

Typesetting language data using ConTeXt

Jelle Huisman - SIL International

E16 & DEtool

EuroTeX2009 & 3rd ConTeXt Meeting



Collecting and publishing language data

- collect and publish: how?
- general information: Ethnologue
- language specific information: dictionary, grammar description
- linguistic analyses: phonetics, grammatical structures, sociolinguistics



Ethnologue: general

- Handbook describing all the world's languages
- 16th edition was published in 2009
 - Part 0: Intro, statistical summaries (50 p.)
 - Part 1: 6909 language descriptions (600 p.)
 - Part 2: Language maps (200 p.)
 - Part 3: Indexes (400 p.)



Ethnologue: data flow

- field linguists collect data on PC based systems
- all data is stored in Oracle database on secure web server
- different output paths with XSLT
 - web version
 - $\text{T}_{\text{E}}\text{X}$ based output for book publication
- download and typeset using Con $\text{T}_{\text{E}}\text{X}$ t it locally (UK/US)



Ethnologue: typesetting

```
\startmode[proofreading] % special layout for proofreading mode
\setuppapersize [letter][letter] % paper size for proofreading mode
\setuplayout[backspace=18mm,
  width=160mm,
  topspace=7mm, top=0mm,
  header=16mm, footer=6mm,
  height=250mm]
\stopmode
```



Ethnologue: typesetting

In project-file:

```
\enablemode[book]  
%\enablemode[proofreading]
```

Proofreading:

- is done by editorial staff in Dallas,
- at least 2000 pages to proofread all language descriptions

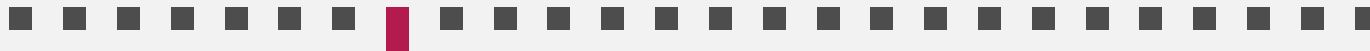


Statistical Summaries (multipage tables)

Table 6. Distribution of living languages by country (continued)

Country	Living languages				Number of speakers			Diversity	
	Total	Percent	Indigenous	Immigrant	Count	Mean	Median	Index	Coverage
Cameroon	279	4.04	278	1	10,751,460	39,968	8,500	0.946	96%
Canada	169	2.45	86	83	33,399,367	219,733	6,330	0.599	90%
Cape Verde Islands	2	0.03	2	0	408,800	204,400	14,800	0.070	100%
Cayman Islands	3	0.04	1	2	43,010	14,337	20,000	0.547	100%
Central African Republic	82	1.19	71	11	3,478,795	46,384	15,000	0.959	91%
Chad	133	1.93	131	2	6,573,351	54,325	12,000	0.944	91%
Chile	16	0.23	9	7	14,049,503	1,561,056	3,390	0.035	56%
China	296	4.28	292	4	1,211,179,642	4,190,933	28,000	0.509	98%
Colombia	83	1.20	80	3	34,596,247	449,302	1,500	0.034	93%
Comoros	7	0.10	6	1	594,100	99,017	1,700	0.545	86%
Congo	66	0.96	62	4	3,942,950	63,596	9,600	0.836	94%
Cook Islands	6	0.09	5	1	12,440	2,073	840	0.602	100%
Costa Rica	13	0.19	10	3	3,385,390	338,539	4,500	0.050	77%
Côte d'Ivoire	93	1.35	77	16	9,220,990	115,262	26,400	0.917	86%
Croatia	22	0.32	7	15	4,488,450	213,736	4,710	0.211	95%
Cuba	4	0.06	2	2	10,003,700	5,003,700	3,500	0.001	90%
Cyprus	6	0.09	4	2	870,040	217,510	2,740	0.331	67%
Czech Republic	20	0.29	10	10	10,019,810	527,318	10,900	0.146	95%
Democratic Republic of the Congo	217	3.14	215	2	40,099,030	199,498	26,000	0.948	93%
Denmark	13	0.19	7	6	5,606,300	566,630	10,000	0.055	77%
Djibouti	10	0.14	5	5	484,900	53,878	10,200	0.571	86%
Dominica	3	0.04	3	0	52,800	17,600	10,000	0.313	100%
Dominican Republic	8	0.12	4	4	7,083,590	1,180,283	8,400	0.053	75%
East Timor	19	0.28	19	0	531,110	29,528	20,000	0.897	95%
Ecuador	75	0.36	29	46	11,347,403	463,317	8,000	0.267	86%
Egypt	27	0.39	11	16	82,130,100	3,422,088	116,000	0.576	89%
El Salvador	7	0.10	5	2	5,912,830	1,382,564	500	0.004	71%
Equatorial Guinea	14	0.20	14	0	393,439	30,265	5,000	0.417	93%
Eritrea	18	0.26	12	6	4,540,300	412,735	100,000	0.627	61%
Estonia	18	0.26	2	16	1,402,170	82,481	1,390	0.454	94%
Ethiopia	88	1.27	85	3	58,433,608	695,638	32,100	0.864	95%
Falkland Islands	1	0.01	1	0	1,990	1,990	1,990	0.000	100%
Fiji	21	0.30	10	11	810,100	62,315	5,300	0.668	62%
Finland	23	0.33	12	11	5,102,260	300,133	4,500	0.148	74%
France	62	0.90	23	39	62,236,150	1,682,058	70,000	0.267	60%
French Guiana	15	0.22	12	3	68,490	6,236	800	0.450	73%
French Polynesia	11	0.16	9	2	192,790	19,279	3,400	0.596	91%
Gabon	43	0.62	42	1	1,275,110	31,300	6,000	0.762	95%

