

# Language Documentation and Description

ISSN 1740-6234

---

This article appears in: *Language Documentation and Description, vol 1*. Editor: Peter K. Austin

## Documenting Lexicons: Chechen and Ingush

JOHANNA NICHOLS, RONALD L. SPROUSE

Cite this article: Johanna Nichols, Ronald L. Sprouse (2003). Documenting Lexicons: Chechen and Ingush. In Peter K. Austin (ed.) *Language Documentation and Description, vol 1*. London: SOAS. pp. 99-121

Link to this article: <http://www.elpublishing.org/PID/010>

This electronic version first published: July 2014

---



This article is published under a Creative Commons License CC-BY-NC (Attribution-NonCommercial). The licence permits users to use, reproduce, disseminate or display the article provided that the author is attributed as the original creator and that the reuse is restricted to non-commercial purposes i.e. research or educational use. See <http://creativecommons.org/licenses/by-nc/4.0/>

---

## EL Publishing

For more EL Publishing articles and services:

Website:	<a href="http://www.elpublishing.org">http://www.elpublishing.org</a>
Terms of use:	<a href="http://www.elpublishing.org/terms">http://www.elpublishing.org/terms</a>
Submissions:	<a href="http://www.elpublishing.org/submissions">http://www.elpublishing.org/submissions</a>

# Documenting Lexicons: Chechen and Ingush

Johanna Nichols and Ronald L. Sprouse<sup>1</sup>

## 1. Introduction.

Producing easily usable, professional-looking descriptive dictionaries on a shoestring budget in a short time span is a priority for documentation, but hard to achieve. The usual procedure for field dictionaries is to compile a target language-to-contact language lexical database (e.g. Chechen-English, in our case) and generate a contact-to-target dictionary or index from the glosses. This is economical but not always fully satisfactory. Here we describe our solutions to some common problems of field lexicography based on several years' experience at compiling, editing, and publishing dictionaries of Chechen and Ingush, close sister languages of the Nakh-Daghestanian language family spoken in the central Caucasus.<sup>2</sup> They are languages with large, literate speech communities for which dictionaries need to be sizable, attractive, and linguistically sophisticated and for which two different alphabets are needed, and we hope that our experience in trying to meet these goals will be helpful to linguists embarking on lexical documentation.

This is not a typical field dictionary project, nor are the Chechen and Ingush languages imminently endangered in the usual sense: there are close to a million speakers of Chechen and 300,000 of Ingush, and there are orthographies, literatures, schools, academics, media, publishing ... and dictionaries. On the other hand the Soviet past lies heavy on the speech communities, intellectual traditions, and facilities for production of such things as dictionaries. Despite excellent lexicography by a number of Ingush and Chechen scholars, state control of publication kept them more symbolic than substantive. There has never been a defining dictionary published in either Chechen or Ingush. An excellent Chechen-Russian dictionary, Maciev 1961, was published in only 9000 copies, making it unavailable to all but a few Chechens; most schools and municipal libraries in Chechnya did not have a copy, let alone individuals. The Ingush glossary of Mal'sagov 1963 (originally published in the then-mandated Latin orthography in 1925, reprinted in Cyrillic in 1963), was excellent but more nearly a lemma list than a regular dictionary, and was published in very few copies. The 1963 reprint was again in a small press run. No Ingush-Russian dictionary was ever published, though Ozdoev et al. 1962 is an Ingush-Chechen-Russian dictionary, small but good, in which for each Ingush word is given one Chechen word and one Russian word as glosses; it has information about inflection but none about

---

<sup>1</sup> The authors' names are in alphabetical order.

<sup>2</sup> The Chechen and Ingush people have nearly identical customs and, because of widespread passive bilingualism, the two languages form a single speech community (though they are not in fact mutually intelligible). In typological terms they are morphologically ergative, verb-final with V2 main clauses, dependent-marking, clause-chaining, with complex verb morphology; they have moderately complex consonant inventories with ejectives, uvulars, pharyngeals, etc. and also complex vowel systems.

valence. Only 1000 copies were published. Russian-Chechen and Russian-Ingush dictionaries were published in 1978 and 1980, again in small runs (Karasaev and Maciev 1978: 18,000 copies; Ozdoeva and Kurkueva 1980: 5000 copies), and with some grammatical information about the Russian headwords but none about the Chechen or Ingush translations. The press-run sizes and one-time-only publications make it obvious that it was never intended that every classroom should have a dictionary and every student, or at least every household, should own one. All of the Soviet publications necessarily contained a good deal of Soviet language that even then was hardly in general use.

Consistent with the goals of Soviet language planning, Chechen and Ingush were treated not as vehicles for transmission and acquisition of information but as ethnographic exotica, and in Soviet times no attempt was ever made to publish bilingual dictionaries of Chechen/Ingush and English or any other major language. The only access to world languages was via Russian. Knowledge of contemporary lexicography was hard to come by and the facilities for computational lexicography unavailable. Our lexicography project began in 1996, as part of a field methods course on Ingush taught at UC Berkeley, and we undertook it partly because we needed to organize our lexical materials and largely because our Ingush consultant, Issa Guliev, considered it a first priority and urged us to do it.<sup>3</sup>

## 2. Early history of the project

Several students in the class were doing lexically oriented research projects and had begun collecting lexical data in spreadsheet, word-processor, or database files and contributed them to the lexical database: Zev J. Handel systematized verb conjugations and made decisions about citation forms and principal parts for verbs (see Handel in press); William F. Weigel worked out the case endings of nouns and the status of post-case clitics; Armindo S. A. Ngunga elicited a large number of nouns with their principal parts and genders; Jan Johnson elicited color terms; Heather Jones elicited motion verbs. Nichols did a complete pass through the glossary of Mal'sagov 1978, extracting all elementary verbs and working out with Guliev their principal parts and meanings. Sprouse began eliciting adjectives. Together with Andrew Dolbey, he worked out the covert possessive classification of Ingush nouns and the different types of phrasal verbs in Ingush.

---

<sup>3</sup> After our project was underway, two Chechen sources appeared. Awde, Nicholas, and Galaev, Muhammad. 1997. *Chechen-English, English-Chechen Dictionary and Phrasebook*. New York: Hippocrene. is a small bilingual dictionary with minimal glosses; the Chechen is in transcription only, in a system using a mix of phonetic symbols, Latin letters and digraphs, and diacritics, underdifferentiating most vowel phonemes and apparently created partly by transliterating items from Maciev 1961 and partly by phonetic transcription of elicited material. It would function well as a tourist's or journalist's phrasebook but was not intended to serve Chechen speakers. Vagapov, Arbi. 1999. *Ingals-noxchiin deshnizhaina / English-Chechen Dictionary*. Soelzha-Ghaala: Chechen State University. is the first source on English designed for use by Chechens. It contains 3000 English words and phrases, each with one or more Chechen translations, with part of speech for the English words and gender for the Chechen ones. Unfortunately, only 1000 copies were printed.

