An Attribute-Sample Database System for Describing Chuvash Affixes

Pavel V. Zheltov

Chuvash State University,
Russia, 428025, Gouzovskogo 21/37,
tchouvachie@narod.ru

Abstract

In the paper is described **"**KÜLEPEK" **–** a database system created by the author for description of Chuvash word-changing and word-forming affixes. The system is based on the attribute-sample model of affixes, which allows to describe affixes and their phonological, morphotactic and orthography rules. The system uses DBase IV database engine and was created in Borland Delphi 7.0 software environment, has a user friendly interface and can run under Windows 98/200/ME/XP. It is applicable for a large number of agglutinative languages with a finite-state morphology and can be used as a part of a morphological parser, and as an independent reference tool as well.

1. The Attribute-Sample Database System's Common Structure and Description

The attribute-sample model of morphology (Zheltov, 2003) is based on following principles:

1) division of affixes to types;
2) conforming to each type a number of patterns, which describe phonological, morphotactic and spelling rules.

Chuvash language[11], being an agglutinative one, is mainly based on affixes and their interaction with stems and each other. In Chuvash morphemics are widely present such phonological variations (Sergeyev, 1992) as:

1) quality synharmonism - "soft" (front vowel) stems agglutinate "soft" allomorphs, while "hard" (back vowel) stems agglutinate "hard" allomorphs. As a rule each affix has minimum two allomorphs – a "soft" one and a "hard" one:

anne "my mother" + e (dat.-gen. affix) = annen**e** "to my mother";

laşa "horse" + a (dat.gen. affix's allomorph) = laşan**a** "to horse".

But some affixes have only "soft" allomorphs (like 3-rd pers. sing. possessive affix -ĕ/-i "his/her"), thus a distorting of. synharmonism is observed sometimes:

šırăvĕ "his/her letter".

2) interphonemes insertion:

anne "my mother" +e (dat.-gen affix) = anne**n**e "to my mother/my mother".

3) vowels reduction (elisia)

a) in word formation:

vat šın < vată šın – "old man",

purnăš < purănăš – "life", from purăn – "to live"

b) in word changing:

vula ("to read") + ăp (future tense 1-st pers. sing. affix) = vulăp " (I) shall read".

4) consonants reduction:

in ten verb stems, ending on -r, *r* is falling out in some verb forms:

pır ("to come") + -t- (past tense affix) + -ăm (1-st pers. sing. affix) = pıtăm "(I) have come".

5) consonants duplication in noun stems, ending on ă/ĕ, combined with final vowel reduction, when placed to dative-genitive case:

tulă "wheat" + -a (dat.-gen. case affix) = tulla.

sĕlĕ "oats" + -e (dat.-gen. case affix) = sĕlle.

6) final vowels alternation:

u – ăv,

ü – ĕv.

šır**u** "letter" + ĕ (3-rd pers. sing. possessive affix) = šır**ăv**ĕ "his/her letter",

věren**ü** "studying" + ĕ (3-rd pers. sing. possessive affix) = věren**ĕv**ĕ "his/her studying".

The ***n*** in the example above as well as in ***laşana*** is an interphoneme, placed when a stem ends on a vowel and the agglutinating to it affix also begins on a vowel. Interphonemes are also being used in Chuvash in some other cases.

From this point of view Tatar language (a neighbour Turkic language of Kipchak group) has also two allomorphs of dative case: -a, -ə, while its others allomorphs -ga, -gə, -ka, -kə, -na, -nə are compound ones, decomposed to the allomorphs -a, -ə and interphonemes -g-, -k-, -n-. But from the formal point of view the representation accepted by Chuvash linguists is more comfortable, especially for computer analysis.

The database system has an interface structure, consisting of 6 tables that can be optionally filled for each affix:

Affix	Allomorphs	Morphologic feature	Type
A	a	Case (dative-genitive)	1
	e		

Table 1: Affixes.

[1] Chuvash language belongs to the Bulgar group of Turkic languages, together with extinguished Bulgar and Hazar and counts near 1,6 millions of speakers, their main part lives in Chuvash Republic and Volga region of Russian Federation. It is considered an endangered one and you can read more about the situation in (Zheltov, 2005).

Type	Stem – Affix	Stem˜ – Affix	Affix˜– Affix	Exception
1	1	2	3	4

Table 2: Type – Application rules.

Left context	Allomorph	Transformation	Result	Example
ВС	а	+а	ВСа	курак-а
FC	е	+е	Все	кӗрӗк-е
Са	а	+н+а	Сана	лаша-на
Се	е	+н+е	Сене	пике-не
Си	е	+е	Сие	пӑри-е

Table 3: Stem–Affix.

