

Машинная Морфологическая База Данных Чеченского Языка*

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

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

Целью этой работы является продемонстрировать применение системы программного обеспечения для создания и использования словаря чеченского и немецкого языков. В настоящее время чеченский язык доступен только через посредство русского. Это проблематично, если лингвист не достаточно знаком с русским.

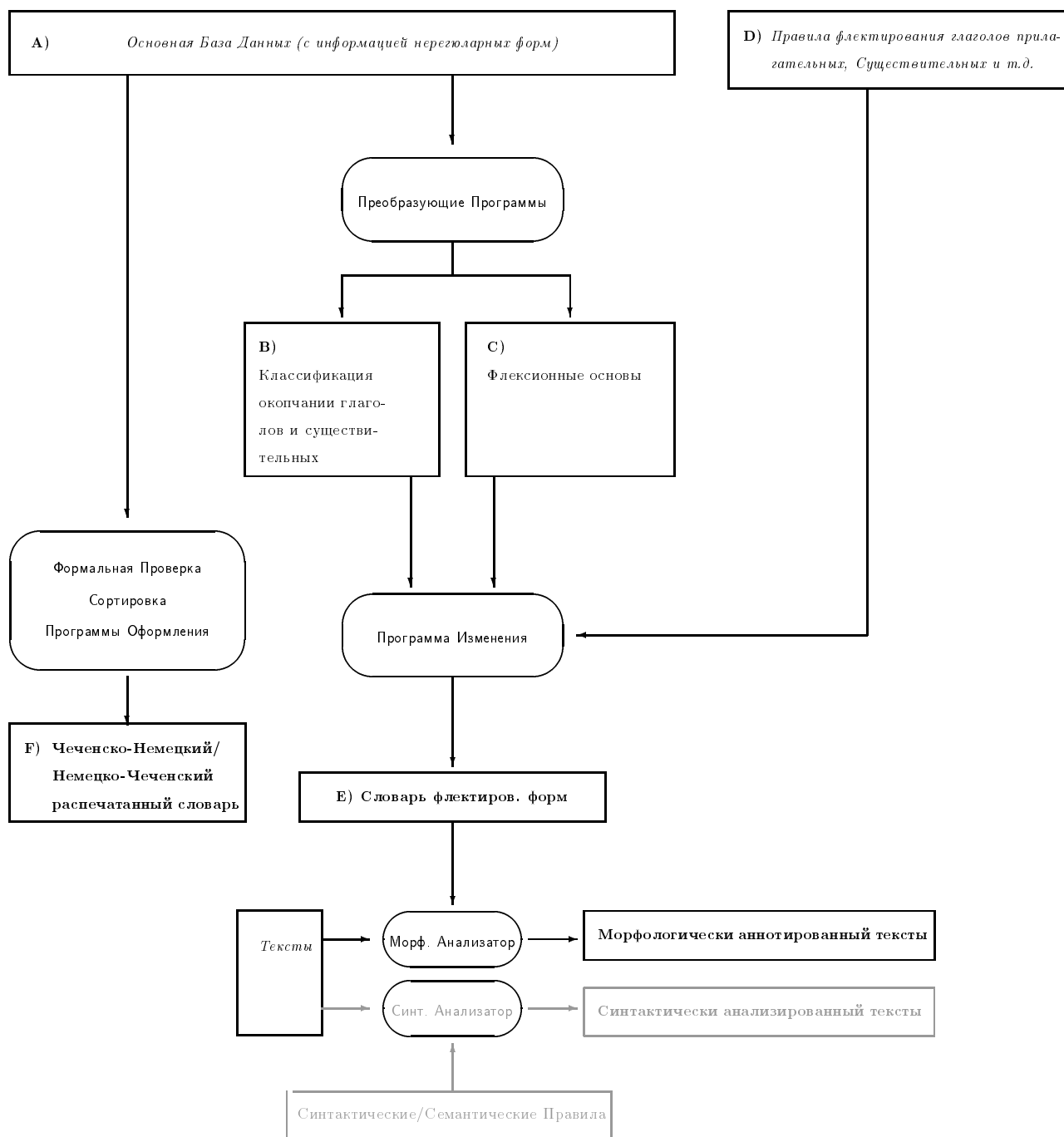
При изучении чеченского было собрано много материала, так что возникла идея создать чеченско-немецкую базу данных, включающую идиомы и пословицы. Для управления этой базой данных была разработана и имплементирована система программного обеспечения. Эта система может также подготовить печатную версию чеченско-немецкого (а также немецко-чеченского) словаря.

