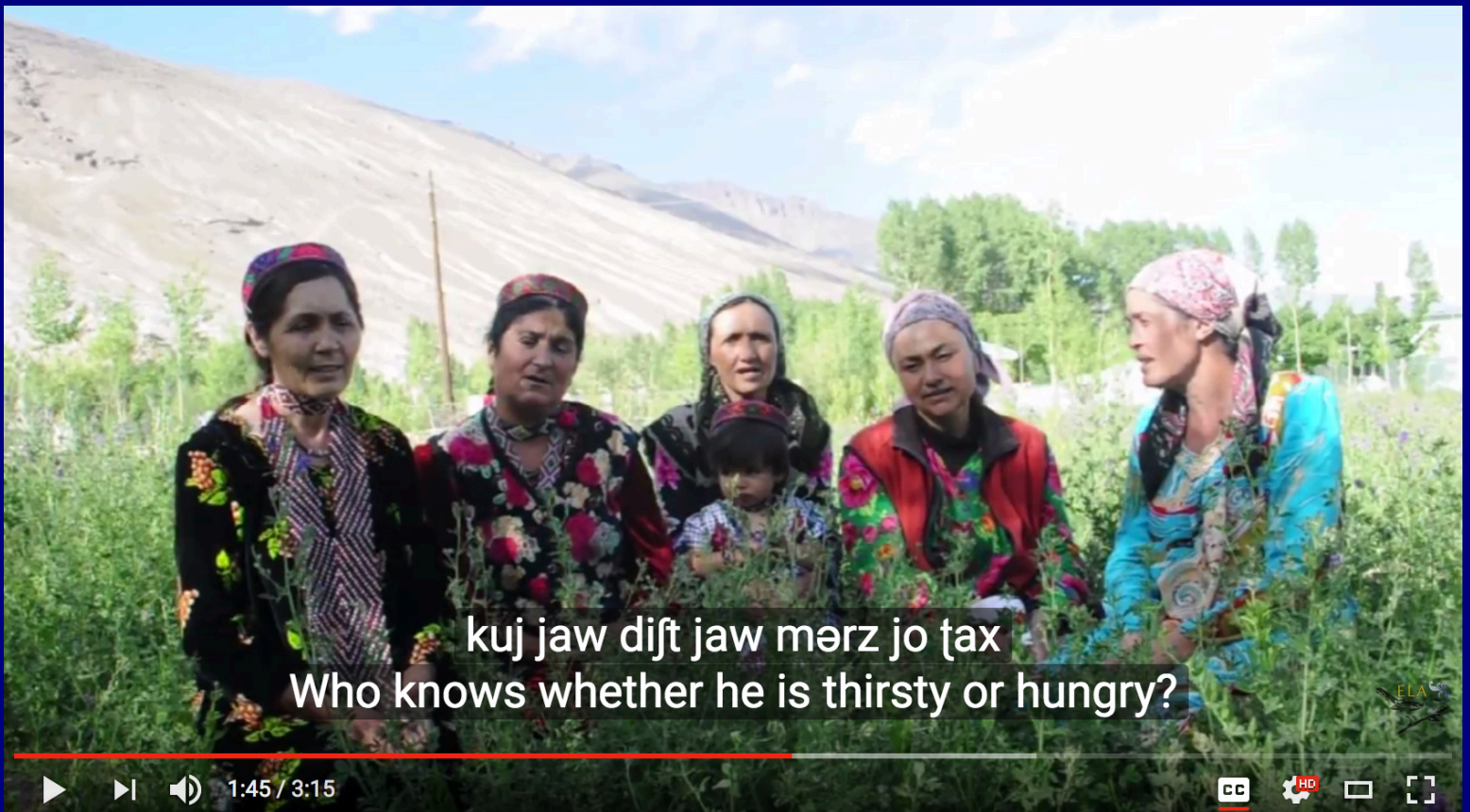
Behind the scenes

Beginnings of the project: A small immigrant community in NYC

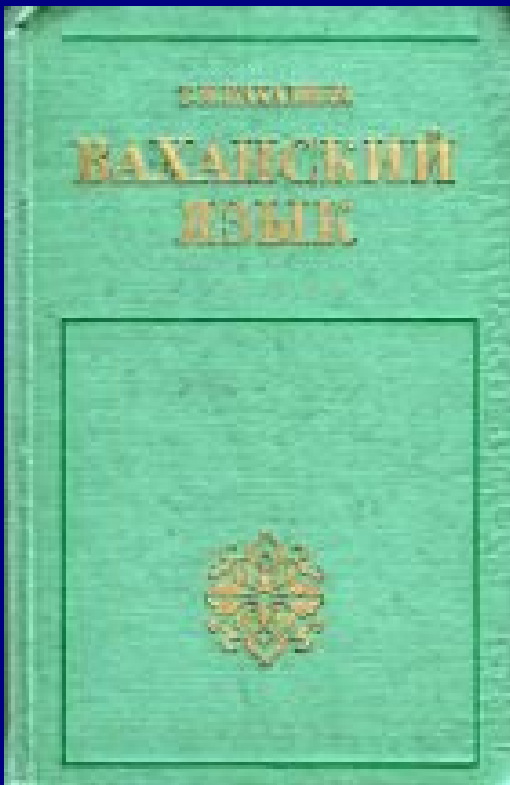


Behind the scenes

Beginnings of the project: From NYC to the Wakhan

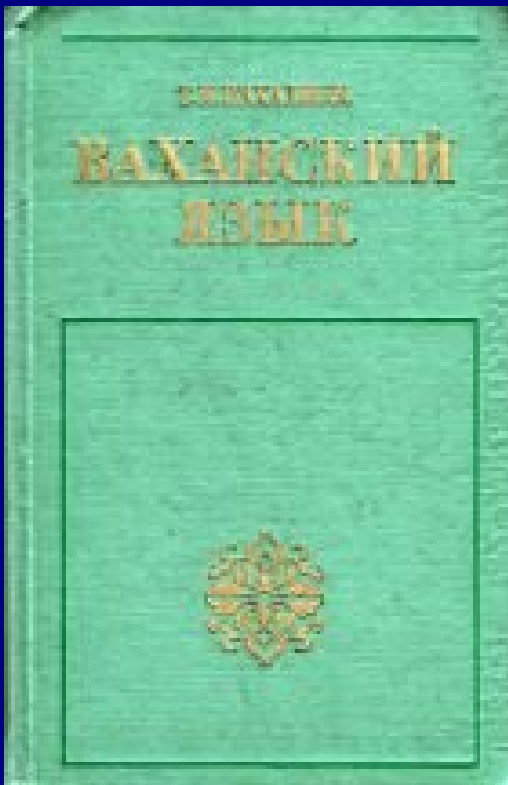


Previous literature