Country	Living languages				Number of speakers			Diversity	
	<i>Total</i>	<i>Per-cent</i>	<i>Indige-nous</i>	<i>Immi-grant</i>	<i>Count</i>	<i>Mean</i>	<i>Median</i>	<i>Index</i>	<i>Cover-age</i>
Cameroon	279	4.04	278	1	10,751,460	39,968	8,500	0.946	96%
Canada	169	2.45	86	83	33,399,367	219,733	6,330	0.599	90%
Cape Verde Islands	2	0.03	2	0	408,800	204,400	14,800	0.070	100%
Cayman Islands	3	0.04	1	2	43,010	14,337	20,000	0.547	100%
Central African Republic	82	1.19	71	11	3,478,795	46,384	15,000	0.959	91%
Chad	133	1.93	131	2	6,573,351	54,325	12,000	0.944	91%
Chile	16	0.23	9	7	14,049,503	1,561,056	3,390	0.035	56%
China	296	4.28	292	4	1,211,179,642	4,190,933	28,000	0.509	98%
Colombia	83	1.20	80	3	34,596,247	449,302	1,500	0.034	93%
Comoros	7	0.10	6	1	594,100	99,017	1,700	0.545	86%
Congo	66	0.96	62	4	3,942,950	63,596	9,600	0.836	94%
Cook Islands	6	0.09	5	1	12,440	2,073	840	0.602	100%



Language description (6909)

409

Ethnologue 16 - data February 13, 2009 - page 409

409

Asia: Pakistan

Ethnologue

409

Farsi, Western [pɛr] (Persian), 25,000 in Oman (1993). Scattered coastal cities, Maltrah area, Jibrow community, toward Kowl. *Lg Use:* Many women speak only Farsi. Men who go outside the community speak Arabic as L2, and many know 2 or 3 other languages. Other: Muslim (Shi'a). See main entry under Iran.

Harasati [has] (Haras' Akort, Harsyat), 600. South central Oman, Dhofar Province, Jiddat al Harasit. *Class:* Afro-Asiatic, Semitic, South, South Arabian. *Dialects:* similar to Mehri [gɛk], but usually considered a separate language. *Lg Use:* Reportedly increasing use and proficiency in Mehri; bilingual in Arabic. Other: Spoken by the Harasit and Har (Afar). *Map:* 818.

Hobeyfi [hob] (Hobeyfi, Hob), 100 in Oman (1998 H, Matrafi). Near Yemen border. Also in Yemen. *Class:* Afro-Asiatic, Semitic, South, South Arabian. *Dialects:* related to Mehri [gɛk], Shehri [shv]. Possibly a mixed language from Shehri and Mehri. Other: Define themselves as belonging to the Maltrah tribe. *Muslim.* *Map:* 818.

Kumzari [zum] (Kumzari), 1,700 (1993 census). North Oman, Musandam Peninsula. *Class:* Indo-European, Indo-Iranian, Iranian, Western, Southwestern, Lari. Other: Distinct from Khansari [kfm] of Iran. Fishermen. *Muslim.* *Map:* 818.

Luwai [lwa], 5,000 (1996). Maltrah walled quarter, facing the old harbor; Muscat and other cities. *Class:* Indo-European, Indo-Iranian, Unclassified. Other: Are called 'Lushaiya', their ancestors reportedly migrated from Iran to Hyderabad, then to Maltrah. Sometimes called 'Persians'. Businessmen. *Muslim (Shi'a).* *Map:* 818.

Mehri [gɛk] (Mahr), 20,800 in Oman (2000). South Oman near Yemen border. *Dialect:* Nagdi. Other: *Muslim.* See main entry under Yemen. *Map:* 818.

Shehri [shv] (Shkhi, Gabbit, Jhali, Jibbali, Qarawi, Sehri, Shahri, Sheret), 25,000 (1993 census). Dhofar, mountains north of Al Gaidat. *Class:* Afro-Asiatic, Semitic, South, South Arabian. *Dialects:* central Jibbali, Eastern Jibbali, Western Jibbali. Eastern Jibbali includes Kurra Muria ('Baby' Jibbali). Reportedly increasingly bilingual in Dhofari Arabic [ad]. Other: Spoken by Qara (Ishkeli, Akhli), Shuhra (Sheri, Shahara), Barahama, Rut, Ash-Shaik, and some Batahira. *Muslim.* *Map:* 818.

Pakistan

Islamic Republic of Pakistan, Islami Jamhuriya-e Pakistan, 138,000,000. National or official languages: Urdu, Sindhi, English. *Literacy rate:* 26%. *Immigrant languages:* Indian Sign Language, Farsi (20,000), Southern Uzbek (50,000), Turkish, Uyghur. Also includes Arabic (122,000), Chinese (6,000). Information mainly from J. Addison 1986; C. Manca 1991; G. Morgenthauser 1974; C. O'Leary 1992; J. R. Payne 1987; R. C. Shackle 1979, 1980; J. C. Sharma 1985; F. Strand 1973. *Blind population:* 1,500,000. *Deaf population:* 7,200,120. *Deaf institutions:* 71. The number of individual languages listed for Pakistan is 72. Of those, all are living languages. See maps beginning on page 816.

Aer [aer], 150 (1998). Women monolingual. 333 around Deh. Lower Sindhi, jirkro (ooh) near Kamri around Deh, Hyderabad, and at Jamesabad. Others reportedly migrated to India at Partition in 1947, living in the Kachi Bilg area in Gujarat. *Class:* Indo-European, Indo-Iranian, Indo-Aryan, Central zone, Gujarati. *Dialects:* jirkro (ooh) Aer, Jamesabad Aer. *Lexical similarity:* 78% with Kachi Meghwar and Kachi Bihl--both dialects of Kachi Koli [gɛk]. 79%-77% with the Rabari dialect of Kachi Koli [gɛk]. 76% with Kachi Koli [gɛk]. *Lg Use:* All ages. Also use Sindhi [sɛd], Western Punjabi [pɛb], or Gujarati. 100% of boys and 25% of girls attend Sindhi medium

schools. *Lg Dev:* *Literacy rate in L1:* 15% in Sindhi. *Writing:* Arabic script. Other: unusual interrogative word suggests possible historical connection with Western Rajasthani group. Speakers in Pakistan running out of marriage possibilities and may have to move to India. The Indian group is most influential. Other Aer people in Nawshahab. Sindhi reportedly speak a different language, dress differently, and do not intermarry with this group. *Hindi.* *Map:* 817.