В дополнение к вышеописанным программам, формализм лексикона, разработанный в университете имени Гумбольдта в рамках проекта "Verbmobil"¹, может генерировать все морфологические формы (глаголов и существительных, включая составные формы), опираясь на информацию в специально созданных структурах признаков и правилах. Это позволяет довольно легко подготовить базу данных для всех флектированных форм, которую можно использовать для получения информации о нерегулярных формах.

* Др. Анатолий Стригин (Берлин) переводил это handout

¹ Проект по машинному переводу устной речи финансируется германским министерством просвещения, науки и технологии (BMBF) Grant Nr. 01 IV 101 G

1 Поток Информации



2 Ступени Обработки Информации

2.1 Основная База Данных (А)

д|ад*бл̄б̄#/dvo|o/#Verb#д|адб̄лад*о#д|адб̄лийра#д|адб̄лийна# #herantreten; beginnen#
д*оца#/dwoca/#Adj#kurz#
дика#/dika/#Adj#gut#Иза (ч|ог|а) дика ду - Das ist (sehr) gut
нохчи#/nwoχtʃi/#Subst#нохчичун#нохчичунна#нохчичб̄#нохчий#1/2#TschetschenIn#
ц|а#/c'a/#Subst#ц|ийнан#ц|енна#ц|ийнб̄#ц|еннош#3#Haus#
нохчо#/nwoχtʃwo/#>nokhchi#TschetschenIn#
б̄да#/bwoda/#Subst#2#б̄данан#б̄данна#б̄данб̄#б̄даш#5#Finsternis, Dunkelheit#
со#/swo/#Pron#сан#с̄уна#ас#соьца#сох#сол#соьга# #ich#

2.2 Классификация Окончаний Глаголов и Существительных (В)

Леммы для существительных дают *Именительный, Родительный, Дательный, Ергатив, Именительный-Множ*. Пример со следующими правилами (25 классов окончаний в настоящее время) генерирует полную парадигму существительных класса “chunna”: *кечдарх̄* “активист”, *нохчи* “чеченец”):

- (1) *Кондиция* ДАТ: -чунна#
⇒ *Тип* chunna
- (2) *Кондиция* ИМ: -К#, ИМ (Мн): -Кш#, ДАТ: -Кна#
⇒ *Тип* 1

У глаголов различаются основные и производные формы. Производные глаголы получаются путем применения различных флектирующих алгоритмов (примеры для трех классов: *гочд*ан* “переводить”, *мал̄б̄* “дать выпить”, *д*б̄ла* “начинать”):

- (3) *Кондиция* Инфинитв: -д*ан#
⇒ *Тип* dan-composed (словосложение д*ан)
- (4) *Кондиция* Инфинитив: -б̄#
⇒ *Тип* kaus-verb (каузативный глагол)
⇒ *Основа* Инфинитив - б̄
- (5) *Кондиция* Инфинитив: Префикс + С[С] + V + С[С] + а/у#
⇒ *Тип* regular-assimilation (закономерный-ассимиляция)
⇒ *Согласные* д*, л
⇒ *Аблаут* Инфинитив: б̄ (а, Гармония Гласных)
Настоящее: у (у, Гармония Гласных)
Очевиднопрошедшее: ü ira
Прошедшее: ö + K2-Ассимиляция + а

2.3 Флекссионные Основы (С)

Вся информация, которая не может быть получена с помощью правил может быть закодирована в базе окончания, напр.: основа используемая при флексии группа окончания основа для парадигмы единственного числа аблаут должны быть закодированы.