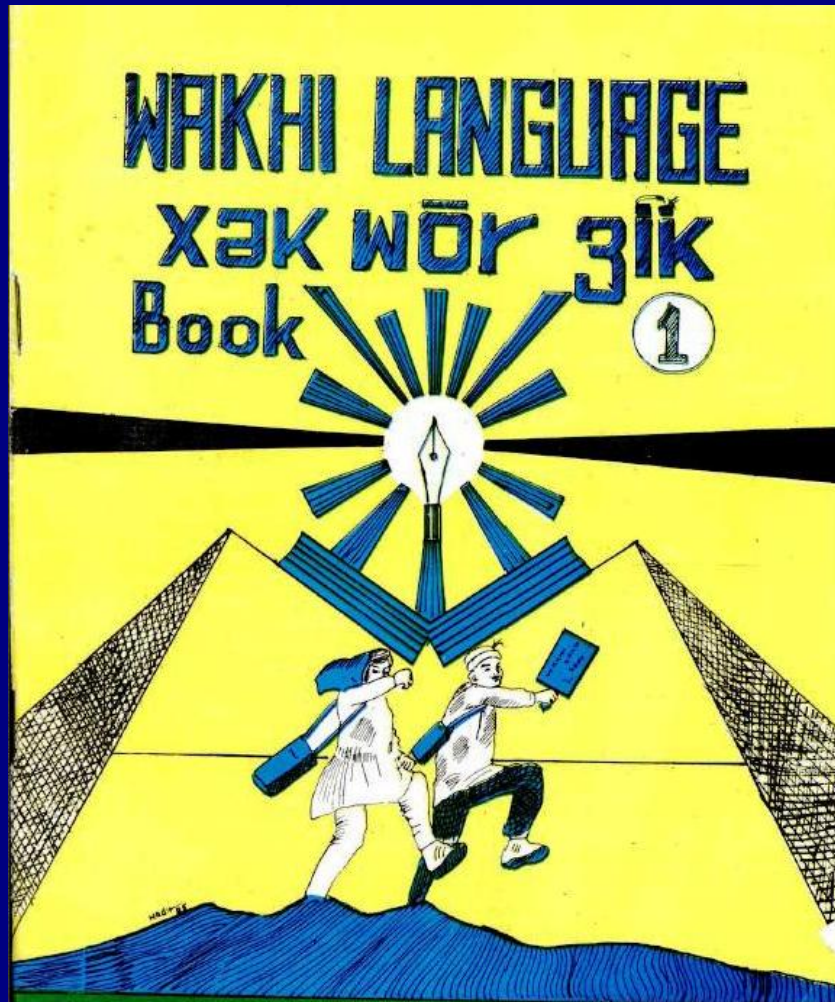
- ▶ Previous work on Wakhi includes Morgenstierne (1938), Lorimer (1958), Pakhalina (1975), Grünberg and Steblin-Kamensky (1988), Bashir (2009), Bashir (1986), Hughes (2011), Reinhold (2006), Steblin-Kamensky (1999), Fuchs (2015), SanGregory (2015).