Meanwhile, since class participants were using email from various platforms to exchange information about Ingush, and since we all wanted to be able to type Ingush examples at a normal rate, we worked out an all-Latin, diacritic-free (i.e. all-lower-ASCII) practical phonemic transcription for Ingush. In April 1996 Nichols and Irwin Komen of SIL, who was working with Chechen, compared and aligned their independently devised but very similar Latin spelling systems for Ingush and Chechen. (The current systems for the Berkeley project can be seen on the world wide web at <http://ingush.berkeley.edu:7061/> and <http://socrates.berkeley.edu/~chechen>.)

Initially as part of a course on computers in linguistics taught by John B. Lowe, Sprouse and Handel designed a FileMaker Pro™ database for the dictionary project. It uses a separate layout for each part of speech, and a radio button drives a script that adjusts the layout to the part of speech of the headword. For the inflecting parts of speech there is a field for every form in the paradigm, and a citation form field separate from these (though in fact the citation form is one of the inflectional forms). There are fields for lexical categories such as gender of nouns, valence and Aktionsart of verbs, etc. The gloss field is an unstructured large text field, and the lexicographer is free to use commas, semicolons, or other devices to separate sub-senses and by-glosses.

Sprouse and Handel also designed a separate database of example sentences, in which the example appeared both as a unit phrase and broken down into separate words, each occupying a field and with an interlinear gloss field corresponding to each word field. This was followed by half a dozen fields for entering citation forms of key words in the example, and another half dozen fields for entering citation forms of words for which the phrase was a good example. This database was queried from the main one, and from any lexical record in the main database one could call up whatever example sentences had among their keywords a form identical to the lexical record's citation form. This was before FileMaker Pro™ was fully relational, so the search could be done only on actual forms. As a result, the example sentences button called up all examples of that word and of any homonyms.

Over the next six months we checked and elicited more words, added Cyrillic spellings of the citation forms and began improving coverage of inflectional forms. In December 1996, about nine months after work on the database began, we produced a sample and prospectus (Nichols et al. 1996) by printing out reports from the database and alphabetizing and editing them in Word. This dictionary prospectus was in three sections: Ingush-English, alphabetized by the Ingush citation form in standard Cyrillic orthography (53 pp.), Ingush-English in Latin transcription, and English-Ingush. The Cyrillic section was for use by Ingush speakers literate in the Cyrillic orthography, and the Latin one was for use by English-speaking learners of Ingush, members of the Ingush diaspora who speak the language but do not know the Cyrillic spelling, and researchers. Cyrillic and Latin sections contained the same information, alphabetized differently. Zev Handel wrote a number of scripts and programs which edited, organized, and sorted the FileMaker Pro™

report, doing such things as boldfacing the citation form, italicizing the principal parts, enclosing the gloss in single quotes, printing the part of speech as an abbreviation in parentheses, etc. He also produced the English-Ingush section from the gloss field by separating that into pieces at commas or semicolons and associating each such piece with the citation form. He also removed ‘little words’ from the alphabetization so that, e.g., *be healthy* is alphabetized and printed as *healthy: be healthy*. The result was less economical of space than a printed dictionary, and at that early stage had various gaps and some typographical errors, but it would have been quite serviceable as a shoestring classroom text except that it contained only about 1000 words, not enough to meet even elementary teaching needs.

In connection with word counts, we note that about 30% of the Ingush verbal roots, including the auxiliaries that are common in phrasal verbs (and therefore more than 30% of the verbal lexemes overall), take gender agreement in the form of mutation of the initial consonant:

<i>vuoda</i>	‘(he) goes’
<i>juoda</i>	‘(she/it) goes’
<i>duoda</i>	‘(child/it) goes’
<i>buoda</i>	‘(it) goes’

We call the genders W, J, D, and B correspondingly. D gender is the standard citation form, but Maciev 1961, Mal’sagov (1925) 1963, and all other Chechen and Ingush publications that we know of include all four gender forms of each gender-inflecting verb as citation forms. This must approximately double the word count, so that e.g. the 20,000 words of Maciev 1961 are really closer to 10,000 lexical items. We have not done this, except in a few cases where different gender forms are lexicalized separately or have very different privileges of occurrence.

### 3. Continuing work

Issa Guliev spent 1995-96 in Berkeley as consultant for the field methods course and returned to Berkeley for a semester each of the next three years for work on the dictionary. During this time he and project participants Nichols, Sprouse, David A. Peterson, and Lisa J. Conathan elicited much lexical and text material. Nichols and Peterson checked valence of verbs and elicited examples showing their case government; Sprouse elicited adjectives and other material; Conathan and Nichols worked through a 5000-word English learner’s dictionary and other basic wordlists. Individually and jointly, all of us worked on texts. Nichols and Guliev did complete passes through the glossary of Mal’sagov 1963 and Ozdoev et al. 1962 and extracted all words known to Guliev. We used these published materials essentially as wordlists, providing glosses and grammatical information from scratch. In the spring of 2001, Ingush writer Vakha Khamkhoev spent some time in Berkeley and reviewed remaining words from Ozdoev et al. 1962 and other materials with

Nichols. Nurdin Kodzoev forwarded his compilation of bird and animal names from the works of Ali Xashagul'gov, and Sultan Mereshkov forwarded his materials on Ingush traditional terminology in various semantic fields, and these materials were added to the database. Nichols wrote an introductory grammatical sketch and appendices illustrating conjugations and declensions. In the spring of 2002, Maryam Yandieva of the Ingush Permanent Representation in Moscow forwarded copies of a late draft printout to several academics and lexicographers in Ingushetia and forwarded their comments back to Nichols, who took them into account in the final revision, which was sent to the publisher in June 2002.

#### **4. Chechen-English dictionary**

Since the Ingush and Chechen languages are closely related they have nearly identical grammatical structures and lexical organization, and the software and analytic decisions produced in the Ingush project could easily be applied to Chechen. Additional funding was secured to begin a parallel Chechen project, and Professor Arbi Vagapov of Chechen State University spent academic year 1999-2000 in Berkeley working on the Chechen-English dictionary of which he is coauthor and serving as consultant for a field methods course on Chechen co-taught by Nichols and Professors Sharon Inkelas and Andreas Kathol. Vagapov drew up a wordlist using his own published and unpublished works (including Vagapov 1999) and Maciev 1961, and over the course of the year entered nearly 5000 lexemes with their grammatical information, which he and Nichols then glossed together. Nichols did some supplementary elicitation with other Chechen speakers after Vagapov left, and wrote an introductory grammatical sketch and appendices illustrating conjugations and declensions. Thus, in about one year's time the project produced a Chechen lexical database comparable in size and nearly identical in format to the Ingush one. The printout was checked and proofread by three other Chechen speakers and the final version sent to the publisher in fall 2002.

