

# Computer Supported Morphological Database for Chechen

Johannes Heinecke

Humboldt-Universität zu Berlin, Computerlinguistik

Jägerstr. 10/11, D - 10099 Berlin

Tel.: +49 - 30 - 20192 553, Fax: +49 - 30 - 20192 560

Email: [heinecke@compling.hu-berlin.de](mailto:heinecke@compling.hu-berlin.de)

8th Caucasian Colloquium, Leiden, June 6-8, 1996

The aim of this paper is to present and to demonstrate the application of a computer software system creating and maintaining a dictionary of the Chechen and German languages.

Currently Chechen language resources can usually only be accessed through the medium of Russian. This may lead to problems especially for linguists who only have a restricted knowledge of this language.

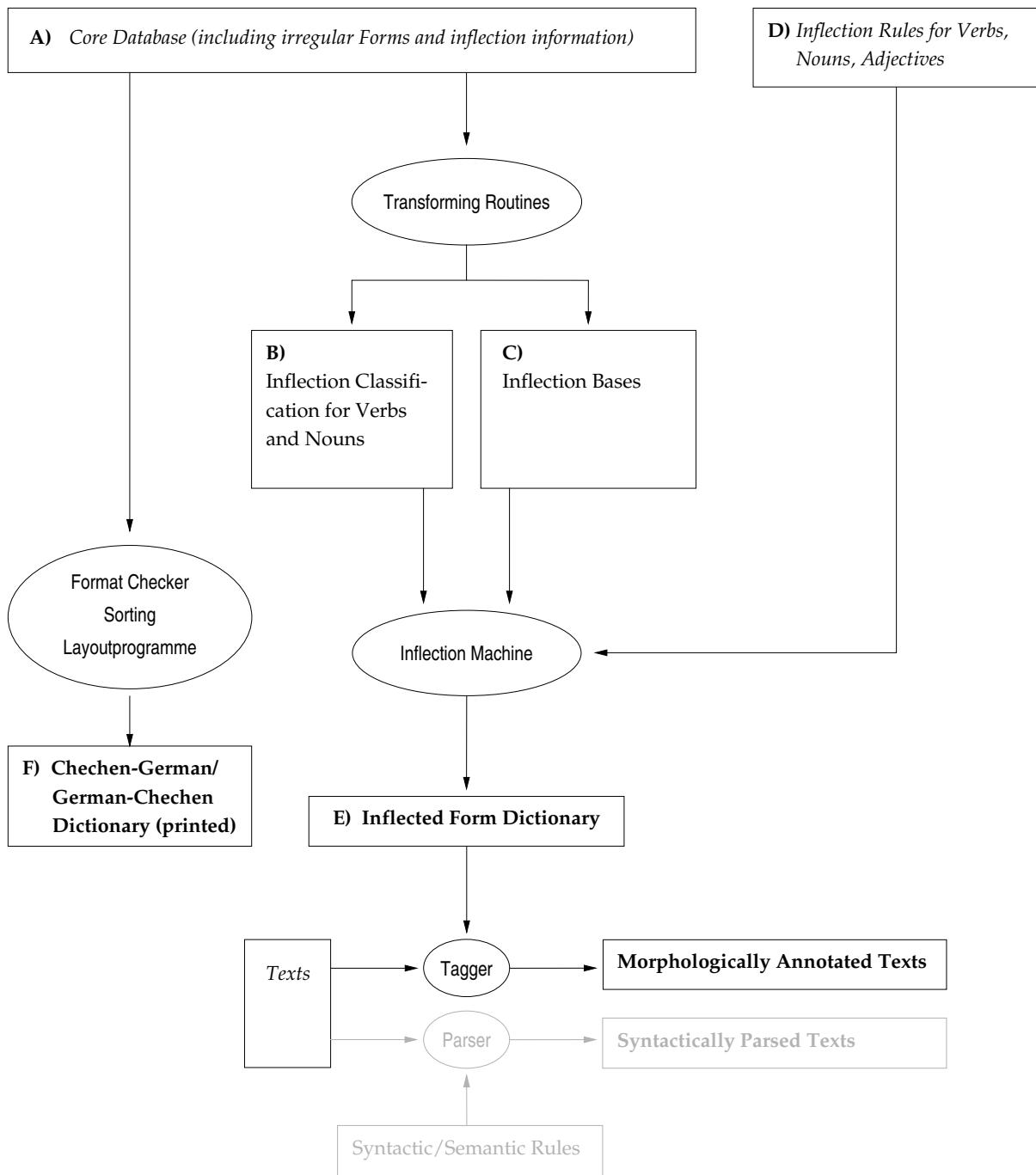
During studies on Chechen a lot of material was collected, so the idea arose to build up a Chechen-German database, including idioms and sayings. In order to administrate the data, a dictionary software has been developed and implemented. This software is also capable to prepare a printed version of a Chechen-German (and German-Chechen) dictionary.