Badeshi [bdz], ethnic population: 2,825 (2000). Bishgram (Chail) Valley upper reaches, east of Madyan, Swat Kohistan. 1 village. *Class:* Indo-European, Indo-Iranian. Unclassified. *Lg Use:* Also use Pashto. Other: Existence unconfirmed. The Torwali say they speak Ushajo, and the Ushajo say they speak Torwali. May be a family name of people who came from Badakhshan. *Muslim (Sunni).* *Map:* 816/25.

Bagri [bɔg] (Bagari, Bagria, Bagris, Bahgri, Baoria, Bawri), 200,000 in Pakistan, 100,000 in Sind Province (1998). Sind Province, Punjab. *Lg Use:* Also use Sindhi [sɛd] or Urdu [urd]. *Lg Dev:* *Literacy rate in L1:* Below 1%. *Literacy rate in L2:* Below 9%. Other: Distinct from Vagari [vɔg]. Nomadic. See main entry under India. *Map:* 817.

Balochi, Eastern [bɔl] (Baloci, Balochi, Baluc, Eastern Hill Balochi), 1,000,000 in Pakistan (1999), 5,000,000 including L2 users of all Balochi languages. *Population total all countries:* 1,800,000. Northwest Balochistan Province, northwest Sind, southwest Punjab. Also in India. *Class:* Indo-European, Indo-Iranian, Iranian, Western, Northwestern, Balochi. A member of macrolanguage Baluchi [bal]. *Lg Dev:* *Literacy rate in L1:* Below 1%. *Literacy rate in L2:* 9%-19%. *Radio programs, Bible portions:* 1815-1906. *Writing:* Arabic script, Nastaliq style. Other: A major language in Pakistan. Distinct from Western Balochi [bɔp] and Southern Balochi [bc]. A small body of literature. *Muslim (Sunni).* *Map:* 817.

Balochi, Southern [bc] (Baloci, Baluchi, Baluc, Makran), 2,770,000 in Pakistan (1998). *Population total all countries:* 3,405,000. South Balochistan, south Sind, Karachi. Also in Iran, Oman, United Arab Emirates. *Class:* Indo-European, Indo-Iranian, Iranian, Western, Northwestern, Balochi. A member of macrolanguage Baluchi [bal]. *Dialects:* Coastal Balochi, Kechi (Kechi, Makran (Jutami). Distinct from Eastern Balochi [bɔp] and fairly distinct from Western Balochi [bɔp]. *Lg Dev:* *Literacy rate in L1:* Below 1%. *Literacy rate in L2:* 9%-15%. *Radio programs, NTI:* 1999. *Writing:* Arabic script, Nastaliq style. Other: *Muslim (Sunni).* *Map:* 817.

Balochi, Western [bɔp] (Baloci, Baluchi, Baluc), 1,120,000 in Pakistan (1998). *Population total all countries:* 1,893,400. Northwest Balochistan Province. Also in Afghanistan, Iran, Tajikistan, Turkmenistan. *Class:* Indo-European, Indo-Iranian, Iranian, Western, Northwestern, Balochi. A member of macrolanguage Baluchi [bal]. *Dialects:* Lashari, Rakhtani (Kashkani), Sarwani. Strongly influenced by Farsi, but not intelligible with Farsi [pɛr]. *Lg Dev:* *Literacy rate in L1:* 18-59%. *Literacy rate in L2:* 9%-15%. *Newspapers, Radio programs, Bible portions:* 1984. *Writing:* Arabic script, Nastaliq style, used in Afghanistan. Other: Balochi is the official spelling in Pakistan. A small body of literature. *Muslim (Sunni).* *Map:* 817.

Balti [bt] (Baltistani, Bhotia of Baltistan, Shukli), 270,000 in Pakistan (1992). *Population total all countries:* 308,800. Primarily northeast Pakistan, Baltistan District, Skardu, Rondo, Shigar, Khaplu, Kharmang, and Gilgari valleys. Also in India. *Class:* Sino-Tibetan, Tibeto-Burman, Himalayish, Tibeto-Kanari, Tibetic, Tibetan, Western. *Dialects:* Chorbak. Is most divergent dialect. *Lexical similarity:* 87%-100% among dialects, 78%-85% with Purik [pɛr]. *Lg Use:* Some also use Shina [sc] or Urdu [urd]. *Lg*

409

E16 updating - 5/24 - On/Off: E16 mobile version - February 13, 2009

409

Farsi, Western [pes] (Persian). 25,000 in Oman (1993). Scattered coastal cities. Mutrah area, Jabroo community, toward Ruwi. *Lg Use*: Many women speak only Farsi. Men who go outside the community speak Arabic as L2, and many know 2 or 3 other languages. *Other*: Muslim (Shi'a). See main entry under Iran.

Harsusi [hss] (Harsi 'Aforit, Hersyet). 600. South central Oman, Dhofar Province, Jiddat al-Harasis. *Class*: Afro-Asiatic, Semitic, South, South Arabian. *Dialects*: Similar to Mehri [gdq], but usually considered a separate language. *Lg Use*: Reportedly increasing use and proficiency in Mehri; bilingual in Arabic. *Other*: Spoken by the Harasis and 'Ifar ('Afar). *Map*: 818.