#### **5. Excursus on the example sentences database and BITC.**

Not long after the lexical database was created, we began work on a small corpus of texts that could be used to illustrate the lexical items in context. The first incarnation of this corpus was the example sentences database described above, implemented in FileMaker Pro. It quickly became apparent that a relational database was poorly suited to the task of collecting and interlinearizing natural language data, so Sprouse began work on the Berkeley Interlinear Text Collector (BITC).

BITC is designed to be a flexible way of collecting and interlinearizing texts. It runs on a web server, which makes it useful for groups of collaborators to share work. Users need only Internet access and a web browser to participate in the project. This sharing is enhanced by semi-automatic glossing from a shared word list that grows as more texts are entered into the database. Every time a user gives an interlinear gloss for a given

word, that gloss is stored in a central word list. Thereafter, that gloss is provided as a suggested gloss the next time the same word form is encountered in a text, regardless of the user entering the data. That user may choose one of the suggested interlinear glosses or provide a new gloss of their own. Access to the suggested glosses has the obvious benefit of increasing the speed at which word forms can be glossed since many glosses can be selected from a list rather than typed in afresh each time the word form occurs. Selecting a form from a word list also helps encourage (but doesn't enforce) consistency in data entry. A sample screenshot of the BITC editor window is shown in Figure 1:

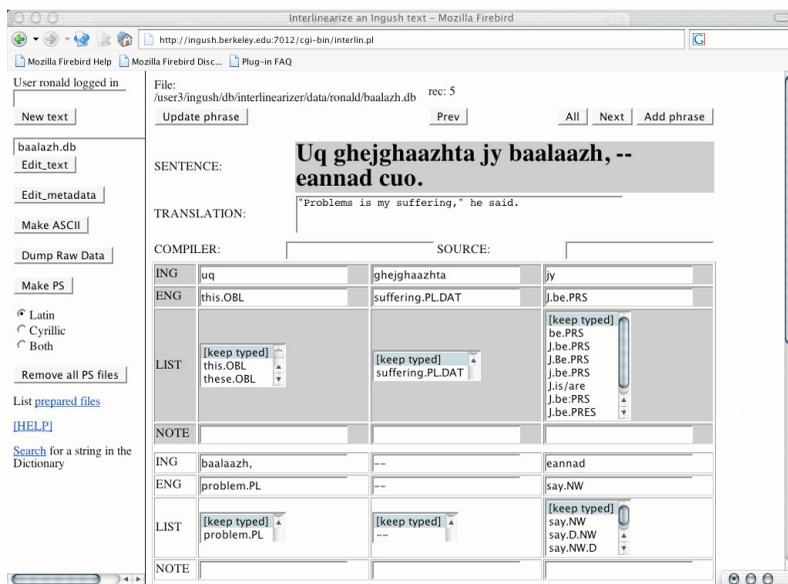


Figure 1. BITC editor window

There are two ways to search a BITC corpus. The first is to search the word list for occurrences of a string. BITC reports each matching word form along with a list of glosses for that form. This list of glosses is the same as the suggested glosses provided when entering data. A second search method yields even better results. This method skips the word list and searches the corpus directly, reporting back a hyperlinked list of word forms so that the user may browse directly to the phrases in which the word is found. This kind of searching essentially allows for on-the-fly concordances to be created so that the contexts in which a lexical item appears may be explored. This feature is also useful for finding and

eliminating common typographical errors. Simply search for the error, browse to the records in which it appears, and correct it.

The BITC search features also allow the user to abstract away from certain problems of Chechen/Ingush morphology. A single meta-character may be used to stand in for one of the four gender markers, which allows a user to find all occurrences of a gender-marked word form regardless of how it is actually marked. An ‘ablaut mercy’ option takes pity on the poor user who can’t remember ablaut patterns and seeks to find word forms that have complementary ablaut vowels with the specified search string.

BITC data can be exported in various ways so that examples from the corpus can be incorporated into publications. The first option produces Postscript copy (with optional LaTeX files). The second option exports tab-delimited text files and is popular with users who do their writing in a word processor. These examples often require some editing to make them fit the requirements of the document into which they are inserted. The last option is a raw text dump that includes all the structural markup used by BITC and is mainly useful for creating human-readable backups.

Overall BITC has been a very successful tool for creating the Chechen and Ingush corpora and for exploring all kinds of linguistic structures from phonology to syntax. At present, however, the BITC corpus is not integrated with the FileMaker Pro lexical database, so it is not yet a complete lexical resource. It has, however, been invaluable for gathering the grammatical information, such as inflectional classes and valence, needed for lexical entries and for revealing the semantic ranges of words.

## 6. Structure of the current lexical databases

The Ingush lexical databases currently contains around 5800 words and the Chechen one about 5600. Both are constantly being edited and expanded, and we plan to continue maintaining and expanding them indefinitely as well as making some changes in their structure. The current structure is largely inherited, with some expansions, from the original structure worked out by Handel and Sprouse. Here we discuss some of those that may have wider relevance to lexical documentation.

Figure 2 shows a noun record from the Chechen lexical database: the word *ch’eepalg* referring to a traditional Chechen pastry. The printout gives the appearance of two pages: a data page with the word and information, and an administrative notes/citations page, actually a separate layout brought up by clicking on the “Admin Notes/Citations” button on the first page. The top of the first page gives the citation form in Latin and Cyrillic, the gloss, the part of speech, and the permanent serial number automatically assigned to this record when it was created. The gloss field visible in the figure is a small window on an unstructured text field. Under the gloss is a “short gloss” field, unfilled on this record and used especially for verbs to extract a single most prototypical gloss for use in citations, for instance in comparative-historical work. Under

that are lemma fields discussed below. In the middle of the page is grammatical information: gender, case forms, declension type, etc. The “Notes” field at the bottom is for information that is to be a permanent part of the record, and often contains encyclopedic information (as here). “Other citations” is for variant spellings and pronunciations. The second page contains various less permanent notes, a record of work on the entry, and sources. The citation form and gloss are repeated at the top of the page, followed by a “Dictionary gloss” field designed to handle some of the problems inherent in using the gloss field to generate a wordlist for (in this case) the English-Chechen dictionary. This is a word in which the gloss is actually a brief encyclopedic entry rather than a true gloss, explaining an element of Chechen culture to a non-Chechen user. There is no English word that should trigger this word as its translation, and the notation “[none]” in the Dictionary Gloss withholds this record from the English-Chechen dictionary compilation.

