

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

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.

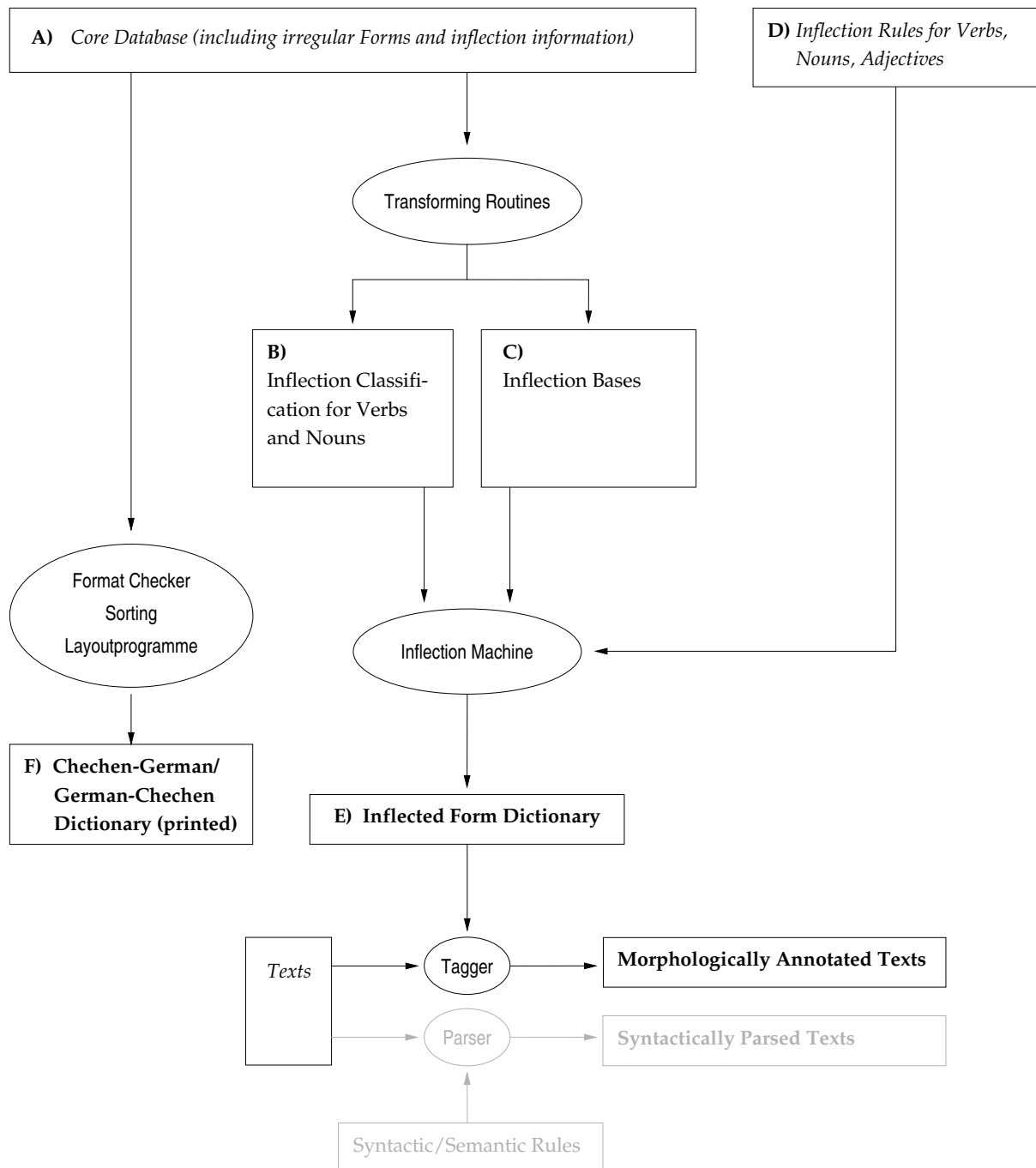
Currently Chechen language resources can usually only be accessed through the medium of Russian. This may lead to problems especially for linguists who only have a restricted knowledge of this language.

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” project¹ 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.

¹A 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)