Hobyót [hoh] (Hewbyót, Hobi). 100 in Oman (1998 H. Mutzafi). Near Yemen border. Also in Yemen. *Class*: Afro-Asiatic, Semitic, South, South Arabian. *Dialects*: Related to Mehri [gdq], Shehri [shv]. Possibly a mixed language from Shehri and Mehri. *Other*: Define themselves as belonging to the Mahra tribe. Muslim. *Map*: 818.

Kumzari [zum] (Kunzai). 1,700 (1993 census). North Oman, Musandam Peninsula. *Class*: Indo-European, Indo-Iranian, Iranian, Western, Southwestern, Luri. *Other*: Distinct from Khunsari [kfm] of Iran. Fishermen. Muslim. *Map*: 818.

Luwati [luv]. 5,000 (1996). Mutrah walled quarter, facing the old harbor; Muscat and other cities. *Class*: Indo-

schools. *Lg Dev*: Literacy rate in L2: 15% in Sindhi. *Writing*: Arabic script. *Other*: Unusual interrogative word suggests possible historical connection with Western Rajasthani group. Speakers in Pakistan running out of marriage possibilities and may have to move to India. The Indian group is most influential. Other Aer people in Nawabshah, Sindh reportedly speak a different language, dress differently, and do not intermarry with this group. Hindu. *Map*: 817.

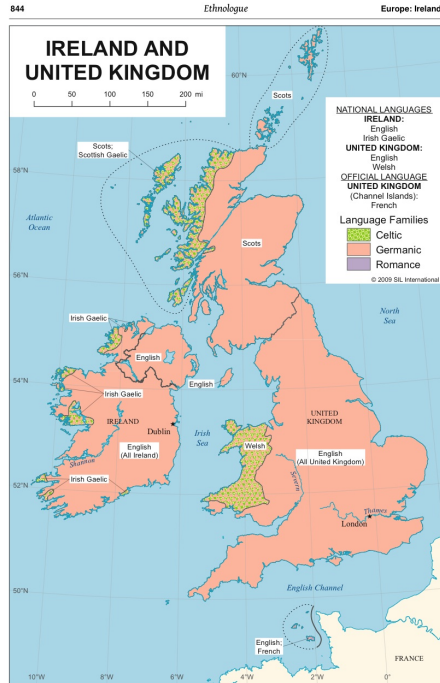
Badeshi [bdz]. Ethnic population: 2,825 (2000). Bishigram (Chail) Valley upper reaches, east of Madyan, Swat Kohistan. 1 village. *Class*: Indo-European, Indo-Iranian, Unclassified. *Lg Use*: Also use Pashto. *Other*: Existence unconfirmed. The Torwali say they speak Ushojo, and the Ushojo say they speak Torwali. May be a family name of people who came from Badakhshan. Muslim (Sunni). *Map*: 816:25.

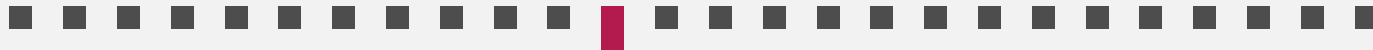
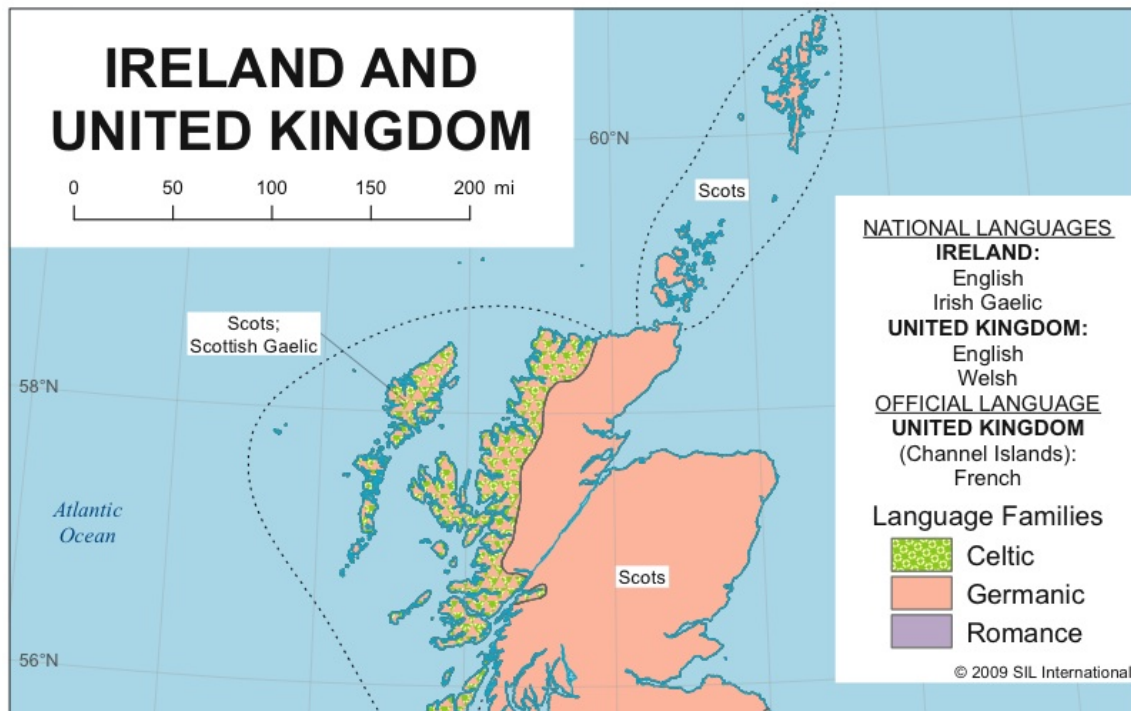
Bagri [bgq] (Bagari, Bagria, Bagris, Bahgri, Baorias, Bawri). 200,000 in Pakistan. 100,000 in Sind Province (1998). Sindh Province; Punjab. *Lg Use*: Also use Sindhi [snd] or Urdu [urd]. *Lg Dev*: Literacy rate in L1: Below 1%. Literacy rate in L2: Below 5%. *Other*: Distinct from Vaghri [vgr]. Nomadic. See main entry under India. *Map*: 817.