Previous literature



- ▶ Previous work on Wakhi includes Morgenstierne (1938), Lorimer (1958), Pakhalina (1975), Grünberg and Steblin-Kamensky (1988), Bashir (2009), Bashir (1986), Hughes (2011), Reinhold (2006), Steblin-Kamensky (1999), Fuchs (2015), SanGregory (2015).
- ▶ For our purposes, Pakhalina (1975) and Grünberg and Steblin-Kamensky (1988) are the most crucial because of the folktales, narratives and songs they recorded. These are inaccessible to the Wakhis of Afghanistan and Pakistan because they are written in Russian.

Wakhi writing systems



Wakhi writing systems

A a	B b	C c	Č č	Ĉ ĉ	D d	Ḑ ḑ	Δ δ	E e	Ə ə
А а	Б б	Ц ц	Ч ч	Ч̣ ч̣	Д д	Ḑ ḑ	Ḑ ḑ	Е е	Ə ə
[a]	[b]	[t͡s]	[t͡ʃ]	[t͡ʃ̣]	[d]	[ḑ]	[ð]	[e]	[ə]
F f	G g	Y y	Ÿ ŷ	H h	I i	J j	ǰ ǰ	K k	L l
Ф ф	Г г	Ғ ғ	Ӏ Ӏ	Х х	И и	Ҷ Ҷ	Ҷ Ҷ	К к	Л л
[f]	[g]	[ɣ]	[y]	[h]	[i]	[d͡ʒ]	[d͡ʒ̣]	[k]	[l]
M m	N n	O o	P p	Q q	R r	S s	Š š	Ṣ ṣ	T t
М м	Н н	О о	П п	Қ қ	Р р	С с	Ш ш	Ṣ ṣ	Т т
[m]	[n]	[o]	[p]	[q]	[r]	[s]	[ʃ]	[ṣ]	[t]
Ṭ ṭ	Θ θ	U u	V v	W w	X x	Ẃ ẃ	Y y	Z z	Ž ž
Ṭ ṭ	Ӧ Ӧ	У у	В в	Ӱ Ӱ	Х х	Ẃ ẃ	Й й	З з	Ж ж
[ṭ]	[θ]	[u]	[v]	[w]	[χ]	[x]	[j]	[z]	[ʒ]
Ž ž	З з	Ы ы							
Ж ж	З з	Ы ы							

Wakhi writing systems

ش	چ	چ	چ	ج	ث	ت	ت	پ	ب	آ	ا
[ts]	[ts]	[tʃ]	[dʒ]	[dʒ]	[θ]	[t]	[t]	[p]	[b]	[o]	[a/Ø]
س	ژ	ژ	ژ	ز	ر	ذ	د	د	خ	ح	خ
[s]	[z]	[ʒ]	[ʒ]	[z]	[r]	[ð]	[d]	[d]	[χ]	[h]	[dʒ]
ق	ڤا	ڤا	ڤا	ع	ظ	ط	ض	ص	ڤس	ش	ڤس
[q]	[v]	[f]	[ɸ]	∅	[z]	[t]	[z]	[s]	[ʃ]	[ʃ]	[x]
ی	ی	و	و	و	ه	ن	م	ل	گا	کا	
[j/e]	[i]	[i]	[u]	[w/ə]	[h]	[n]	[m]	[l]	[g]	[k]	