Left context	Allomorph	Transformation	Result	Example	Exception
Сӳ	е	-ӳ ,+ӗв,+е	СӗВе	пӗлӳ–пӗлӗве	–
Су	а	-у,+ӑв,+а	СӑВа	çыру–çырӑва	–
ВСӑ	а	-ӑ,+с,+а	ССа	пулӑ–пулла	–
FСӗ	е	-ӗ,+с,+е	Ссе	сӗлӗ – сӗлле	–
ССӑ	а	-ӑ,+а	Сса	карланкӑ – карланка	пуртӑ

Table 4: Stem˜– Affix.

Left context	Allomorph	Transformation		Result	Example
		Left context	Transformation		
ӳ	е	F	+н+е	ӳне	аннӳ–аннӳне
		FC	-ӳ, +н+е	Сне	киӳ– кинне
		ВСь	-ӳ,+ь,+н+е	Сьне	мӑкӑнӳ– мӑкӑньне
у	а	В	+н+а	уна	хулу– хулуна
		ВС	-у,+н+а	Сна	арӑму–арӑмна

Table 5: Affix˜ – Affix.

Exception	Allomorph	Transformation	Result
пуртӑ	а	-ӑ, +с, +а	пуртта

Table 6: Exceptions.

The table "Affixes" contains the affixes (their lexical representation)[2], their allomorphs (surface representation of affixes), their morphological feature (in Chuvash there is usually 1 : 1 relation between an affix and morphological features it expresses) and the optional field "Type". The field "Type" is used when in our database there exist already patterns of rules corresponding to the newly inputted affix, so we can just fill the type and the tables below will be filled automatically from existing patterns, with which the type value is related in the table "Type – Application rules".

The table "Type – Application rules" is filled automatically by the system for each affix, after the user has filled phonological rules in the tables "Stem – Affix", "Stem˜ – Affix", "Affix˜ – Affix", "Exception".

The table "Stem – Affix", describes contexts in which the current affix, when being glued to a stem, doesn't cause any changes in the last one. The field "Left context" contains the finals of the stem, to which the current allomorph can be glued. In the field "Allomorph" the user enters an allomorph of the current affix, which can be glued in this context. The field "Transformation" contains the transformations, which have to be done to glue the allomorph. The sign "-" means reduction of a symbol before it, while "+" means addition. The Latin capital letter "B" means back vowel, "F" means front vowel and "C" consonant. In the field "Exception" are listed stems, the interaction of which with the current allomorph is exception to these rules. The field "Example" illustrates the application of each phonological rule.

The table "Stem˜ – Affix" describes contexts in which the current affix causes the change of a stem it is glued to.

"Affix˜ – Affix" is a table describing contexts when an interaction between the current affix and the affix that has been glued to the stem before it causes phonological changes in the last one.

While describing morphonological rules for Chuvash affixes, we have also encountered the recursion phenomenon.

The phenomenon of recursion on the morphology level is present both in Chuvash and Tatar, as well as in other Turkick languages. They are formed in Chuvash by:

1) relative affixes -ti/-çi.

Such recursive structures are translated into English with the means of relative pronouns – "who", "which", "what" and with the means of demonstrative pronouns "that", "those", "these".

a) Yal+ti – "that, who/which is in the village".
Village+ti.

b) Yaltisene – "to those who/which are in the village"
Village+ti+plural affix+dat-dir.

c) Yaltisençine –"to that, which/who is by those, who are in the village".
Village+ti+plural+çi +dat-dir.

d) Yaltisençisene – "to those, who/which are by those, who/which are in the village".
Village+ti+plural+çi +plural+dat-dir.

2) by the possesivity affix -*Ăn* (-*ăn*/-*ĕn*/-*n*). This affix is closely related with the possesivity's category.

Let us have a hierarchy of possesivity's relations: A⊂B⊂C… Then a following structure is possible: A-*Ăn*⊂B-*Ăn*⊂C-*Ăn*…:

Tăvan appăşĕn ıvălĕn açin mănukĕ – "the grandson of his/her elder sister's son's son (the grandson of cousin's son)";

3) by the affix of dative-accusative case -*A* (-a/-e). Let us have an hierarchy of spatial relations. Then is possible a recursive structure A-*NA*⊂B-*NA*⊂C-*NA*…: ***Aytar Parişa universiteta texnika fakultetne vĕrenme kayrĕ***. – "Aydar has gone to study **to** Paris, **to** university, **to** the technical department".

4) by the locative case affix -*TA* (-*ta*/-*te*/-*ra*/-*re*/-çe).

Aytar Parişra universitetra texnika facultetĕnçe vĕrenet. – "Aydar studies **in** Paris, **at** university, **at** the nical department".