Balochi, Eastern [bgp] (Baloci, Baluchi, Baluci, Eastern Hill Balochi). 1,800,000 in Pakistan (1998). 5,000,000 including



Language Map (200)





Language Name Index (50.000)

1000

Ethnologue 16 - date: February 13, 2010 - page: 1000

1000

1000	Ethnologue	Language Name Index
Kachei , <i>alt. dial. Nicobarese, Central [kəʃ]</i> , 397	Kadazan Tanggar , <i>alt. Kadazan, Coastal [kəʃ]</i> , 472	Kafa [kəʃ], 125, 699 <i>dial. Kof [kəʃ]</i> , 411
Kachepo , <i>alt. dial. Kacipo-Balesi [kəʃ]</i> , 124	Kadazan, Coastal [kəʃ], 472, 811	Kafanchan , <i>dial. Tyap [kəʃ]</i> , 187
Kachi , <i>alt. Kachi-Khi [kəʃ]</i> , 379, 551, 210, 503 <i>alt. Koli, Kachi [gk]</i> , 502, 383 <i>dial. Koli, Kachi [gk]</i> , 502, 383	Kadazan, Klias River [kəʃ], 472, 811	Kafia , <i>alt. Kafa [kəʃ]</i> , 125
Kachi Bñil , <i>dial. Koli, Kachi [gk]</i> , 502, 383	Kadazan-Kinabatangan [kəʃ], 472, 811	Kaficho , <i>alt. Kafa [kəʃ]</i> , 125
Kachi Gujaraŋi , <i>alt. Koli, Kachi [gk]</i> , 502, 383	Kadazan-Tagaru , <i>dial. Dusun, Central [kəʃ]</i> , 472	Kafia , <i>alt. dial. Bidjogo [kəʃ]</i> , 88
Kachi Koli , <i>see Koli, Kachi [gk]</i> , 502, 383	Kade-Kade , <i>alt. dial. Batak Alas Kluet [kəʃ]</i> , 450	Kafir , <i>dial. Kadaz [kəʃ]</i> , 202
Kachi Meghwar , <i>alt. dial. Koli, Kachi [gk]</i> , 502, 383	Kadenba , <i>alt. Bwamu, Láá Láá [kw]</i> , 60	Kafire , <i>dial. Senoufo, Cebaara [sf]</i> , 104
Kachib , <i>dial. Kadaz [kəʃ]</i> , 177	Kadero , <i>alt. Kadaru [kəʃ]</i> , 202	Kafira [kəʃ], 423, 794
Kachichere , <i>dial. Tyap [kəʃ]</i> , 187	Kaderu , <i>alt. Kadaru [kəʃ]</i> , 202	Kafu , <i>dial. Bulom So [bu]</i> , 194
Kachin , <i>alt. Jingpho [kəʃ]</i> , 479, 344 <i>alt. Jingpho [gk]</i> , 493	Kadi , <i>dial. Senoufo, Sayire [pp]</i> , 155	Kafu , <i>dial. Gbiri-Niragu [gr]</i> , 172
Kachina , <i>dial. Khakas [kəʃ]</i> , 520	Kadian , <i>alt. dial. Brunoi [kəʃ]</i> , 333, 471	Kag , <i>dial. ut. Ma'in [gə]</i> , 188
Kachmere , <i>alt. dial. Karanga [kəʃ]</i> , 91	Kadien , <i>alt. dial. Brunoi [kəʃ]</i> , 333, 471	Kaga , <i>dial. Kamri, Central [kəʃ]</i> , 177, 71, 91, 203
Kachuama , <i>alt. Kaxauina [kəʃ]</i> , 510	Kadin-Kaban , <i>dial. Cakfem-Mushere [kəʃ]</i> , 168	<i>dial. Mono [mh]</i> , 113
Kacipo , <i>dial. Kacipo-Balesi [kəʃ]</i> , 124	Kadina , <i>dial. Galefa [gə]</i> , 414	Kagaba , <i>alt. Kogi [kəʃ]</i> , 257
Kacipo-Balesi [kəʃ], 202, 124, 699, 728	Kadingi , <i>alt. Tur [tə]</i> , 202	Kagama , <i>alt. dial. Kamri, Central [kəʃ]</i> , 177, 72, 91, 203
Kacmiri , <i>alt. Kashmiri [ka]</i> , 380, 502	Kadiro , <i>dial. Mori [mɔ]</i> , 204	Kagan Kalagan , <i>see Kalagan, Kagan [kə]</i> , 512
Kacoʔ [kəʃ], 314, 778	Kadiwəu [kəʃ], 239, 743	Kagan Kalagan , <i>alt. Kalagan, Kagan [kə]</i> , 512
Kad Chemu , <i>alt. truda [trud]</i> , 378	Kadjakə , <i>alt. Kajakə [kəʃ]</i> , 90	Kagan , <i>alt. Hindko, Northern [hn]</i> , 501
Kada , <i>alt. Gidar [gə]</i> , 71, 89	Kadjala , <i>dial. Lama [la]</i> , 219, 55, 135	Kaganan , <i>dial. Hamaruo [hm]</i> , 510
<i>alt. Kadar [kəʃ]</i> , 379	Kadjang , <i>alt. dial. Konjo, Coastal [kəʃ]</i> , 445	Kageri , <i>alt. Kanjari [kəʃ]</i> , 380