Wakhi writing systems

6. Rahim Khon – PM000000120000001530 7, 2008



Xhik Murdumver slom; Ghafch bef yark oghozve xhetk, zu ghenaver saver muboraki. Zunini her yi xhik murdumer yi arz woz swol, cum batkenep sak sach vir re samander mulung wozan din woz dak vanden ki koipe wizat xe dhastep saker rashufd, wazet oghoz ceren, terleman wocn. Wodh ki nei peghrep sak xath xath te xhu dhasten nusun. her yiu shenoxti yo zik, wazeth xhath ce neshaken shetraxven.

Salam to Wakhi people around the globe, very nice effort is going on, many congratulation to all of you. I have a request and quest from each Wakhi people, For how long we will swim alone in the Ocean and wait for others to lend their hand to us for help? Lets come together and start a sincere effort. If not today tomorrow we will kill ourselves with our own hands. Language is identity of any individual, lets come forward and

Exploring Wakhi morphosyntax with regular expressions

- ▶ Fuchs (2015) shows that certain clitics in Wakhi can be doubled and tripled within a single clause:

(1) wuz = əm ç̣i şew-i = m ş̣kəndevdi
 1SG.NOM=1SG SELF.POSS horns-ACC=1SG break.PST
 ‘I broke my horns.’ (Fuchs 2015:151)

(2) jan də səḅıq = əş lup = əş ç̣at-i ḳı nəyardum ətʃ waxt də
 then in old.time=IPFV big=IPFV say-PST COMP bear no time with
 məltıq-ən ja-r q̣rib mə rəʃ̣-əv
 shotgun-ABL 3SG-DAT near NEG.IMP go-2PL.SAGR
 ‘Then in old times the old people said: Bears, never go near them (not even) with a shotgun.’ (SanGregory 2015:7).

Exploring Wakhi morphosyntax with regular expressions

- ▶ Fuchs (2015) shows that certain clitics in Wakhi can be doubled and tripled within a single clause:
- ▶ She also shows that the positioning of the clitics may depend on focus and other semantic factors (see also, Erschler 2010; Beck 2013; Hughes 2011; Bashir 1986).

(1) wuz = əm ç̣i şew-i = m ş̣kəndevdi
 1SG.NOM=1SG SELF.POSS horns-ACC=1SG break.PST
 ‘I broke my horns.’ (Fuchs 2015:151)

(2) jan də səḅıq = əş lup = əş ç̣at-i ḳı nəyardum əṭʃ waxt də
 then in old.time=IPFV big=IPFV say-PST COMP bear no time with
 məltıq-ən ja-r q̣rib mə rəṭʃ-əv
 shotgun-ABL 3SG-DAT near NEG.IMP go-2PL.SAGR
 ‘Then in old times the old people said: Bears, never go near them (not even) with a shotgun.’ (SanGregory 2015:7).

Exploring Wakhi morphosyntax with regular expressions

- ▶ Fuchs (2015) shows that certain clitics in Wakhi can be doubled and tripled within a single clause:
- ▶ She also shows that the positioning of the clitics may depend on focus and other semantic factors (see also, Erschler 2010; Beck 2013; Hughes 2011; Bashir 1986).

(1) wuz =ə**m** ç*i* şew-i =**m** şkəndevdi
 1SG.NOM=1SG SELF.POSS horns-ACC=1SG break.PST
 ‘I broke my horns.’ (Fuchs 2015:151)

(2) jan də səbıq =ə**ş** lup =ə**ş** çat-i k*i* nəyardum ətʃ waxt də
 then in old.time=IPFV big=IPFV say-PST COMP bear no time with
 məltıq-ən ja-r qrib mə rəʃ-ə**v**
 shotgun-ABL 3SG-DAT near NEG.IMP go-2PL.SAGR
 ‘Then in old times the old people said: Bears, never go near them (not even) with a shotgun.’ (SanGregory 2015:7).