In addition to the software described above, the lexicon formalism, developed at Humboldt-University Berlin within the "Verbmobil" project<sup>1</sup> is able to generate all inflected forms (verbs and nouns; including compound forms), drawing the information from specially designed feature structures and rules. This makes it relatively easy to prepare a database for all inflected forms which can be accessed to retrieve information of non-standard forms.

---

<sup>1</sup>A project funded by the German Federal Ministry of Education, Science, Research and Technology (BMBF) on machine translation of spontaneous speech, Grant-No. 01 IV 101 G

# 1 Flow of Information



## 2 States of Information Processing

### 2.1 Core Database (A)

```
дIад*блō#/dvolvo/#Verb#дIадблад*о#дIадблийра#дIадблийна# # #herantreten; beginnen#
д*оца#/dwoca/#Adj#kurz#
дика#/dika/#Adj#gut#Иза (чIогIа) дика ду - Das ist (sehr) gut
нохчи#/nwoχt/i/#Subst#нохчичуи#нохчичуи#нохчиб#нохчиий#1/2#TschetschenIn#
ЦIа#/c'a/#Subst#ЦIийнан#ЦIенна#ЦIийн#ЦIенош#3#Haus#
нохчи#/nwoχt/wo/#>nokhchi#TschetschenIn#
ббда#/buoda/#Subst#2#ббданан#ббданна#ббданб#ббдам#5#Finsternis, Dunkelheit#
со#/swo/#Pron#сан#сүна#ас#соыца#сох#сол#соыга# #ich#
```

### 2.2 Inflection Classifications for Verbs and Nouns (B)

Noun entries supply *NOM*, *GEN*, *DAT*, *ERG*, *NOM-PL*. The first of the following example rule systems (currently 25 noun inflection classes) generates the full paradigm for nouns of the “chunna” class (e.g. кечдархб “activist”, нохчи “chechen”):

- (1) *Condition* DAT: -чунна#  
⇒ *Type* chunna
- (2) *Condition* NOM: -C#, NOM PL: -CШ#, DAT: -Сна#  
⇒ *Type* 1

Verbs are distinguished into derived and base forms. Derived verbs will be inflected using different inflection algorithms (examples for the three types are гочд\*ан “to translate”, малб “to make (somebody) drink”, д\*бла “to begin”):

- (3) *Condition* INF: -д\*ан#  
⇒ *Type* dan-composed
- (4) *Condition* INF: -б#  
⇒ *Type* kaus-verb  
⇒ *Stem* INF – -б
- (5) *Condition* INF: PREFIX + C[C] + V + C[C] + a/u#  
⇒ *Type* regular-assimilation  
⇒ *Consonants* д\*, л  
⇒ *Ablaut* Infinitive: б (a, Vowelharmony)  
Present: у (u, Vowelharmony)  
Witnessed Past: ѿ ira  
Perfective Past: ѿ + C2-Assimilation + a

## 2.3 Inflection Bases (C)

All information which cannot be retrieved from rules must be coded into the inflection bases, e.g.:

The following examples are coded using the lexiconformalism (programming language) *L<sup>E</sup>V44*:

```

base 'дIад*блō' :<< verb_c >>:
    deutsch: 'herantreten; anfangen' &
    phonol: 'dvolvo' &
    konjug: 'kaus-verb' &
    praefix: 'дIа' &
    c1: 'д*бл' .                                % Stem to be used for inflection

base 'нохчи' :<< subst_c >>:
    deutsch: 'TschetschenIn' &
    phonol: 'nwoχtfi' &
    flex: 'chunna' &                            % Inflection group
    nom: 'нохчи' &
    nom_pl: 'нохчий' &
    klasse: ('1' \ '2')                          % Class (v/db or j/db)
    pl_stamm: 'нохчи' .                         % Stem for plural paradigm

base 'дIа' :<< subst_c >>:
    deutsch: 'Haus' &
    phonol: 'ça' &
    flex: 'vsh' &                               % Inflection group
    nom: 'дIа' &
    gen: 'дIийнан' &                           % Ablaut phenomena must be coded
    dat: 'дIенна' &
    erg: 'дIийнб' &
    nom_pl: 'дIенош' &
    klasse: ('3') &                            % Class (d/d)
    sg_stamm: 'дIен' &
    pl_stamm: 'дIен' .
```

```

base 'д*оца' :<< adj_c >>:
    deutsch: 'kurz' &
    phonol: 'd*woca' &
    nom: 'доца' .
```

## 2.4 Inflection Rules for Verbs, Adjectives, Nouns etc. (D)

### • Adjective Inflection Rules

After being checked, whether the adjective is gender agreeable the full inflection suffixes are appended for both, dependent and independent forms:

Firstly, the gender agreement is affigated to the root:

$$(6) \text{ \texttt{д}*+ROOT} \Rightarrow \text{б+ROOT, в+ROOT, д+ROOT, й+ROOT (я/ю/e)}$$

e.g.:

$$(7) \text{ \texttt{д}*оца ("short")} \Rightarrow \text{боца, воца, доца, йоца}$$

$$(8) \text{ \texttt{д}*аъсса ("empty, free")} \Rightarrow \text{баъсса, ваъсса, даъсса, яъсса}$$

Secondly, the case suffixes are added onto the stem:

(9)	independent		dependent
	SG:	PL:	SG/PL:
NOM:	-ниг	-наш	
GEN:	-чун	-чёран, -чийн	-чу
ERG:	-чб	-чара, -чуша	
DAT:	-чунна	-чарна, -чушка	
INSTR:	-чуюнца	-чаърца, -чушца	
MAT:	-чух	-чарах, -чех	
KOMP:	-чул	-чарал, -чел	
ALL:	-чуюнга	-чаърга, -чушка	

### • Noun Inflection Rules

The traditional division into four inflection classes unfortunately does not hold for computerized processing. Even without taking um- and ablaut phenomena into account 25 different classes are necessary for automatically inflecting nouns. Generally the dative stem can be taken for the singular paradigm, the nominative plural for the plural paradigm.

(10)	Type	chunna
	⇒ SG-stem	NOM
	⇒ PL-stem	NOM PL -ой
	⇒ Paradigm	MAT-SG: SG-stem + -чух KOMP-SG: SG-stem + -чул INSTR-SG: SG-stem + -чуюнца ALL-SG: SG-stem + -чуюнга
	GEN-PL:	PL-stem + -ойн
	DAT-PL:	PL-stem + -ошна
	ERG-PL:	PL-stem + -бша
	MAT-PL:	PL-stem + -ойх
	KOMP-PL:	PL-stem + -ойл
	INSTR-PL:	PL-stem + -ошца
	ALL-PL:	PL-stem + -ошка
	...	