д|а|д*|б|л|о|/#|d|v|o|l|o|/#|V|e|r|b|#|д|а|д|б|л|а|д*|о|#|д|а|д|б|л|и|й|р|а|#|д|а|д|б|л|и|й|н|а|#|#|h|e|r|a|n|t|r|e|t|e|n|;|b|e|g|i|n|n|e|n|#
д*|о|ц|а|#/#|d|w|o|c|a|/#|A|d|j|#|k|u|r|z|#
д|и|к|а|#/#|d|i|k|a|/#|A|d|j|#|g|u|t|#|И|з|а|(ч|л|o|г|I|a)|д|и|к|а|д|у|-|D|a|s|i|s|t|(s|e|h|r)|g|u|t
н|о|х|ч|и|#/#|п|w|o|χ|f|i|/#|S|u|b|s|t|#|н|о|х|ч|и|ч|у|н|#|н|о|х|ч|и|ч|у|н|н|а|#|н|о|х|ч|и|ч|б|#|н|о|х|ч|и|й|#|1|/|2|#|T|s|c|h|e|t|s|c|h|e|n|I|n|#
ц|I|а|#/#|c'|a|/#|S|u|b|s|t|#|ц|и|й|н|а|н|#|ц|е|н|н|а|#|ц|и|й|н|б|#|ц|е|н|о|ш|#|3|#|H|a|u|s|#
н|о|х|ч|о|#/#|п|w|o|χ|f|w|o|/#>|n|o|k|h|c|h|i|#|T|s|c|h|e|t|s|c|h|e|n|I|n|#
б|б|д|а|#/#|b|v|o|d|a|/#|S|u|b|s|t|#|2|#|б|б|д|а|н|а|#|б|б|д|а|н|н|а|#|б|б|д|а|н|б|#|б|б|д|а|ш|#|5|#|F|i|n|s|t|e|r|n|i|s|,|D|u|n|k|e|l|h|e|i|t|#
с|о|#/#|s|w|o|/#|P|r|o|n|#|с|а|н|#|с|у|н|а|#|а|с|#|с|о|б|ц|а|#|с|о|х|#|с|о|л|#|с|о|б|г|а|#|#|i|c|h|#

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: -с#, NOM PL: -сш#, 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* d*, l
⇒ *Ablaut* Infinitive: б (a, Vowelharmony)
Present: u (u, Vowelharmony)
Witnessed Past: ü ira
Perfective Past: ö + 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) $\mathcal{L}\mathcal{V}4$:

```
base 'д|а|д*о|л|о' :<< verb_c >>:
  deutsch: 'herantreten; anfangen' &           % German translation
  phonol: 'dʋolʋo' &
  konjug: 'kaus-verb' &                       % Inflection set
  praefix: 'д|а|' &
  c1: 'д*о|л|' .                               % Stem to be used for inflection

base 'но|х|чи' :<< subst_c >>:
  deutsch: 'TschetschenIn' &
  phonol: 'nʋoxʃi' &
  flex: 'chunna' &                             % Inflection group
  nom: 'но|х|чи' &                             % Stem for singular paradigm
  nom_pl: 'но|х|чий' &
  klasse: ('1' \ '2')                          % Class (v/db or j/db)
  pl_stamm: 'но|х|чи' .                       % Stem for plural paradigm

base 'ц|а|' :<< subst_c >>:
  deutsch: 'Haus' &
  phonol: 'ʃa' &
  flex: 'vsh' &                                 % Inflection group
  nom: 'ц|а|' &
  gen: 'ц|и|й|н|а|н' &                         % Ablaut phenomena must be coded
  dat: 'ц|е|н|н|а|' &
  erg: 'ц|и|й|н|о|' &
  nom_pl: 'ц|е|н|о|ш|' &
  klasse: ('3') &                               % Class (d/d)
  sg_stamm: 'ц|е|н|' &
  pl_stamm: 'ц|е|н|' .

base 'д*о|ц|а|' :<< adj_c >>:
  deutsch: 'kurz' &
  phonol: 'd*ʋoca' &
  nom: 'до|ц|а|' .
```

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 affiged to the root:

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

e.g.:

$$(7) \text{ д}^* \text{оца (‘‘short’’)} \Rightarrow \text{боца, воца, доца, йоца}$$

$$(8) \text{ д}^* \text{аьсса (‘‘empty, free’’)} \Rightarrow \text{баьсса, ваьсса, даьсса, яьсса}$$

Secondly, the case suffixes are added onto the stem:

(9)	independent		dependent
	SG:	PL:	SG/PL:
NOM:	-ниг	-наш	
GEN:	-чун	-чѣран, -чийн	-чу
ERG:	-чѣ	-чѣра, -чуша	
DAT:	-чунна	-чарна, -чушка	
INSTR:	-чуьнца	-чаьрца, -чушца	
MAT:	-чух	-чѣрах, -чех	
КОМП:	-чул	-чѣрал, -чел	
ALL:	-чуьнга	-чаьрга, -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 + -чуьнца
		ALL-SG: SG-stem + -чуьнга
		GEN-PL: PL-stem + -ойн
		DAT-PL: PL-stem + -ошна
		ERG-PL: PL-stem + -оша
		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*оход*ан “to warm up”), the inflection is carried out before replacing ‘д*’, e.g.:

(11)	Type	regular
	⇒ Present	consonants + present-vowels
	⇒ Witnessed Past	consonants + with.-past-schema
	⇒ Perfective Past	consonants + perf.-past-schema
	⇒ Paradigm	Close Past
		Witnessed Past – -pa
		Remote Past
		Perfective Past + -ē-
		Repeated Past
		Present + -pa
		...
		Present Participle
		Present + case-suffixes
		Perfect Participle
		Witnessed Past + case-suffixes
		Masdar
		Present + case-suffixes

(12)	Type	kaus-verb
	⇒ Paradigm	Present
		stem + ад*о
		Close Past
		stem + ий
		Perfective Past
		Close past + на
		Witnessed Past
		Close past + pa
		Remote Past
		Close past + нēpa
		Future
		stem + ад*ep
		...
		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
дѡабѡлабо # VERB praesens b VAR # дѡад*ѡлѡ # anfangen
дѡабѡлѡ # VERB infinitiv b VAR # дѡад*ѡлѡ # anfangen
дѡабѡлѣ # VERB imperativ b VAR # дѡад*ѡлѡ # anfangen
...