ID Num 5274 [Chechen Lexicon: Noun](#) [Prev](#) [Next](#)

Part of Speech Noun class traditional Chechen dish (pasta shells filled with cottage cheese, cheese, or

Quotation Form **ch'eepalg** short\_gloss **ɣlɛtɑm**

[Ge](#) [Ge](#) [Ge](#) lemma

Gender angular  plural  [Example Sentences](#) [Admin Notes/Citations](#)

opl\_posession\_type  POSITIVE  ALLATIVE  POSITIVE/ALLATIVE  UNKNOWN

noun\_stem  irregular  Yes  No (Suppletive)

noun\_root  Edit by hand

declension type  Plan (number)

Case	angular	plural
NOM	ch'eepalg	ɣlɛtɑm
GEN	ch'eepalgan	ɣlɛtɑmɑn
DAI	ch'eepalgana	ɣlɛtɑmɑna
ERG	ch'eepalguo	ɣlɛtɑmu
ALL	ch'eepalgie	ɣlɛtɑme
INS	ch'eepalgca	ɣlɛtɑmca
LAT	ch'eepalgax	ɣlɛtɑmax
CMN	ch'eepalgal	ɣlɛtɑmal
ADV		

Adv gloss

[See Other Noun Forms](#)

notes Traditionally and prototypically, the filling is cottage cheese.

other\_quotations

declension type

decl ID

all decl type

declension desc

Yes

cv stem

Yes

Fig. 2a. Main page of a noun record.

Web?  Yes      ID Num 5274 [Chechen Lexicon: Notes/Citations](#)

Part of Speech	Noun	gloss	traditional Chechen dish (pasta, shells filled with cottage cheese, cheese, or
Chuvon Form	ch'eeppalg	Dictionary Gloss	[none]

---

Secondary Fields

food		
------	--	--

notes\_temporary

notes\_methodological

etymology

compiler

AY 4-15-00, JN

source

<input checked="" type="checkbox"/> Waclov <input type="checkbox"/> Allnovev. Cit. <input type="checkbox"/> Vagapov <input type="checkbox"/> KPR <input type="checkbox"/> K & M (FC) <input type="checkbox"/> Innalshvili <input type="checkbox"/> Other <input type="checkbox"/> Undocumented	Other Source Info	<div style="border: 1px solid black; width: 200px; height: 30px;"></div>
---	-------------------	--

[Return to data page](#)

Qualification

phonological info  
for each item, please (add) pronunciation –  
use Cof-rob (FC) or Opho-rob (Misc)  
to enter a rob character

Fig. 2b. Administrative page of a noun record.

We also use this field to remove parts of glosses. For instance, Ingush *laattar* is glossed ‘stand, be standing; consist; progressive auxiliary’, where the first three are proper glosses and *progressive auxiliary* is a description. The Dictionary Gloss field for this word consists of only ‘stand; be standing; consist’. Use of this field to alphabetize glosses properly is described in section 7.1 below.

The Dictionary Gloss field thus enables us to work with almost no constraints on the wording and format of glosses. The only constraint is that, since stretches separated by commas or semicolons are used to create entries in the English dictionary, gloss-internal commas cannot be used. Thus, though it is tempting to use telegraphic wording such as ‘cultivate earth, garden’, we use the alternative ‘cultivate (earth or garden)’ (parenthesized elements are all removed from English headwords, so this yields simply *cultivate*). When coordinating parts of glosses seems necessary, as in ‘fall into a crevice, bluff, or abyss’, we use it but factor out the glosses in the Dictionary Gloss field: ‘fall into a crevice; fall off a cliff’. This small additional time expenditure makes it possible to use the glosses to generate a usable English-target language dictionary that is very close to a professional bilingual dictionary.

Under the Dictionary Gloss field is a repeating field showing up to three entries for Semantic Fields. Entries are selected from a pull-down menu of broad semantic fields: agriculture, anatomy, animal name, architecture, bird name, cards, chemical, clothing ... The menu is built up as work progresses and expanded as needed. By searching on these fields one can assemble a complete set of technical vocabulary in an area. We have used these fields for research on word formation, ethnosemantics, and distribution of genders.

As the dictionary became larger we realized we were putting idioms and other phraseological units into the Notes field as illustrations of how headwords are lexicalized. Since these needed to be brought together and printed out, we created a separate Phraseological Units field and began entering these items there, in both Cyrillic and Latin and with glosses. In production of a dictionary printout we print these as subwords under the headword. This tactic is workable for a small dictionary, but not ideal as most lexicalized phrases are equally related to all of the words that appear in them. For instance, *dogha delx* ‘rain falls, it rains’ is associated with both *dogha* ‘rain’ and *d.elxar* ‘cry, weep’, or *xin kertie* ‘headwaters, source (of river)’ with both *xii* ‘water; river’ and *kuorta* ‘head’. In the future we will turn the phraseological units into separate records, relationally linked to each relevant headword and displayed in a portal on that headword’s records. This will have the added advantage of providing for grammatical information, notes, sources, etc. on the phraseological unit as a whole.

In Ingush and Chechen, simple verbs are a closed class and the main source of new verbs is compounds and phrases with a conjugated verb as second element and often a noun or fossilized noun as first element: *nab jar* ‘sleep, take a nap’, lit. ‘do sleeping’; *kedzh jaqqar* ‘come to a boil’, lit. ‘take a boil’, *cerg tuoxar* ‘bite’, lit. ‘tooth strike’, *xaattar dar*

‘ask’, lit. ‘make a question’, etc. In all of these the first element is a former direct object now lexicalized as part of the compound verb but exhibiting properties of direct objects such as triggering gender agreement on the verb and blocking the presence of any other direct object. In valence notations we cite the first element itself with a pound sign before it: *nab jar* ‘sleep, take a nap’ (Erg #nab); *cerg tuoxar* ‘bite’ (Erg Dat #cerg), etc. Strictly speaking this is not necessary, since the entire phrase or compound is cited as the verb, but we find it clearer and assume that non-linguists will too. In addition, this notation distinguishes these verbs from others whose first element does not control agreement, e.g. *ch’orma-d.aaqqar* ‘peel, remove peeling’ (Erg Abs), where the second element agrees not with the first element but with the external, free direct object.

Figure 3 shows the first page for a simple verb and Figure 4 the record for a compound verb, both from Ingush. For the simple verb, all inflectional forms are filled in. For the compound verb, only those that serve as principal parts are filled in. For all non-simple verbs, the radical (for verbs that are non-simple because of directional/locational prefixation) or auxiliary (for other compounds and phrases) is identified near the top of the page, and one can consult that record by clicking the “Go” button next to that field. The button activates a script which searches for the auxiliary or radical in the citation form field of all other records, and as a result it returns not only the desired radical or auxiliary but all other non-simple verbs built on it. This link will be relationalized in the near future, though we note that in its present form it is quite serviceable, and one can sort through the returned records quickly to find the simple root. In our earliest years of lexical analysis this even had its beneficial side, as regularities of derivation quickly became evident from looking over a set of related derivatives.