Эти примеры на языке программирования *ℒA*:

```
base 'ḍIaḍ*ōlō' :<< verb_c >>:
  deutsch: 'herantreten; anfangen' &          % Немецкий перевод
  phonol: 'dvolxo' &
  konjug: 'kaus-verb' &                       % Класс спряжения
  praefix: 'ḍIa' &
  c1: 'ḍ*ōl' .                               % Основа для спряжения
```

```
base 'нохчи' :<< subst_c >>:
  deutsch: 'TschetschenIn' &
  phonol: 'nwoχtʃi' &
  flex: 'chunna' &                            % Класс склонения
  nom: 'нохчи' &                               % Основа для склонения един. числа
  nom_pl: 'нохчий' &
  klasse: ('1' \ '2') % Грамматический Класс (e/ḍb или ĩ/ḍb)
  pl_stamm: 'нохчи' . % Основа для склонения много. числа
```

```
base 'цIa' :<< subst_c >>:
  deutsch: 'Haus' &
  phonol: 'ça' &
  flex: 'vsh' &                                % Класс спряжения
  nom: 'цIa' &
  gen: 'цIийнан' &                             % Нет правил для аблаута
  dat: 'цIенна' &
  erg: 'цIийнō' &
  nom_pl: 'цIēнош' &
  klasse: ('3') &                               % Грамматический Класс (ḍ/ḍ)
  sg_stamm: 'цIен' &
  pl_stamm: 'цIēн' .
```

```
base 'ḍ*оца' :<< adj_c >>:
  deutsch: 'kurz' &
  phonol: 'ḍ*woca' &
  nom: 'ḍоца' .
```

2.4 Правила Флектирования Глаголов, Прилагательных, Существительных и т.д. (D)

- **Правила флектирования прилагательных** установить, согласуется ли прилагательное в роде, затем полностью прибавить все суффиксы окончания для зависимых и независимых форм:

сначала к корню добавляется суффикс согласования в роде

(6) $\partial^* + \text{КОРЕНЬ} \Rightarrow \text{б} + \text{КОРЕНЬ}, \text{в} + \text{КОРЕНЬ}, \text{д} + \text{КОРЕНЬ}, \text{й} + \text{КОРЕНЬ}$ (*я/ю/е*)

например:

(7) $\partial^* \text{оца}$ (“короткий”) \Rightarrow *боца, воца, доца, йоца*

(8) $\partial^* \text{аьсса}$ (“пустой, свободный”) \Rightarrow *баьсса, ваьсса, даьсса, яьсса*

затем к основе добавляются суффиксы падежей

	независимый		зависимый
	Ед. число	Мн. число	Ед./Мн. число
Именительный:	<i>-ниг</i>	<i>-наш</i>	
Родительный:	<i>-чун</i>	<i>-чѣран, -чийн</i>	<i>-чу</i>
Ергатив:	<i>-чѳ</i>	<i>-чѳра, -чуша</i>	
Дательный:	<i>-чунна</i>	<i>-чарна, -чушка</i>	
Творительный:	<i>-чуьнца</i>	<i>-чаьрца, -чуица</i>	
Вещественный:	<i>-чух</i>	<i>-чѳрах, -чех</i>	
Сравнительный:	<i>-чул</i>	<i>-чѳрал, -чел</i>	
Местный:	<i>-чуьнга</i>	<i>-чаьрга, -чушка</i>	

- **Правила флектирования существительных** традиционное подразделение на четыре класса окончания к сожалению не годится для машинной обработки. Даже если не учитывать явления аблаута, необходимо 25 различных классов для автоматического флектирования существительных. Обычно можно взять основу в дательном падеже для парадигмы единственного, и именительного множественного для парадигмы множественного числа.