Kadadjara , <i>alt. dial. Mataru Wangka [mɔ]</i> , 587	Kado , <i>alt. Caddo [kəʃ]</i> , 308	Kagste [yɔ], 489, 813
Kadagbe , <i>dial. Gbe, Atyo [əb]</i> , 54	<i>alt. Kado [kəʃ]</i> , 479	Kagste Bhoté , <i>alt. Kagste [yɔ]</i> , 489
Kada-Obe , <i>alt. dial. Gbe, Atyo [əb]</i> , 54	<i>alt. Umananda [um]</i> , 449	Kagayan , <i>alt. Mayan [mɔ]</i> , 511, 473
Kadagi , <i>alt. Kodaru [kəʃ]</i> , 382	<i>poj. alt. Herdē [həd]</i> , 90	Kagayanen , <i>alt. Kagayanen [gə]</i> , 512
Kadai [kəʃ], 415, 791	<i>poj. alt. Pévé [pə]</i> , 95	Kagbaaga , <i>dial. Bidjogo [bj]</i> , 140
<i>dial. Galefa [gə]</i> , 414	Kádo , <i>alt. Pévé [pə]</i> , 80	Kagbo , <i>dial. Godé [gə]</i> , 102
Kadalian , <i>alt. dial. Brunoi [kəʃ]</i> , 333	Kádo Herdē , <i>alt. Herdē [həd]</i> , 90	Kagbaba , <i>alt. Kog [kəʃ]</i> , 257
<i>dial. Brunoi [kəʃ]</i> , 471	Kádo Ngwəsh , <i>alt. Ngwə [nm]</i> , 95	Kagban , <i>alt. Hindko, Northern [hn]</i> , 501
Kadalkan , <i>dial. Finiligi [kəʃ]</i> , 509	Kádo Pévé , <i>alt. Pévé [pə]</i> , 95	Kagbu , <i>dial. Bertomno [bt]</i> , 406
Kadalkan-Barig Bontoc , <i>alt. Finaligi [kəʃ]</i> , 509	Kadohəshəsho , <i>alt. Caddo [kəʃ]</i> , 308	Kagluong , <i>alt. Kayong [kəʃ]</i> , 539
Kadam , <i>alt. Gimonite [gm]</i> , 71	Kadotho , <i>dial. Ot Danum [ot]</i> , 411	Kagoma [kəʃ], 177, 716
Kadar [kəʃ], 379	Kadu [kəʃ], 479	Kagora [kəʃ], 154, 707
Kadara [kəʃ], 177, 716	Kadu Kurumba , <i>alt. Kurumba, Betta [kəʃ]</i> , 385	<i>dial. Tyap [kəʃ]</i> , 187
Kadaru , <i>alt. Kadaru [kəʃ]</i> , 202	Kadu Sholigar , <i>alt. Sholiga [sh]</i> , 482	Kagoué , <i>alt. dial. Dida, Yocoboué [yɔ]</i> , 101
Kadaru [kəʃ], 202, 729	Kadu-Gaman , <i>alt. Kadu [kəʃ]</i> , 479	Kapu , <i>alt. dial. Gbiri-Niragu [gr]</i> , 172
Kadau , <i>alt. Puyuma [py]</i> , 527	Kadugi , <i>dial. Katcha-Kadugli-Miri [kəʃ]</i> , 203	Kagulu [kəʃ], 210, 731
<i>alt. Puyuma [py]</i> , 527	Kadukak , <i>alt. Kuruk [kəʃ]</i> , 386	Kaguru , <i>alt. Kagulu [kəʃ]</i> , 210
Kadazan , <i>alt. Dusun, Central [kəʃ]</i> , 471	Kadumodi , <i>alt. Krongo [kəʃ]</i> , 203	Kagwahimb , <i>alt. Jima [jə]</i> , 239
Kadatus , <i>dial. Muna [mh]</i> , 446	Kadun , <i>alt. dial. Vaghai-Ya-Bijim-egeri [bj]</i> , 188	Kagwahip , <i>alt. Jima [jə]</i> , 239
Kadepurhata , <i>alt. dial. Ruvuma [rw]</i> , 237	Kaduna , <i>dial. Obagyi [gb]</i> , 172	<i>alt. Tenharim [tə]</i> , 244
Kadovu , <i>dial. Fijian [fj]</i> , 593	Kaduo [kəʃ], 344, 465, 780, 807	Kagwahiva , <i>alt. Karipuna [kəʃ]</i> , 240
Kadayan , <i>alt. Dusun, Central [kəʃ]</i> , 471	Kadyan , <i>alt. dial. Brunoi [kəʃ]</i> , 333, 471	<i>alt. Ito'wa'wa'wa [itə]</i> , 245
<i>alt. Dusun, Sogot [sɔ]</i> , 471	Ká'e , <i>alt. dial. Kadaru [kəʃ]</i> , 202	Kah So , <i>alt. Sə [sə]</i> , 467
<i>alt. dial. Brunoi [kəʃ]</i> , 333, 471	Kaesab , <i>dial. Cia-Cia [cia]</i> , 444	Kahalin Iway , <i>dial. Tagon [tə]</i> , 421
	Kaeb , <i>poj. alt. Mandobo Atas [am]</i> , 434	Kahali , <i>alt. Khali [kəʃ]</i> , 381, 239
	<i>poj. alt. Mandobo Bawah [bw]</i> , 434	Kahayan , <i>dial. Ngaju [nj]</i> , 411
		Kahayan Kapuas , <i>dial. Ngaju [nj]</i> , 411