ID Num	9	Verb?	<input checked="" type="checkbox"/> Yes	Ingush Lexicon: Verb	Prev	Next
Part. of Speech	Verb	Prog. form	laatt	short_Labels/AUX	stand	
Verb (Present)	laatt	<input type="checkbox"/> stand, be standing; consist; <input type="checkbox"/> progressive auxiliary		cross_ref	<input type="checkbox"/>	
lemma	laatt	Radical	laatt	Auxiliary	Simple? Yes	
reference	Nom	<a href="#">Example Sentences</a> <a href="#">Admin Notes/Citations</a>				
aspect: discontinuous <input type="checkbox"/> Moment <input checked="" type="checkbox"/> Duration <input type="checkbox"/> Repetition <input type="checkbox"/> Other <input type="checkbox"/> Unknown aspect: am iter <input type="checkbox"/> Simultaneous <input type="checkbox"/> Iterative <input type="checkbox"/> Neither <input checked="" type="checkbox"/> Unknown <input type="checkbox"/> simultaneous <input type="checkbox"/> iterative						
number: sing pl <input type="checkbox"/> Singular <input type="checkbox"/> Plural <input type="checkbox"/> Neutral <input type="checkbox"/> Unknown <input type="checkbox"/> singular <input type="checkbox"/> plural						
CS	laatta-d.u	gloss	keep, maintain; make stand;			
C3rd		gloss				
C3rdI		gloss				
INCP3 I		gloss				
INCH		gloss				
type: laatt irregular <input type="checkbox"/> Yes <input checked="" type="checkbox"/> No						
INFINFY	laatta	laatt	IMPV(2)			
PRS	laatt	laatt	VN	laattar	laattap	
PR3:SGC			VP	laattar	laattap	
IMPV	laattar	laattap	MM	laattaav	laattax	
PUR	laattagva	laattax	PPV I	laattaavar	laattaxp	
CYam	laattazh	laattax	CYent	laattaa	laattax	
			on temp	laattar	laattap	
notes	As progressive auxiliary, usually a witnessed or ongoing true progressive. In meaning 'stand', is inherently progressive.					<input type="checkbox"/>
other_relations						<input type="checkbox"/>

Figure 3. Record for a simple verb.

ID Num	150	Web?	<input type="checkbox"/> Yes	Ingush Lexicon: Verb	Prev	Next																												
Part of Speech	Verb	Pragm	short_gloss																															
Verb (Present)	bwaguor d.oal	get dizzy, be dizzy; be a bit crazy, be peculiar			cross_ref	<input type="checkbox"/>																												
IPA	→ðj̥wəːʔ f̥j̥iːʔ ʔʷʷʌʔʷ	gloss		Simple?	No																													
lemma		Go	Radical	Go	Auxiliary	d.oal																												
valence	Nom	Example Sentences		Admin Notes/Citations																														
<input type="checkbox"/> aspect: dissociation <input type="checkbox"/> Moment <input type="checkbox"/> Duration <input type="checkbox"/> Repetition <input type="checkbox"/> Other <input type="checkbox"/> Unknown <input type="checkbox"/> aspect: sem iter <input type="checkbox"/> Simulactive <input type="checkbox"/> Iterative <input type="checkbox"/> Neither <input type="checkbox"/> Unknown <input type="button" value="Go"/> simulative <input type="button" value="Go"/> iterative																																		
number: sing pl <input type="checkbox"/> Singular <input type="checkbox"/> Plural <input type="checkbox"/> Neutral <input type="checkbox"/> Unknown <input type="button" value="Go"/> singular <input type="button" value="Go"/> plural																																		
CS		gloss																																
CS:nd		gloss																																
CS:bl		gloss																																
INCP:ɾ		gloss																																
INCH		gloss																																
		gloss																																
<table border="1"> <thead> <tr> <th>type</th> <th>irregular</th> <th><input type="checkbox"/> Yes</th> <th><input checked="" type="checkbox"/> No</th> </tr> </thead> <tbody> <tr> <td>IMP:IMP</td> <td>bwaguor d.sala</td> <td>бӀагӀо р х.а.а.а.</td> <td>IMP(IMP)</td> </tr> <tr> <td>PRE</td> <td>bwaguor d.oal</td> <td>бӀагӀо р х.о.а.а.</td> <td>YN</td> </tr> <tr> <td>PRE:NEG</td> <td></td> <td></td> <td>YN</td> </tr> <tr> <td>IMP</td> <td></td> <td></td> <td>YN</td> </tr> <tr> <td>PL</td> <td></td> <td></td> <td>PP:ɾ</td> </tr> <tr> <td>CS:m</td> <td></td> <td>CS:m</td> <td>bwaguor d.eanna бӀагӀо р х.а.а.а.а.</td> </tr> </tbody> </table>							type	irregular	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No	IMP:IMP	bwaguor d.sala	бӀагӀо р х.а.а.а.	IMP(IMP)	PRE	bwaguor d.oal	бӀагӀо р х.о.а.а.	YN	PRE:NEG			YN	IMP			YN	PL			PP:ɾ	CS:m		CS:m	bwaguor d.eanna бӀагӀо р х.а.а.а.а.
type	irregular	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No																															
IMP:IMP	bwaguor d.sala	бӀагӀо р х.а.а.а.	IMP(IMP)																															
PRE	bwaguor d.oal	бӀагӀо р х.о.а.а.	YN																															
PRE:NEG			YN																															
IMP			YN																															
PL			PP:ɾ																															
CS:m		CS:m	bwaguor d.eanna бӀагӀо р х.а.а.а.а.																															
on temp	bwaguor d.aalar	бӀагӀо р а.а.а.а.																																
notes	Sy kuorta bwaguor ux / boal. Aspect: bwaguor boal 'get dizzy'; bwaguor boagha 'am dizzy (right now)'																																	
other_notations	bwaguor																																	

Figure 4. Record for a compound verb.

A number of verbs distinguish the pluractional categories of iterative and/or plural S/O, which are usually marked by vowel ablaut. Here too we have hand-entered the corresponding iterative for simulactives, simulactive for iteratives, etc. and provided the

fields with “Go” buttons. These again are not relationalized yet, and return not just the counterpart to the verb in question but all parallel derivatives.

The lemma fields, mentioned earlier, have mostly not been filled yet. What is entered there is the ultimate root for each morpheme. A singular-plural pair or a simulfactive-iterative pair share the same lemma, and so do all their prefixal and phrasal derivatives, deverbal nouns, lexicalized participles, etc. The root of any prefix, any auxiliary, or any derivational suffix is also entered. Eventually, when the lemma fields are all filled, the user will be able to bring together all words sharing a single root morpheme, for vocabulary enhancement or morphological research. In December 1999 Sprouse (who created this part of the database and began filling it) printed out an 84-page draft lemma-based dictionary (Sprouse 1999) which grouped words together under the lemmas for each of their component morphemes and printed the words in citation form with part of speech and gloss.