дѡайѡлайо # VERB praesens j VAR # дѡад*ѡлѡ # anfangen
дѡайѡлѡ # VERB infinitiv j VAR # дѡад*ѡлѡ # anfangen
...

(14) Inflected Form # Wordclass Case Number Class # Baseform # German
цѡенах # SUBST materialis sg d # цѡа # Haus
цѡенал # SUBST komparativ sg d # цѡа # Haus
цѡенца # SUBST instrumental sg d # цѡа # Haus
цѡѣнийн # SUBST genitiv pl d # цѡа # Haus
...

(15) Inflected Form # Wordclass Case Number Class State # Baseform # German
боѡачѣран # ADJ genitiv pl b unabh # д*ѡца # kurz
боѡачѡ # ADJ ergativ sg b unabh # д*ѡца # kurz
боѡачаѣрца # ADJ instrumental pl b unabh # д*ѡца # kurz
боѡачаѣрга # ADJ lokativ pl b unabh # д*ѡца # kurz
боѡачарна # ADJ dativ pl b unabh # д*ѡца # kurz
боѡачех # ADJ materialis pl b unabh # д*ѡца # kurz
боѡачел # ADJ komparativ pl b unabh # д*ѡца # kurz
боѡачийн # ADJ genitiv pl b unabh # д*ѡца # kurz
боѡачу # ADJ ^nominativ pl b abh # д*ѡца # kurz
боѡачу # ADJ ^nominativ sg b abh # д*ѡца # kurz
боѡачух # ADJ materialis sg b unabh # д*ѡца # kurz
боѡачул # ADJ komparativ sg b unabh # д*ѡца # kurz
...

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

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

(16)	Imperative	д*ола
	Present	д*улу
	Future	д*улур
	Close Past	д*уьйли
	Perfective Past	д*оьлла
	Witnessed Past	д*уьйлира
	Repeated Past	д*улура
	...	

Participles (Present, 23 Forms, with class agreement \Rightarrow 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*улучо	д*улучёра, д*улучуша	
	Dative	д*улучунна	д*улучарна, д*улучушна	
	Instrumental	д*улучуьнца	д*улучаьрца, д*улучушца	
	Materialis	д*улучух	д*улучех	
	Comparative	д*улучул	д*улучел	
	Allative	д*улучуьнга	д*улучаьрга, д*улучушка	
	...			

(Past 23 Forms, with class agreement \Rightarrow 92 Forms):

(18)		Independent		Dependent
		Singular	Plural	Singular/Plural
	Nominative	д*оьлларг	д*оьлларш	д*оьлла
	Genitive	д*оьллачун	д*оьллачёран, д*оьллачийн	д*оьллачу
	Ergative	д*оьллачо	д*оьллачёра, д*оьллачуша	
	Dative	д*оьллачунна	д*оьллачарна, д*оьллачушна	
	Instrumental	д*оьллачуьнца	д*оьллачаьрца, д*оьллачушца	
	Materialis	д*оьллачух	д*оьллачех	
	Comparative	д*оьллачул	д*оьллачел	
	Allative	д*оьллачуьнга	д*оьллачаьрга, д*оьллачушка	
	...			

Masdar (16 Forms, with class agreement \Rightarrow 64 Forms):

(19)		Singular	Plural
	Nominative	д*ӕлар	д*ӕларш
	Genitive	д*ӕларан	д*ӕларийн
	Ergative	д*ӕларӕ	д*ӕларша
	Dative	д*ӕларна	д*ӕларшна
	Instrumental	д*ӕларца	д*ӕларшца
	Materialis	д*ӕларах	д*ӕларех
	Comparative	д*ӕларал	д*ӕларел
	Allative	д*ӕларӕ	д*ӕларшка
	...		

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 Caucasologica I: Proceedings of the Third Caucasian Colloquium, Oslo, July 1986.* Norwegian University Press: Oslo 1988.
- Gebhardi, Gunter; Heinecke, Johannes:** Lexikonformalismus &U4. Verbmobil Technisches Dokument 19. Humboldt-Universität zu Berlin 1995 .
- Nichols, Johanna:** Chechen. In: *Smeets, Riëks (Hrsg.): The Indigenous 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-Галгӕайн 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.