1000

EOL eppanternet - 5/24 - On/Off: 216 months version - February 13, 2010

1000

- Kachel**, *alt. dial.* Nicobarese, Central [ncb], 397
- Kachepo**, *alt. dial.* Kacipo-Balesi [koe], 124
- Kachi**, *alt.* Kachchi [kfr], 379, 151, 210, 501
alt. Koli, Kachi [gjk], 502, 383
dial. Koli, Kachi [gjk], 502, 383
- Kachi Bhil**, *dial.* Koli, Kachi [gjk], 502, 383
- Kachi Gujarati**, *alt.* Koli, Kachi [gjk], 502, 383
- Kachi Koli**, *see* Koli, Kachi [gjk], 502, 383
- Kachi Meghwar**, *alt. dial.* Koli, Kachi [gjk], 502, 383
- Kachia**, *dial.* Kadara [kad], 177
- Kachichere**, *dial.* Tyap [kcg], 187
- Kachin**, *alt.* Jingpho [kac], 479, 344
alt. Singpho [sgp], 403
- Kachin**, *dial.* [kac], 479, 344
- Kadazan Tangaa'**, *alt.* Kadazan, Coastal [kzj], 472
- Kadazan, Coastal** [kzj], 472, 811
- Kadazan, Klias River** [kqt], 472, 811
- Kadazan, Labuk-Kinabatangan** [dtb], 472, 811
- Kadazan-Tagaro**, *dial.* Dusun, Central [dtp], 471
- Kade-Kade**, *alt. dial.* Batak Alas-Kluet [btz], 450
- Kadenbà**, *alt.* Bwamu, Láá Láá [bwj], 60
- Kadero**, *alt.* Kadaru [kdu], 202
- Kaderu**, *alt.* Kadaru [kdu], 202
- Kadi**, *dial.* Senoufo, Supyire [spp], 155
- Kadian**, *alt. dial.* Brunei [kxd], 333, 471
- Kadien**, *alt. dial.* Brunei [kxd], 333, 471
- Kadim-Kaban**, *dial.* Cakfem-Mushere [kdm], 122
- Kafa** [kbr], 125, 699
dial. Foi [foi], 611
dial. Kafa [kbr], 125
- Kafanchan**, *dial.* Tyap [kcg], 187
- Kaffa**, *alt.* Kafa [kbr], 125
- Kaficho**, *alt.* Kafa [kbr], 125
- Kafila**, *alt. dial.* Bidiyo [bid], 88
- Kafir**, *dial.* Kadaru [kdu], 202
- Kafire**, *dial.* Senoufo, Cebaara [sef], 104
- Kafoa** [kpu], 423, 794
- Kafu**, *dial.* Bullom So [buy], 194
- Kafugu**, *alt. dial.* Gbiri-Niragu [grh], 172
- Kag**, *dial.* ut-Ma'in [gel], 188
- Kaga**, *dial.* Kanuri, Central [knc], 177, 72, 91, 203
dial. Mono [mnh], 113
- Kagaba**, *alt.* Kogi [kog], 257
- Kagama**, *alt. dial.* Kanuri, Central [knc], 177, 72, 91, 203



Dictionary

A a

- =a [a] int Eng question marker added to consonant final words Pdg *dispela i makim askim* Bua **Gëyam-a?** Eng Have you come? Bua **Mağ gelë Abraham-a?** Eng Did you maybe see Abraham?
- a- [a] V Eng First person plural marker prefixed to verbs stressed on the first syllable of the root in first person dual, trial and plural forms. It is often not marked on the plural exclusive unless something else intervenes between the pronoun he and the verb Bua **He su ayoh vu bë adah vun in ham rë.** Eng We are not able to hide it from you Bua **Hil ana sepa atag vaha.** Eng Let's go and follow my mother's tracks Bua **Alu aluğ dub loğ ayed sënë** Eng We two went to school in our own language
- aba [a.ˈba] noun(al) Eng cob, core, the central section of a cob of corn or of pandanus fruit after the outer seeds or kernels are removed Pdg *bun bilong kon o marita* Bua **Aba neggëp loğ sakom ayo peggo** Eng There is an inner core inside a cob of corn Bua **Desoh tagee gedetë aba ya** Eng They removed the fruit of the pandanus and threw the inner core away.
- abel [a.ˈmbe] noun(al) Eng line, row Pdg *lain* noun(al) Eng pole, the lowest of the three horizontal poles to which a fence is tied and which form the main horizontal framework for the fence. This is the biggest of the three Pdg *stik-longpela diwai o mambu ol i yusim bilong pasim banis i go pas long en*
- abel **abel** [a.ˈmbe] a.ˈmbe] Nc Eng lines, rows Pdg *lain*
- abes [a.ˈmbe] noun(al) Eng having a flat tail-as of eels, tadpoles, catfish Pdg *stretpela tel-olsen tel bilong maleo, pis, samting* Bua **Beg-göb rak bël rabu rak abes** Eng Eels are able to climb waterfalls by using their tails Bua
- Mëğov abes pepöö ya gevaha los nema rak** Eng Tadpoles' tails fall off and they grow legs and hands
- abiba** [a.ˈbi.ˈba] noun(al) Eng yam species Pdg *yam*
- abu** [a.ˈbu] noun(al) Eng large brown flying beetle Pdg *kain binatang em i save plal long nait*
- abuhëk** [a.ˈbu.ək] noun(al) Eng edible wild sugarcane-Saccharum edule; this is the generic term, many different varieties are grown Pdg *ol pitpit bilong kaikai*
- abuung** [a.ˈbuŋ] noun(al) Eng A flat sheet used as a rubbish carrier. They are usually used in pairs and are made from any suitable material such as the sago palm trunk, the sheath of a banana palm, or in more recent times from a couple of pieces of flat iron or plywood. Pdg *pangal bilong saksak ol meri save kisim bilong bungim pipia long en.*
- adiiv** [a.ˈdi:β] prop Eng Angans-people of the Angan family of languages, formerly referred to as the Kukukuku people Pdg *man-meri Watut inap i go long Asiki*
- ading** [a.ˈdiŋ] 1) adj Eng long, tall Pdg *longpela, longwe* Bua **Maluh seyu anöö rak aggis ading.** Eng The man tied the dog up with a long rope. Bua **Nabë sa nanër dus, më ading, oy iğ yoh vu!** Eng Whether I speak briefly or at length, it doesn't matter. Bua **Bël tuus geseya ading rot.** Eng The water had dried up and I went a very long way. 2) Adv Eng far, distant Pdg *longwe* Bua **He naya nyëğ ading rot.** Eng We are going to a very distant place. Bua **Böök yu ti vare deneruuk rak kedu sënë medenare ading teka.** Eng A herd of pigs were rooting on this mountain and they were standing at a little distance. 3) adj Eng tall
- adingnë** [a.ˈdiŋnë] Adv Eng far away, at a distance Pdg *longwe tru* Bua **Gwetah hong na adingnë** Eng Move well back/far away Bua