The verb record page also has fields for the various causative and inceptive derivatives of Ingush verbs. (These are regular and numerous in Ingush, but are being filled in gradually as we come across them in texts, rather than mechanically elicited for each verb or mechanically taken from existing dictionaries, which usually cite them with single glosses that hide the considerable extent of their lexicalization, idiomaticity, deponence, and other irregularities.) This part of the database is now relationalized, so any entries (such as the causative verb on Fig. 3, the record for *laattar* ‘stand’) actually displays in portals which access the separate records of those verbs. Given the high morphological productivity and the significant lexical idiosyncrasy of these derivatives, hand-entering the information on the records of their radicals would be a poor use of time and duplicate information needed on their own records.

Returning to matters of glossing, most of our records have several glosses, so that the English dictionaries contain about 8000 words each, as against respectively 5600 and 5800 for Chechen and Ingush. This was done deliberately, and we often used a thesaurus in glossing precisely in order to expand the English glosses. This way, the usefulness of the dictionaries to members of the Chechen and Ingush communities who need them to read English grows faster than their usefulness to English-speakers studying Chechen and Ingush, at very little extra effort on our part.

Cyrillic spellings were generated automatically after the dictionaries had reached about 5000 records, with a program written by Sprouse. The Cyrillic orthographies are ingenious and economical in their use of Cyrillic letters to write the Chechen and Ingush phonemes (which considerably outnumber the Cyrillic letters), but underrepresent the vowel phonemes, generally not indicating length and merging some vowels with diphthongs. Our Latin system distinguishes these. Were it just a matter of converting paradigmatic representation of phoneme inventories from Latin to Cyrillic, the matter would be straightforward, as the Cyrillic spelling represents loss of information. (That statement describes Chechen. Ingush has had undergone some sound changes that further

distance it from the orthography and require more elaborate contextual specifications for autotransliteration.) But in syntagmatic context matters are much more complex, as both languages have a schwa which is often unpronounced but influences the pronunciation of a preceding vowel, opening the syllable and allowing length to surface (while shortening occurs in closed syllables). Some minimally contrasting words, with phonetic transcriptions, from Ingush:

tq'aam	[tʰɑ́ ʔám]	'wing'
chaam	[tʰɑ́ ʔám]	'taste'
gaam	[tʰɑ́ ʔám]	'chisel'
zaama	[tʰɑ́ ʔám]	'time, era'
ghaama	[tʰɑ́ ʔám]	'shock of hay'

When endings are added, schwas prove to be vowel-zero alternations, surfacing as zero in open syllables and centralized short vowel in closed syllables, as in these case forms from Chechen:

transliterated Cyrillic	phonetic transcription	our spelling		
bugha	[bʊɑ́]	bugha	'bull'	nominative
bughan	[bʊɑ́ ʔn]	bughan		genitive
bughanna	[bʊɑ́ ʔn . ʔn]	bughanna		dative
bughanuo	[bʊɑ́ ʔn . ʔn]	bughnuo		ergative
bughanie	[bʊɑ́ ʔn . ʔn]	bughnie		allative

The vowel-zero alternations seem to have been fully phonemicized in the standard dialect of Chechen, so our spelling generally reflects them. The Cyrillic orthography writes them all, reflecting pronunciation that can still be heard in some dialects. Converting automatically from Latin to Cyrillic in forms like these is obviously not straightforward, and the Cyrillic autotransliteration required a good deal of checking and editing, though of course it was much faster than hand entry. Automatic addition of Cyrillic spellings was a one-time undertaking. Since then we have hand-entered Cyrillic. Since the rate of lexical growth has slowed down to a few hundred new words per year, the additional time is not burdensome.

Overall, the Cyrillic orthographies for Chechen and Ingush are non-phonemic in various ways: they fail to distinguish several of the vowel phonemes and their spelling of schwa is more abstract than phonemic. Maciev 1961 indicates phonetic length but makes no other phonemic clarifications, and other dictionaries do not even do this. In addition to these problems of inferring pronunciation from spelling, morphological change has altered conjugations and declensions in many words, often to the point of obscuring the original

ablaut classes and setting up phonological reanalysis as well. Even the phonemically transcribed words and paradigms in Imnaishvili 1977 contain enough inconsistencies and typographical errors that they had to be checked with speakers. For these reasons, every word from published sources needs to be checked with one or more speakers. Lexicography therefore proceeds more slowly than it would were the languages a century or two either behind or ahead in their morphophonemic developments.

## 7. Generating dictionaries from the database

A great advantage of storing lexical data in a relational database is that multiple customized presentations of the lexicon can be produced in a more-or-less automatic fashion. For the Chechen and Ingush projects, we have two primary kinds of presentation: 1) publication-quality traditional print dictionaries; 2) on-line searchable dictionaries.

### 7.1 The print dictionary

The print dictionaries are produced from FileMaker Pro exported merge files (i.e. delimited text files). The raw data is processed with several Perl scripts that output LaTeX files from which publication-quality Postscript files are produced.

The dictionaries consist of three chapters each: 1) Chechen/Ingush-English with headwords in Latin orthography; 2) Chechen/Ingush-English with headwords in Cyrillic orthography; 3) English-Chechen/Ingush. These chapters contain similar elements, but each requires special treatment. The English-Chechen/Ingush chapter is the most complicated to produce, as English headwords are automatically generated from the glosses of Chechen/Ingush lexical entries, a less-than-ideal but serviceable solution. A sample of the dictionary is in Figure 5. Each column represents one of the three chapters identified at the beginning of this paragraph, respectively.

<p>sid'a<sup>+</sup> v. <i>sid'a</i> (Erg-Nom) sid'a<sup>+</sup> Lift, lift up, raise. ardam n. (nb) [1] sid'a<sup>+</sup> sid'a<sup>+</sup> Action, activity, operation; mathe- mational operation. ■ deq'wasa ardam deqsarw ardam Di- vision (as a mathe- mational operation) ■ toq'wasa ardam toq'wasa ardam Mul- tiplication (as a mathematical oper- ation)</p>	<p>sid'a<sup>+</sup> v. <i>sid'a</i> (Erg-Nom) sid'a<sup>+</sup> Lift, lift up, raise. ardam n. (nb) [1] ardam sid'a<sup>+</sup> sid'a<sup>+</sup> sid'a<sup>+</sup> Action, activity, operation; mathe- mational operation. ■ deqsarw ardam ardam deq'wasa ardam Division (as a mathematical oper- ation) ■ toq'wasa ardam toq'wasa ardam Multiplication (as a mathematical operation)</p>	<p>lift v. (lf) 1. AΓa<sup>+</sup> sid'a. 2. Γa<sup>+</sup>. sid'a. 3. Kuzkalla da<sup>+</sup>, kurzalla da<sup>+</sup>. 4. AΓa<sup>+</sup>: sid'a<sup>+</sup>. Lift up v. 1. AΓa<sup>+</sup>. sid'a<sup>+</sup>. 2. AΓa<sup>+</sup>. sid'a. 3. Γa<sup>+</sup>. sid'a. action, activity n. Ardam. ardam. operation; mathe- mational operation n. Ardam. ardam.</p>
--	---	--

Figure 5. Sample columns from the three dictionary sections.