- Verb Inflection Rules

As verbs can have more than one slot for gender agreement (cf. *д\*охд\*ан* “to warm up”), the inflection is carried out before replacing ‘д\*’, e.g.:

(11)	Type	regular	
	⇒ Present	consonants + present-vowels	
	⇒ Witnessed Past	consonants + witn.-past-schema	
	⇒ Perfective Past	consonants + perf.-past-schema	
	⇒ Paradigm	Close Past Remote Past Repeated Past ...	Witnessed Past – -pa Perfective Past + -ē- Present + -pa
		Present Participle	Present + case-suffixes
		Perfect Participle	Witnessed Past + case-suffixes
		Masdar	Present + case-suffixes
(12)	Type	каус-verb	
	⇒ Paradigm	Present stem + ад*о Close Past stem + ий Perfective Past Close past + на Witnessed Past Close past + ра Remote Past Close past + нēра Future stem + ад*ер ...	
		Present Participle Present + case-suffixes	
		Perfect Participle Witnessed Past + case-suffixes	
		Masdar Present + case-suffixes	

Finally the class marker ‘д\*’ is replaced by its appropriate forms (cf. Adjective Inflection Rules).

### 3 Automatically generated results

#### 3.1 Morphological Database (E)

(13) Inflected Form # Wordclass Tense Class (Aspect) # Baseform # German  
длабôлабо # VERB praesens b VAR # длад\*блô # anfangen  
длабôлô # VERB infinitiv b VAR # длад\*блô # anfangen  
длабôлë # VERB imperativ b VAR # длад\*блô # anfangen

...

длайблайо # VERB praesens j VAR # длад\*блô # anfangen  
длайблô # VERB infinitiv j VAR # длад\*блô # anfangen

...

(14) Inflected Form # Wordclass Case Number Class # Baseform # German  
ценах # SUBST materialis sg d # цла # Haus  
ценал # SUBST komparativ sg d # цла # Haus  
ценца # SUBST instrumental sg d # цла # Haus  
ценийин # SUBST genitiv pl d # цла # Haus

...

(15) Inflected Form # Wordclass Case Number Class State # Baseform # German

боذاчёран # ADJ genitiv pl b unab # д\*оца # kurz  
боذاчô # ADJ ergativ sg b unab # д\*оца # kurz  
боذاчарьца # ADJ instrumental pl b unab # д\*оца # kurz  
боذاчарьга # ADJ lokativ pl b unab # д\*оца # kurz  
боذاчарна # ADJ dativ pl b unab # д\*оца # kurz  
боذاчех # ADJ materialis pl b unab # д\*оца # kurz  
боذاчел # ADJ komparativ pl b unab # д\*оца # kurz  
боذاчийн # ADJ genitiv pl b unab # д\*оца # kurz  
боذاчу # ADJ ^nominativ pl b abh # д\*оца # kurz  
боذاчу # ADJ ^nominativ sg b abh # д\*оца # kurz  
боذاчух # ADJ materialis sg b unab # д\*оца # kurz  
боذاчул # ADJ komparativ sg b unab # д\*оца # kurz

...