(10)	Тип	chunna
	⇒ Основа ед. числа	Именительный
	⇒ Основа мн. числа	Именительный Мн. -ой
	⇒ Парадигма	ВЕЩ (Ед.): Основа ед. + -чух СРАВ. Ед.: Основа ед. + -чул ТВОР. Ед.: Основа ед. + -чуьнца МЕСТ. Ед.: Основа ед. + -чуьнга РОД. Мн.: Основа мн. + -ойн ДАТ. Мн.: Основа мн. + -ошна ЕРГ. Мн.: Основа мн. + -оша ВЕЩ. Мн.: Основа мн. + -ойх СРАВ. Мн.: Основа мн. + -ойл ТВОР. Мн.: Основа мн. + -ошца МЕСТ. Мн.: Основа мн. + -ошка ...

• Правила флектирования глаголов

Поскольку у глаголов может быть более одного места для согласования в роде (ср. $\partial^*ox\partial^*an$, “разогреть”), флектирование проводится до замены ‘ ∂^* ’, напр.:

(11)	Тип	regular (закономерный)
	⇒ Настоящее	Согласные + Гласные Наст.
	⇒ Очевиднопрош.	Согласные + Гласные Очевиднопрош.
	⇒ Прошедшее	Согласные + Гласные Прошедшего
	⇒ Парадигма	Недавнопрош. Очевиднопр. - -ра Давнопрош. Прошедшее + -ē- Прош. Многократ. Наст. + -ра ... Причастие Наст. Наст. + Суффиксы падежей Причастие Прош. Очевиднопрош. + Суффиксы падежей Масдар Наст. + Суффиксы падежей
(12)	Тип	kaus-verb (каузативный-глагол)
	⇒ Парадигма	Наст. Основа + ad^*o Недавнопрош. Основа + $ий$ Прош. Недавнопрош. + $на$ Очевиднопрош. Недавнопрош. + $ра$ Давнопрош. Недавнопрош. + $нēра$ Будущее Основа + ad^*er ... Причастие Наст. Наст. + Суффиксы падежей Причастие Прош. Очевиднопрош. + Суффиксы падежей Масдар Наст. + Суффиксы падежей

Наконец, указатель класса ‘ ∂^* ’ заменяется на одны из требуемых форм (ср. правила флектирования прилагательных).

3 Результаты, Полученные Автоматически

3.1 Морфологическая База Данных (E)

(13) *Слов # Часть речи Время Грамм. класс (Вид) # Основа # Немецкий*
д|а|б|о|ла|б|о # VERB praesens b VAR # д|а|д*|о|л|о # anfangen
д|а|б|о|л|о # VERB infinitiv b VAR # д|а|д*|о|л|о # anfangen
д|а|б|о|л|е # VERB imperativ b VAR # д|а|д*|о|л|о # anfangen
...

д|а|й|о|л|а|й|о # VERB praesens j VAR # д|а|д*|о|л|о # anfangen
д|а|й|о|л|о # VERB infinitiv j VAR # д|а|д*|о|л|о # anfangen
...

(14) *Слов # Часть речи Падеж Число Грамм. класс # Основа # Немецкий*
ц|е|н|а|х # SUBST materialis sg d # ц|а # Haus
ц|е|н|а|л # SUBST komparativ sg d # ц|а # Haus
ц|е|н|ц|а # SUBST instrumental sg d # ц|а # Haus
ц|е|н|и|й|н # SUBST genitiv pl d # ц|а # Haus
...