Logically, a print dictionary is composed of a list of entries, each of which is composed of a number of elements. These elements are therefore the primitives used to create the dictionary and include categories such as ‘headword’, ‘part of speech’, ‘gloss’, and ‘pronunciation’. A dictionary reader recognizes these elements through their presentation, as encoded by their relative order and by stylistic elements (e.g. character script, type size, bold type face, italic type face), or by their membership in a closed class of possible values (e.g. typical part-of-speech abbreviations; a range of Roman numerals denoting declension classes). A simplified list of the primitives used in the Chechen/Ingush dictionaries is included in Figure 6. Each has a counterpart LaTeX command:

headword  
 alternate orthography (Latin or Cyrillic)  
 part of speech  
 pronunciation  
 gender  
 aspect/number  
 conjugation class  
 valence  
 declension class  
 principal parts  
 gloss

Figure 6: Partial list of elements used in the Chechen/Ingush print dictionaries

The elements not listed in Figure 6 are mostly redundant categories that additionally specify the orthography, e.g. principal parts in Cyrillic script vs. the Latin orthography, or that pertain to subwords listed under a headword, e.g. the part of speech pertaining to the subword.

The primary purpose of the Perl scripts is to translate the semantics of the lexical database into the semantics of traditional dictionary entries. In the simplest of cases, this translation is straightforward, with a single database record mapping directly to a dictionary entry and individual fields mapping onto one of the elements. The only exception to this one-to-one mapping is gender, which is composed of two database fields collapsed into the argument of one LaTeX command. In more complicated cases, one or more database records may contain a pointer to another record, indicating that it is derived from that record. When this happens, the derived forms map onto the ‘subword’ primitives and are grouped under the entry corresponding to the headword from which they are derived.

Another kind of complication is that the content mapped to the dictionary primitives differs, depending on the chapter and part of speech. For instance, the headword of a verb entry takes the infinitive form in the first chapter (Latin orthography). The second chapter, in which headwords use the traditional Cyrillic orthography, draws from the

database's Cyrillic infinitive field. The Perl scripts include mapping tables that handle these kinds of details.

The general approach taken in creating the dictionary almost completely separates content and presentation, and the LaTeX typesetting program makes this approach easy to implement. The primitives listed in Figure 6 are defined as LaTeX commands, and these are what the Perl scripts produce as output. The details of the presentation of the dictionary elements are provided in the LaTeX preamble, where all the commands are defined. The result is that all dictionary elements are displayed in the same manner. The `\headword{}` command, for example, applies bold face to its argument.

With content and presentation separated, it is quite easy to create updated dictionaries easily. The overall look of the dictionary can be changed by simply altering the definitions of the LaTeX commands that correspond to the dictionary elements and producing a new Postscript file. Updating the content is a complementary process consisting of remapping the contents of the lexical database to the dictionary elements without worrying about the details of the presentation.

In addition to mapping the lexical database to a traditional dictionary structure, the Perl scripts perform several other important tasks. First, the database records are not sorted in proper dictionary order, and separate sorting routines have been created for the Latin and Cyrillic orthographies. The scripts also automatically generate English headwords for the English-Chechen/Ingush chapter from the gloss fields of the lexical database. In many cases, the automatic generation works perfectly, as for Chechen *zhwaela* 'dog', which appears under the headword 'dog' in the English-Chechen chapter. A significant number of glosses yield awkward results, however, and these need to be fixed. An example of an awkward gloss is one that starts with a 'little word' such as 'be', as in Chechen *qoeru* 'be afraid', which is better sorted and displayed under the headword 'afraid' than 'be'. These awkward glosses are handled by a simple rewrite rule that places the contentful word first, as in 'be afraid' → 'afraid: be afraid'. Compounds and other complex glosses present a more difficult problem. For these we take a less general but simple approach. We create a separate gloss field, the Dictionary Gloss field, containing exactly the headwords desired for the English-Chechen/Ingush chapter. If this field is non-empty, no glosses are automatically generated for that record. An example of this type is Chechen *ardam*, which includes the gloss 'mathematical operation'. Since we want this gloss to be under the English headword 'operation', the Dictionary Gloss field contains 'operation: mathematical operation'. One of the glosses for Ingush *bwaguor d.aalar* is 'be a bit crazy'. Automatic generation will remove 'be' but leave 'a bit crazy', which will be alphabetized under "a" rather than by "crazy", so the Dictionary Gloss field is used to enter 'crazy: be a bit crazy'. While this approach requires a bit of extra data entry, we find that this overall approach to automatic gloss creation strikes a comfortable balance for a project of our size between reduced programming complexity and fully-automated processing.

The Latin orthographies devised for Chechen and Ingush produce spellings that, with few exceptions, provide reliable guides to pronunciation. English orthography, of course, doesn't share this trait, and so we have attempted to provide pronunciations for Chechen/Ingush speakers of as many English headwords as possible. We have accomplished this by matching pronunciations in the machine-readable CMU Pronouncing Dictionary (<http://www.speech.cs.cmu.edu/cgi-bin/cmudict>) with our headwords. The CMU pronunciations have been recoded with the same symbols used in the Chechen/Ingush Latin orthographies so that speakers of those languages need learn only one set of symbols to take advantage of all parts of the dictionary. Fortunately, the Chechen and Ingush phonemic segment inventories are large enough to include almost all English phonemes, so English pronunciations can be reasonably approximated with the Latin orthographies. Like our automatic gloss generation, this matching yields good but imperfect results.

## **7.2 The on-line dictionary**

Sprouse also created Web-accessible on-line dictionaries (Sprouse 1997ff., Sprouse 2000ff.) from delimited text files exported from the FileMaker Pro lexical database. It consists of a very simple search interface that allows the user to enter a string to be found in the dictionary, either in Chechen/Ingush or English. The user may narrow the query by targeting only words of the specified part of speech, or by searching only through citation forms. The on-line dictionary is at its best, however, when searches are not narrowed in this way. As linguists, we often find it useful to be able to find words or parts of words regardless of their morphological form, and having the ability to search through all inflectional categories is helpful for this task. Performing the same kind of task in the FileMaker Pro™ database is much more awkward and inefficient, requiring the search string to be entered for every field to be searched.