5) by ablative case affix -*TAn* (-*tan*/-*ten*/-*ran*/-*ren*/-*çen*). ***Francinçen Parişran universitetran Aytartan şıru kilçĕ.*** – "A letter has come **from** France, **from** Paris, **from** university, **from** Aydar".

As it can be seen from the examples above, the recursion phenomenon is an important one while parsing. In our database we describe it by adding the word "recursion" into the morphologic feature of these affixes.

Affix	Allomorphs	Morphologic feature	Type
Ти (Ti)	ти (ti) чи (çi)	Relative affix (recursion)	3
Ăн (Ăn)	ăн (ăn) ĕн (ĕn)	Possesivity affix (recursion)	4
А	а е	Dative-accusative case affix (recursion)	4
ТА	та (ta) те (te) ра (ra) ре (re) че (çe)	Locative affix (recursion)	4
ТАн (ТАn)	тан (tan) тен (ten) ран (ran) рен (ren) чен (çen)	Ablative case affix (recursion)	4

Table 7: Affixes.

Using our system we have created a database, which describes over 120 word-forming and over 50 word-changing Chuvash affixes, and are planning to create on its basis a morphological parser of Chuvash language. We have also compiled on its basis a reference tool (Table 8).

[2] The lexical representation is an abstract uniting entry for a group of allomorphs, expressing the same morphologic features. For example we can lexically represent the allomorphs' set {-a,-e} of dative-genitive case like -A. The concrete allomorphs {-a,-e} are surface representations of the abstract affix –A.

Lexical representation	Allomorph	Context		
		Vowel quality	Vowel / Consonant	Rejection /Addition
A	е (кĕнеке+н+е)	F	е	The interphoneme ‘н’ is glued after the stem.
	е (пĕрч+е, сĕлл+е)	F	ĕ	Ĕ is rejected. If the word ends on. Ĕ preceded by a consonant this consonant is duplicated in two syllable words.
	а (карланк+а, пулл+а, пуртт+а)	B	ă	ă falls out. If the word ends on ă preceded by a consonant this consonant is duplicated in two syllable words. The exception is the word пуртă.
	а (ача+н+а)	B	а	‘н’ is glued after the stem. The exceptions are loanwords from Russian ending on ‘а’. In them ‘а’ is rejected and changed on ‘ă’ (машина – машинăна).
	е (пĕлĕв+е, анне – аннӳ – аннӳ+н+е, ĕне – ĕнӳ – ĕнӳ+н+е)	F	ӳ	‘ӳ’ is rejected. Exceptions are words, where ‘ӳ’ is the possessive affix of the 2-nd person singular. In them ‘ӳ’ is not rejected and the interphoneme ‘н’ is glued after the stem, before the allomorph ‘е’.
	а (ывăл – ывăлу – ывăл-н-а) the possessive affix of the 2-nd person singular	B	the stem ends on a consonant	‘у’ falls out: ‘н’ is added after the stem.
	е (кин – кинӳ – кин+н+е) the possessive affix of the 2-nd person singular	F	the stem ends on a consonant	‘ӳ’ falls out: ‘н’ is added after the stem.
	а (кино+н+а)	B	о	‘н’ is added after the stem
	е (медаль – медал-е, мăкăнь– мăкăн-е, тетрадь – тетрад+е)	B	дь, ль. нь	‘ь’ falls out (медаль - медале)
	е (мăнукĕ – мăнук+н+е) the possessive affix of the 3-rd person singular	B	ĕ	‘ĕ’/’и’ fall out, ‘н’ is added after the stem.
	е (çулçи – çулçи+н+е) the possessive affix of the 3-rd person singular	B	и	‘н’ is added after the stem.
	е (тетрачĕ – тетрадь+н+е, турачĕ – турат+н+е) the possessive affix of the 3-rd person singular	B	ĕ	If the original stem ends on ‘т’/‘ть’/ ‘д’/‘дь’, then ‘ч’ which appeared due to the possessive affix of 3-rd person singular is rejected. The stem is transacted back and the interphoneme ‘н’ appears.

Table 8: The table of a reference tool for Chuvash affixes.

2. References

Zheltov, P.V. (2003). Attribute-Sample Model of Lexics in a Comparative-Correlative Aspect. *Newsletter of Chuvash State University, Natural and Technical science,* № 2. Cheboksary: Chuvash State University Press, pp.131-136.

Zheltov, P.V. (2005). Minority Languages and Computerization. Situation in Russian Federation. *Ogmios Newsletter*,3.03 (#27), pp. 8-11.

Sergeyev, V.I. (1992). *Modern Chuvash Language. A Handbook of Morphology*. Cheboksary: Chuvash State University Press.