(15) *Слов # Часть речи Падеж Число Грамм. класс (Не)зав. # Основа # Немецкий*

...
б|о|ц|а|ч|е|р|а|н # ADJ genitiv pl b unabh # д*|о|ц|а # kurz
б|о|ц|а|ч|о # ADJ ergativ sg b unabh # д*|о|ц|а # kurz
б|о|ц|а|ч|а|ь|р|ц|а # ADJ instrumental pl b unabh # д*|о|ц|а # kurz
б|о|ц|а|ч|а|ь|р|г|а # ADJ lokativ pl b unabh # д*|о|ц|а # kurz
б|о|ц|а|ч|а|р|н|а # ADJ dativ pl b unabh # д*|о|ц|а # kurz
б|о|ц|а|ч|е|х # ADJ materialis pl b unabh # д*|о|ц|а # kurz
...
в|о|ц|а|ч|е|л # ADJ komparativ pl v unabh # д*|о|ц|а # kurz
в|о|ц|а|ч|и|й|н # ADJ genitiv pl v unabh # д*|о|ц|а # kurz
в|о|ц|а|ч|у # ADJ ^nominativ pl v abh # д*|о|ц|а # kurz
в|о|ц|а|ч|у # ADJ ^nominativ sg v abh # д*|о|ц|а # kurz
в|о|ц|а|ч|у|х # ADJ materialis sg v unabh # д*|о|ц|а # kurz
в|о|ц|а|ч|у|л # ADJ komparativ sg v unabh # д*|о|ц|а # kurz
...

3.2 Образец Словарной Страницы Распечатанного Словаря (F)