## **8. Conclusions**

Our Ingush dictionary has taken several years to produce, but our experience with Chechen showed that, once the database structure and other organizational matters were in place, a decent lexical database could be compiled and printed out in a year. We hope that our detailed description of some of our problems and solutions will expedite lexicography for other linguists.

By now our project has surveyed all known lexical sources and has largely exhausted elicitation as a means of expanding the database. Work on texts generates new lexical items at a rate of about 2%-3% and is now becoming a large concern of our project.

In a typical case of lexical documentation of an endangered language, the availability of a comprehensive dictionary increases the language's prospects for survival by providing a teaching resource, visibly demonstrating the reality and importance of the language, and expanding the language's informational functions to include word

definitions (for defining dictionaries) or access to world languages (for bilingual dictionaries). In the case of Chechen and Ingush, the role of the external linguist is not to do typical basic description but to contribute to de-Russification and decolonialization of the language and its descriptive tradition. In ex-Soviet regions there is often interest in latinization of orthographies, but in most cases the proposed latinizations are mechanical transliterations of Cyrillic orthographies, when in fact the design principles of the Cyrillic and Latin alphabets are quite different. As linguists experienced in phonology and literate in various Latin-alphabet languages we have been able to contribute knowledge of how schwas, vowel length, diphthongs, glottal stops, [j], and other things which Cyrillic handles poorly, and awareness of how vowel and consonant digraphs function in various languages, to the task of designing an easy-to-use, unambiguous practical transcription that could well function as an orthography. Whether or not our system is ever adopted by the Chechen and Ingush communities, it has been tested as no other proposed system has and is useful as a pronunciation guide even in the Cyrillic sections of the dictionaries.

Russian uses extensive prefixal and suffixal derivation of words, including suffixal accommodation of borrowed words to Russian; both nouns and verbs are open classes. Hence the headwords in a Russian dictionary are mostly single orthographic words and we suspect that the result of several academic generations of forced confinement to Russian lexicography is an expectation that headwords in an ordinary dictionary are single orthographic words and phrasal expressions are either illustrations of these, idioms and the like, or material for phraseological dictionaries, a common Russian genre. In Chechen and Ingush, however, many lexical items consist of two or more orthographic words, much as in languages with extensive verb serialization and phrasal compounding: e.g. the elementary notions ‘sit’ and ‘hang’ in Ingush are:

<i>wa-xeina</i>	<i>v-aaghar</i>
down-having.sat	GENDER-be.sitting
lit. ‘sits (having) sat down’	

<i>hwal-'ellaa</i>	<i>ullar</i>
up-strung	lie
lit. ‘lies strung up, lies draped’	

Very many of our headwords are like these – phrasal-looking but single, often very basic, lexical items. Restricting a dictionary mostly to single orthographic words gives the impression that Chechen and Ingush have smaller lexicons than, say, Russian, and we hope that our inclusion of multi-word lexical items will make clear just how rich and nuanced the vocabularies of these languages are.

Finally, we note that we have found a mixture of free software and commercial database software to be excellent for our purposes. Perl, LaTeX, and Apache (for web serving) are all free software. In an ideal world a research project such as ours would use free software

exclusively, primarily because free software by its very nature supports open standards, which prolongs the shelf-life of the project's data. Free software is also available at low or no cost, an important fact for speech communities that may not be able to afford expensive commercial packages. Nevertheless, we find that for everyday work a well-documented, well-supported commercial application makes an excellent tool and interface. For long-term storage one can always export the data to non-proprietary structured text format.

**Acknowledgments.** This material is based upon work supported by the National Science Foundation under Grant No. 9616448. Additional funding for the field methods courses and the early stages of the Ingush dictionary came from the Committee on Research, the (then) Center for Slavic and East European Studies, the Dean of Humanities, the Dean of Social Sciences, the Dean of the Graduate Division, the Department of Linguistics, and private sources. We thank all of the individuals mentioned in the paper for their assistance. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the other sources of funding, the other participants in the project, or our consultants.

## References

- Awde, Nicholas, and Galaev, Muhammad. 1997. *Chechen-English, English-Chechen Dictionary and Phrasebook*. New York: Hippocrene.
- Handel, Zev J. in press. Ingush inflectional verb morphology. In *Current trends in Caucasian, East European, and Inner Asian linguistics: Papers in honor of Howard I. Aronson*, eds. Dee Ann Holisky and Kevin Tuite. Philadelphia: Benjamins.
- Imnaishvili, D. S. 1977. *Istoriko-sravnitel'nyj analiz fonetiki naxskix jazykov*. Tbilisi: Mecniereba.
- Karasaev, A. T., and Maciev, A. G. 1978. *Oersiin-noxchiin slovar' / Russko-chechenskij slovar'*. Moscow: Russkij jazyk.
- Maciev, A. G. 1961. *Noxchiin-oersiin slovar' / Chechensko-russkij slovar'*. Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovarej.
- Mal'sagov, Z. K. 1963. *Grammatika ingushskogo jazyka: Checheno-ingushskij nauchno-issledovatel'skij institut: Trudy, V*. Grozny: Checheno-Ingushskoe knizhnoe izdatel'stvo.
- Nichols, Johanna, Guliev, Issa, Handel, Zev J., and Sprouse, Ronald L. 1996. *Ingush-English and English-Ingush lexical materials: Prospectus and sample printout for an*

- Ingush-English, English-Ingush dictionary. Berkeley: University of California, Berkeley.
- Ozdoev, I. A., Maciev, A. G., and Dzhamalxanov, E. D. 1962. Ingushsko-chechensko-russkij slovar'. Groznyj: Checheno-Ingushskoe knizhnoe izdatel'stvo.
- Ozdoeva, F. G., and Kurkueva, A. X. 1980. Ersii-ghalghaai slovar' / Russko-ingushskij slovar'. Moscow: Russkij jazyk.
- Sprouse, Ronald L. 1997ff. Search the Ingush lexical database: On-line Ingush-English, English-Ingush dictionary.  
<http://ingush.berkeley.edu:7012/cgi-bin/search.pl?lang=Ingush>
- Sprouse, Ronald L. 1999. Ingush-English Lemma-based Dictionary. Berkeley: University of California, Berkeley.
- Sprouse, Ronald L. 2000ff. Search the Chechen lexical database: On-line Chechen-English, English-Chechen dictionary.  
<http://ingush.berkeley.edu:7012/cgi-bin/search.pl?copyright=agree;lang=Chechen>
- Vagapov, Arbi. 1999. Ingals-noxchiin deshnizhaina / English-Chechen Dictionary. Soelzha-Ghaala: Chechen State University.