## 3.2 Sample page of the printed dictionary (F)

дІА	дІахъажо
<b>дІа-</b> /d̥a:/ <i>Präfix: dort-, dahin-, weg-, (vom Sprecher).</i>	<b>дІахъача</b> /d̥aqa:t/a/ <i>Verb (Präs. дІахъбчу, Prätbez. дІахъчира, Präpf. дІахъчна, Akt. SG): ankommen. (russ.: прийти, дойти, доехать.)</i>
<b>дІа-</b> /d̥a:/ <i>Adv.: dort, dorthin.</i>	<b>дІахъийса</b> /d̥aqi:sa/ <i>Verb (Präs. дІахъуйсу, Prätbez. дІахъйсира); werfen, schleudern. (russ.: бросить.)</i>
<b>дІад*ала</b> /d̥ada:l/a/ <i>Verb (Präs. дІадблу, Prätbez. дІадлира, Präpf. дІадълла, Fut. дІадльр, Akt. SG): zu Ende sein, sterben, weggehen. (russ.: кончиться, умереть, уйти.)</i>	<b>дІанисд*ан</b> /d̥anisdan/ <i>Verb (Präs. дІанисд*о, Prätbez. дІанисд*ира, Präpf. дІанисд*ина): unterbringen, aufstellen. (russ.: разместить.)</i>
<b>дІад*ан</b> /d̥ada:n/ <i>Verb (Präs. дІало, Prätbez. дІадера, Präpf. дІадеана, Akt. SG): losgehen. (russ.: покинуть, идти.)</i>	<b>дІарани</b> /d̥aranig/ <i>Pron. (G дІарачун, D дІарачунна, Е дІарачо, I дІарачунца, МАТ дІарачух, К дІарачул, А дІарачунга, PL N дІаранаш): jener. (Genitiv PL дІарачэн, дІарачин; Ergativ PL дІарачара, дІарачушна; Dativ PL дІарачарна, дІарачушна; Instrumental PL дІарачарьца, дІарачушца; Material PL дІарачарех, дІарачех; Komparativ PL дІарачэрел, дІарачел; Allativ PL дІарачьрга, дІарачушка)</i>
<b>дІад*ан</b> /d̥adan/ <i>Verb (Präs. дІадо, Prätbez. дІадира, Präpf. дІадина, Fut. дІадер, Akt. SG): machen.</i>	<b>дІаса</b> /d̥asa/ <i>Präfix: mal. (russ.: раз.)</i>
<b>дІад*аха</b> /d̥ada:xa/ <i>Verb (Präs. дІадъду, Prätbez. дІадхира, Präpf. дІадахна, Akt. PL): sich ausziehen. (russ.: раздеваться.)</i>	<b>дІасхъа-</b> /d̥asha/ <i>Präfix: hin/weg. (Bewegung zum/vom Sprecher)</i>
<b>дІад*аха</b> /d̥ada:xa/ <i>Verb (Präs. дІадхара, Prätbez. дІадхарна, Akt. SG): weggehen (vom Sprecher weg). (russ.: отойти.)</i>	<b>дІасхъакхийса</b> /d̥ashaq:sa/ <i>Verb (Präs. дІасхъакхуйсу, Prätbez. дІасхъакхийсира): gleichmäßig durcheinander bringen. (russ.: разбросывать.)</i>
<b>дІадър</b> /d̥adär/ <i>siche дІад*ала (zu Ende sein, sterben, weggehen)</i>	<b>дІасхъакхъаста</b> /d̥ashaq:asta/ <i>Verb (Präs. дІасхъакхъаста, Prätbez. дІасхъакхъистири): auseinander gehen. (in verschiedene Richtungen)</i>
<b>дІад*ийца</b> /d̥adic:a/ <i>Verb (Präs. дІадувьцу, Prätbez. дІадийцира, Präpf. дІадийцина): erzählen.</i>	<b>дІахъажа</b> /d̥ahaga/ <i>Verb (Präs. дІахъожу, Prätbez. дІахъажира, Präpf. дІахъажна, Akt. SG): nicht beachten (NEG); eine Zeitlang sehen. (russ.: посмотреть.)</i>
<b>дІадъоб</b> /d̥advolvo/ <i>Verb (Präs. дІадблод*, Prätbez. дІадблрия, Präpf. дІадблйна): herantreten; beginnen, anfangen. (russ.: приступить, начать.)</i>	<b>дІахъажо</b> /d̥ahago/ <i>Verb (Präs. дІахъожу, Prätbez. дІахъажад*, Präpf. дІахъажира,</i>
<b>дІад*олор</b> /d̥adcolor/ <i>Masdar (Kl. 3): Unternehmen, Vorhaben.</i>	
<b>дІакхайкха</b> /d̥aqajqa/ <i>Verb (Präs. дІакхойку, Prätbez. дІакхайкира, Präpf. дІакхайкхина): besiege rufen. (um</i>	

## 4 Some Numbers

Currently the Core Database contains 2,000 entries, including 500 verbs, 600 nouns and 170 adjectives, which are inflected.

The automatically generated Morphological Database contains 95,000 forms, including 10,000 nominal, 76000 verbal 7,600 adjectival forms.

The rich morphology, especially the verbal paradigm, is responsible for the huge number of generated forms, cf. (д\*бла “to begin, to start”):

Tenses/Moods (7 Forms, with class agreement ⇒ 28 Forms)

(16)	Imperative	д*бла
	Present	д*улу
	Future	д*улур
	Close Past	д*уйли
	Perfective Past	д*оълла
	Witnessed Past	д*уйлира
	Repeated Past	д*улура
	...	