дІа	дІахъажō
дІа- /dʃa/ <i>Präfix</i> : dort-, dahin-, weg-, (vom Sprecher).	ihm etwas mitzuteilen)
дІа /dʃa/ <i>Adv.</i> : dort, dorthin.	дІакхъача /dʃaqa:ʃa/ <i>Verb</i> (<i>Präs.</i> дІакхъачу, <i>Prätbez.</i> дІакхъачира, <i>Prätpf.</i> дІакхъачна, <i>Akt. SG</i>): ankommen. (russ.: <i>прибѣтъ, доѣѣтъ</i> .)
дІад*ала /dʃadala/ <i>Verb</i> (<i>Präs.</i> дІадōлу, <i>Prätbez.</i> дІадēлира, <i>Prätpf.</i> дІадъялла, <i>Fut.</i> дІадъяр, <i>Akt. SG</i>): zu Ende sein, sterben, weggehen. (russ.: <i>кончитъся, умереть, ѣѣтъ</i> .)	дІакхийса /dʃaʒi:sa/ <i>Verb</i> (<i>Präs.</i> дІакхъѣсу, <i>Prätbez.</i> дІакхийсира); werfen, schleudern. (russ.: <i>бросать</i> .)
дІад*ала /dʃadala/ <i>Verb</i> (<i>Präs.</i> дІало, <i>Prätbez.</i> дІаделира, <i>Prätpf.</i> дІаделла, <i>Akt. SG</i>): stehenbleiben; überlaufen. (russ.: <i>остать, стоять, передать</i> .)	дІанисд*ан /dʃanisdan/ <i>Verb</i> (<i>Präs.</i> дІанисд*о, <i>Prätbez.</i> дІанисд*ира, <i>Prätpf.</i> дІанисд*ина); unterbringen, aufstellen. (russ.: <i>разместить</i> .)
дІад*ан /dʃadan/ <i>Verb</i> (<i>Präs.</i> дІадōлу, <i>Prätbez.</i> дІадєара, <i>Prätpf.</i> дІадєана, <i>Akt. SG</i>): losgehen. (russ.: <i>пойѣтъ, ѣѣтъ</i> .)	дІараниг /dʃaraniɡ/ <i>Pron.</i> (<i>G</i> дІарачун, <i>D</i> дІарачунна, <i>E</i> дІарачō, <i>I</i> дІарачуьнца, <i>MAT</i> дІарачух, <i>K</i> дІарачул, <i>A</i> дІарачуьнга, <i>PL</i> <i>N</i> дІаранаш); jener. (<i>Genitiv PL</i> : дІарачєран, дІарачийн; <i>Ergativ PL</i> : дІарачāра, дІарачушна; <i>Dativ PL</i> : дІарачарна, дІарачушна; <i>Instrumental PL</i> : дІарачаьрца, дІарачушца; <i>Materialis PL</i> : дІарачāрех, дІарачех; <i>Komparativ PL</i> : дІарачāрел, дІарачел; <i>Allativ PL</i> : дІарачаьрга, дІарачушка)
дІад*аха /dʃada:xa/ <i>Verb</i> (<i>Präs.</i> дІадōбьду, <i>Prätbez.</i> дІадєхира, <i>Prätpf.</i> дІадъхна, <i>Akt. PL</i>): sich ausziehen. (russ.: <i>раздеѣтъся</i> .)	дІаса- /dʃasa/ <i>Präfix</i> : mal. (russ.: <i>раз</i> .)
дІад*аха /dʃada:xa/ <i>Verb</i> (<i>Präs.</i> дІадōбьду, <i>Prätbez.</i> дІадахаара, <i>Prätpf.</i> дІадахарна, <i>Akt. SG</i>): weggehen (vom Sprecher weg). (russ.: <i>оѣѣтъ</i> .)	дІасхъа- /dʃasha/ <i>Präfix</i> : hin/weg. (<i>Bewegung zum/vom Sprecher</i>)
дІадаьр /dʃadār/ <i>siche</i> дІад*ала (zu Ende sein, sterben, weggehen)	дІасхъакхийса /dʃashaʒi:sa/ <i>Verb</i> (<i>Präs.</i> дІасхъакхъѣсу, <i>Prätbez.</i> дІасхъакхийсира); gleichmäßig durcheinander bringen. (russ.: <i>разбрасывать</i> .)
дІад*ийца /dʃadi:ca/ <i>Verb</i> (<i>Präs.</i> дІадуьйцу, <i>Prätbez.</i> дІадийцира, <i>Prätpf.</i> дІадийцина); erzählen.	дІасхъакъаста /dʃashaʒa:sta/ <i>Verb</i> (<i>Präs.</i> дІасхъакъаста, <i>Prätbez.</i> дІасхъакъастира); auseinander gehen. (in verschiedene Richtungen)
дІад*ōлō /dʃadoolo/ <i>Verb</i> (<i>Präs.</i> дІадōлад*о, <i>Prätbez.</i> дІадōлийра, <i>Prätpf.</i> дІадōлийна); herantreten; beginnen, anfangen. (russ.: <i>приступитъ, начать</i> .)	дІахъажа /dʃahaʒa/ <i>Verb</i> (<i>Präs.</i> дІахъожу, <i>Prätbez.</i> дІахъажира, <i>Prätpf.</i> дІахъажна, <i>Akt. SG</i>): nicht beachten (NEG); eine Zeitlenge sehen. (russ.: <i>неосмотреть</i> .)
дІад*ōлор /dʃadoolor/ <i>Masdar</i> (<i>Kl. 3</i>): Unternehmen, Vorhaben.	дІахъажō /dʃahaʒo/ <i>Verb</i> (<i>Präs.</i> дІахъажад*о, <i>Prätbez.</i> дІахъажйра,
дІакхайкха /dʃaʒaʒa/ <i>Verb</i> (<i>Präs.</i> дІакхойкху, <i>Prätbez.</i> дІакхайкхира, <i>Prätpf.</i> дІакхайкхина); besette rufen. (um	

21

4 Некоторые Количественные Данные

В настоящее время основная база данных содержит 2 000 словарных статей, в том числе 500 глаголов, 600 существительных и 170 прилагательных с окончаниями.

Автоматически генерированная Морфологическая база данных содержит 95 000 словоформ, включая 10 000 форм существительных, 76 000

форм глагола, 7600 форм прилагательного.

Причиной такого множества генерированных форм является богатая морфология, в особенности глагольная парадигма (напр. (д*о́ла “начинать”):

Временные формы/наклонения (7 форм, плюс согласование в классе ⇒ 28 форм)

(16)	Императив	<i>д*о́ла</i>
	Настоящее	<i>д*улу</i>
	Будущее	<i>д*улуp</i>
	Недавнoпрошедшее	<i>д*уьйли</i>
	Прошедшее	<i>д*оьлла</i>
	Очевиднопрошедшее	<i>д*уьйлиpа</i>
	Прошедшее Многократнов	<i>д*улуpа</i>
	...	