Fieldworks

- FW Data Notebook
- FW WorldPad
- FW Language Explorer
 - lexicon
 - interlinear tool
 - grammar tool
 - data integration



www.sil.org/computing/fieldworks/

The screenshot displays the FieldWorks Language Explorer interface. The main window is titled "Simple A - FieldWorks Language Explorer" and features a menu bar (File, Edit, View, Data, Insert, Format, Tools, Parser, Window, Help) and a toolbar. The interface is divided into several panes:

- Lexicon:** A sidebar on the left with a tree view containing items like "Lexicon Edit", "Browse", "Dictionary", "Categorized Entry", "Classified Dictionary", "Bulk Edit Entries", "Bulk Edit Senses", "Reversal Indexes", and "Bulk Edit Reversal Entries". Below this are buttons for "Lexicon", "Texts", "Words", "Grammar", and "Lists".
- Entries:** A central table with columns: "Headword", "Lexeme Form", "Grammatical Inf...", "Glosses", and "Semantic Dom...". The table lists various words such as "aardvark", "bank", "bark", "beaver", "box", "break", "bright", "bring", "cat", "dog", "elephant", "fox", "giraffe", "horse", "ring", "-s", and "wood".
- Entry:** A detailed view on the right for the selected word "horse". It shows:
 - Headword: **horse** / A large hoofed animal (scientific name *Equus caballus*).
 - Lexeme Form: Eng **horse**, Morph Type: stem.
 - Citation Form: Eng.
 - Entry Type: Main Entry.
 - Note: Eng.
 - Summary Definition: Eng.
 - Sense 1: Gloss: Eng **hoofed animal**; Definition: Eng A large hoofed animal (scientific name *Equus caballus*).
 - Grammatical Info.: Noun.
 - Example: (empty).
 - Semantic Domains: (empty).
 - Lexical Relations: (empty).
 - Alternate Forms: (empty).
 - Grammatical Info. Details: Category Info. Noun.

At the bottom of the window, the status bar shows the date "27/Sep/2007", a queue status "Queue: [:-/-] No Parser Loaded", and a page indicator "15/18".



Dictionary Express

■ File > Print as Dictionary ?

■ Dictionary Express



Dictionary Express

- ConT_EXt-based tool for DE: DEtool
- data from database,
- in proper format,
- for typesetting system to produce pdf
- (behind the scenes)



Dictionary output

■ DEoutput



Dictionaries: typesetting

- data from database: converters to ODF and $\text{T}_{\text{E}}\text{X}$ -format
- in proper format: using TeX-tagged Dictionary Data (T2D2)
- for typesetting system to produce pdf: Con $\text{T}_{\text{E}}\text{X}$ t-Based Library for Typesetting (CoBaLT)
- (hidden for user): embedded minimal Con $\text{T}_{\text{E}}\text{X}$ t-distribution (miniCTX) (nota bene: work in progress)



Dictionaries: sample tags

- headword (hw): this is the word that this particular entry is about,
- pronunciation (pr): the proper pronunciation of the word written using the International Phonetic Alphabet (IPA),
- part of speech (ps): the grammatical function of the word,
- language tag (lt): the language of the definition or example,
- definition (de): meaning of the headword,
- example (ex): example of the word used in a sentence.



Dictionaries: sample entry

```
\Bentry  
\Bhw{abel}\Ehw  
\marking[guidewords]{abel}  
\Bpr{a.bl}\Epr  
\Bps{noun(al)}\Eps  
\Blt{Eng}\Elt  
\Bde{line, row}\Ede  
\Blt{Pdg}\Elt  
\Bde{lain}\Ede  
\Eentry
```



adiiv [a.¹ⁿdi:β] prop Eng Angans-people of
the Angan family of languages, formerly re-
ferred to as the Kukukuku people PdG man-
meri Watut inap i go long Asiki



Dictionaries: final

- most of the required features are implemented
 - font selection (including the use of Graphite fonts)
 - basic dictionary layout and picture support
- some features easy to implement, others not (page wide pictures that keep floating to the next page)
- remaining challenge: ultra light version of ConT_EXt
 - dealing with the Ruby dependency
 - stripping the T_EX-tree
- work in progress



Comments or Questions?

www.ethnologue.com

www.sil.org/computing/fieldworks

jelle_huisman@sil.org