Exploring Wakhi morphosyntax with regular expressions

- ▶ Looks for all examples with a PROGRESSIVE clitic...
 - utterance-initial: `^=PROG`
 - after the first word: `^\b\w+\b =PROG`
 - after the second word: `^\b\w+\b \b\w+\b =PROG`
 - after the third word: `^\b\w+\b \b\w+\b \b\w+\b =PROG`
 - in multiple positions: `=PROG.*=PROG`

Exploring Gurung phonology with regular expressions

- ▶ Looking for clusters: `[pɓtdkgwɟ][pɓtdkgwɟ]`
- ▶ Looking for geminates: `(.)\1`
- ▶ Looking for reduplication: `(.!*)(.!*)\g1\g2`
- ▶ Comparing frequencies of ND and NT clusters: `[mnŋ][pɓtk]` vs. `[mnŋ][bdg]`
- ▶ Looking for post-stopped nasals with following nasal vowels: `[mnŋ][bdg]~`
- ▶ Looking for nasals with following oral vowels: `[mnŋ][^~\swjrd]`

An open challenge

“Apart from technical challenges, there is also an important sociological challenge to create maximally interoperable language analysis software. To imagine this can be done simply by adopting common file formats, or by operating an in-house software development lifecycle using project funds, or by invoking the XML family of buzzwords is to misunderstand the nature of the problem. Instead, we need to foster new research collaborations involving computational linguists and field linguists, leading to new understanding about how to collect and analyze corpora of data from endangered languages.”

(Bird 2009:473)

Behind the scenes

Thank you!



Kaufman&Finkel

Kratylos

March 2, 2017

31 / 48

Bibliography

- Bashir, Elena. 1986. Beyond split-ergativity: Subject marking in Wakhi. In Chicago Linguistic Society, Papers from the General Session at the Twenty-Second Regional Meeting, 14–35. Chicago.
- Bashir, Elena. 2009. Wakhi. In The Iranian Languages, ed. Gerhardt Windfuhr, 825–862. Routledge.
- Beck, Simone. 2013. Backgrounding and foregrounding, prominence, and highlighting in Afghan Wakhi. Master's thesis, University of Gloucestershire, Cheltenham, UK.
- Bird, Steven. 2009. Natural language processing and linguistic fieldwork. Computational Linguistics 35:469–474.
- Erschler, David. 2010. On optionality in grammar: The case of East Iranian almost Wackernagel clitics. Handout from Syntax of the World's Languages IV, Lyon.
- Fuchs, Zuzanna. 2015. Second position and “floating” clitics in Wakhi. In Proceedings of the Annual Meeting of the Berkeley Linguistics Society, ed. Anna E. Jurgensen, Hannah Sande, Spencer Lamoureux, Kenny Baclawski, and Alison Zerbe, volume 41, 133–154. Berkeley: Berkeley Linguistics Society. URL <http://escholarship.org/uc/item/2965n622>.
- Grünberg, Aleksander Leonovich, and Ivan M. Steblin-Kamensky. 1988. La langue Wakhi, volume 1: Corpus de littérature orale. Paris: Maison des Sciences de l'Homme.
- Hughes, Todd R. 2011. Wakhi agreement clitics. Master's thesis, University of Florida.
- Lorimer, David Lockhart Robertson. 1958. The Wakhi Language, volume I (Introduction, Phonetics and Texts). London: School of Oriental and African Studies, University of London.
- Morgenstierne, Georg. 1938. Indo-Iranian frontier languages, 2: Iranian Pamir languages. Oslo: H. Aschehoug & Co. (W. Nygaard).
- Pakhalina, Tatyana N. 1975. Vaxanskij jazyk. Moscow: Nauka.
- Reinhold, Beate. 2006. Neue Entwicklungen in der Wakhi-Sprache von Gojal (Nordpakistan). Weisbaden: Harrasowitz Verlag.
- SanGregory, Erin. 2015. Towards a new analysis of Wakhi clitics. Occasional Papers in Applied Linguistics.
- Steblyn-Kamensky, Ivan M. 1999. Etimologičeskij slovar' vachanskogo jazyka [Etymological dictionary of the Wakhi language]. St. Petersburg: Vostokoved.