Причастие (Наст., 23 форм, плюс согласование в классе ⇒ 92 форм)

(17)		Независимый		зависимый
		Ед. число	Мн. число	Ед./Мн. число
	Именительный	<i>д*улуpг</i>	<i>д*улуpи</i>	<i>д*улу</i>
	Родительный	<i>д*улучун</i>	<i>д*улучѐран, д*улучийн</i>	<i>д*улучу</i>
	Эргатив	<i>д*улучѓ</i>	<i>д*улучѐpа, д*улучуша</i>	
	Дательный	<i>д*улучунна</i>	<i>д*улучарна, д*улучушина</i>	
	Творительный	<i>д*улучуьнца</i>	<i>д*улучаьpца, д*улучушца</i>	
	Вещественный	<i>д*улучух</i>	<i>д*улучех</i>	
	Сравнительный	<i>д*улучул</i>	<i>д*улучел</i>	
	Местный	<i>д*улучуьнга</i>	<i>д*улучаьpга, д*улучушка</i>	
	...			

(Прош., 23 форм, плюс согласование в классе ⇒ 92 форм)

(18)		Независимый		зависимый
		Ед. число	Мн. число	Ед./Мн. число
	Именительный	<i>д*оьлларг</i>	<i>д*оьллари</i>	<i>д*оьлла</i>
	Родительный	<i>д*оьллачун</i>	<i>д*оьллачѐран, д*оьллачийн</i>	<i>д*оьллачу</i>
	Эргатив	<i>д*оьллачѓ</i>	<i>д*оьллачѐpа, д*оьллачуша</i>	
	Дательный	<i>д*оьллачунна</i>	<i>д*оьллачарна, д*оьллачушина</i>	
	Творительный	<i>д*оьллачуьнца</i>	<i>д*оьллачаьpца, д*оьллачушца</i>	
	Вещественный	<i>д*оьллачух</i>	<i>д*оьллачех</i>	
	Сравнительный	<i>д*оьллачул</i>	<i>д*оьллачел</i>	
	Местный	<i>д*оьллачуьнга</i>	<i>д*оьллачаьpга, д*оьллачушка</i>	
	...			

Масдар (Наст., 16 форм, плюс согласование в классе ⇒ 64 форм)

(19)		Ед. число	Мн. число
	Именительный	<i>д*ōлар</i>	<i>д*ōлари</i>
	Родительный	<i>д*ōларан</i>	<i>д*ōларийн</i>
	Эргатив	<i>д*ōларō</i>	<i>д*ōларша</i>
	Дательный	<i>д*ōларна</i>	<i>д*ōларшна</i>
	Творительный	<i>д*ōларца</i>	<i>д*ōларшца</i>
	Вещественный	<i>д*ōларах</i>	<i>д*ōларех</i>
	Сравнительный	<i>д*ōларал</i>	<i>д*ōларел</i>
	Местный	<i>д*ōларē</i>	<i>д*ōларшка</i>
	...		

В общей сложности это дает 276 форм, причем сюда не входят герундии, и т.д.

Библиография

Beerle, Werner: A contribution to the Morphology of the Simple Verb in Chechen. In: *Thordarson, F. (Hrsg.): Studia Caucasologica I: Proceedings of the Third Caucasian Colloquium, Oslo, July 1986.* Norwegian University Press: Oslo 1988.

Gebhardi, Gunter; Heinecke, Johannes: Lexikonformalismus $\mathcal{L}\mathcal{V}4$. Verbmobil Technisches Dokument 19. Humboldt-Universität zu Berlin 1995 .

Nichols, Johanna: Chechen. In: *Smeets, Riëks (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-ПалгIайн 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 Иностраных 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.