Participles (Present, 23 Forms, with class agreement ⇒ 92 Forms):

(17)	Independent			Dependent Singular/Plural
		Singular	Plural	
Nominative	д*улург	д*улурш		д*улу
Genitive	д*улучун	д*улучёран, д*улучийн		д*улучу
Ergative	д*улучб	д*улучара, д*улучуша		
Dative	д*улучунна	д*улучарна, д*улучушна		
Instrumental	д*улучувьнца	д*улучарьрца, д*улучушца		
Materialis	д*улучух	д*улучех		
Comparative	д*улучул	д*улучел		
Allative	д*улучувьнга	д*улучарьрга, д*улучушка		
...				

(Past 23 Forms, with class agreement ⇒ 92 Forms):

(18)	Independent			Dependent Singular/Plural
		Singular	Plural	
Nominative	д*оълларг	д*оълларш		д*оълла
Genitive	д*оъллачун	д*оъллачёран, д*оъллачийн		д*оъллачу
Ergative	д*оъллачб	д*оъллачара, д*оъллачуша		
Dative	д*оъллачунна	д*оъллачарна, д*оъллачушна		
Instrumental	д*оъллачувьнца	д*оъллачарьрца, д*оъллачушца		
Materialis	д*оъллачух	д*оъллачех		
Comparative	д*оъллачул	д*оъллачел		
Allative	д*оъллачувьнга	д*оъллачарьрга, д*оъллачушка		
...				

Masdar (16 Forms, with class agreement ⇒ 64 Forms):

	Singular	Plural
Nominative	д*бла́р	д*бла́рш
Genitive	д*бла́ран	д*бла́рийн
Ergative	д*бла́рб	д*бла́рша
Dative	д*бла́рна	д*бла́ршина
Instrumental	д*бла́рца	д*бла́рща
Materialis	д*бла́рах	д*бла́рех
Comparative	д*бла́рал	д*бла́рел
Allative	д*бла́ре	д*бла́ршка
...		

This results in 276 Forms, having not yet taken into account gerunds etc.

## References

- Beerle, Werner:** A contribution to the Morphology of the Simple Verb in Chechen.  
In: *Thordarson, F. (Hrsg.): Studia Caucasiologica I: Proceedings of the Third Caucasian Colloquium, Oslo, July 1986.* Norwegian University Press: Oslo 1988.
- Gebhardi, Gunter; Heinecke, Johannes:** Lexikonformalismus *L&gt;V44*. Verbmobil  
Technisches Dokument 19. Humboldt-Universität zu Berlin 1995 .
- Nichols, Johanna:** Chechen. In: *Smeets, Rieks (Hrsg.): The Indigenious Languages of the Caucasus, Vol 4: The North East Caucasian Languages II.* Caravan Books: Delmar 1994.
- Арсаханов, Исрапил:** Хинцалера Нохчийн Мотт. Лексикологи, Фонетика,  
Морфологи. Нохч-Галгайн Книжни Издательство: Грозный 1965.
- Дешериев, Юнус Д.:** Чеченский Язык. In: *Виноградов, В.В. et al. (eds): Языки народов СССР, Том 4: Иберийско-Кавказские Языки.* Академия Наук: Москва 1967.
- Карасаев, А. Т., Мациев, Ахмат Гехаевич:** Русско-Чеченский Словарь. Русский Язык: Москва 1978.
- Мациев, Ахмат Гехаевич:** Чеченско-Русский Словарь. Государственное Издательство Иностранных и Национальных Словарей: Москва 1961.
- Мациев, Ахмат Гехаевич:** Очерки Лексикологии Современного Чеченского Языка. Грозный 1973.
- Мациев, Ахмат Гехаевич; Оздоев, И.А.:** Русско-Чеченско-Ингушский Словарь. Чечено-Ингушское Книжное Издательство: Грозный 1966.
- Яковлев, Николай Феофанович:** Морфология Чеченского Языка. Чечено-Ингушское Книжное Издательство: Грозный 1960.