Caucasian Languages

[B.] George Hewitt FBA, Professor of Caucasian Languages, NME Dept., SOAS, Thornhaugh Street, Russell Square, London WC1H 0XG, UK.

Around 38 languages are deemed to be indigenous to the Caucasus; often difficult demarcation between language and dialect explains the uncertainty. The ancestral homelands are currently divided between:

Russia's north Caucasian provinces (Circassian, Abaza, Ingush, Chechen, Avaro-Ando-Tsezic, Lako-Dargic, northern Lezgic)

de facto independent Abkhazia (Abkhaz, Mingrelian, Svan, Georgian, Laz)

Georgia (Georgian, Mingrelian, Svan, Laz, Bats, Chechen, Avar, Udi)

Azerbaijan (Lezgi, Budukh, Kryts’, Khinalugh, Rutul, Ts’akhur, Avar, Udi)

Turkey (Laz, Georgian)

Diaspora-communities of north (especially north-west) Caucasians can be found across former Ottoman territories, particularly Turkey, where the majority Circassian and Abkhazian populations reside and where the term 'Cherkess' often indiscriminately applies to any north Caucasian. Circassians are found in Syria, Israel and Jordan, home also to a significant Chechen population. Speaker-numbers range from 500 (Hinukh) to 3-4 million (Georgian). Many of the languages are endangered.

Three families are usually recognised:

A. South Caucasian (Kartvelian)

Georgian

Svan

Mingrelian (Megrelian)

Laz (Ch’an)

[Scholars in Georgia regard Mingrelian and Laz as co-dialects of Zan]

B. North West Caucasian

Abkhaz

Abaza

Ubykh (extinct from 1992)
West Circassian (Adyghe)

East Circassian (Kabardian)

C. Nakh-Daghestanian

(a) Nakh (North Central Caucasian)

Chechen

Ingush

Bats (Ts’ova Tush)

(b) Daghestanian (North East Caucasian)

1. Avaro-Ando-Tsezic(/Didoic):

   Avaric: Avar

   Andic: Andi, Botlikh, Godoberi, K’arat’a, Akhvakh, Bagvalal, T’indi, Ch’amaral

   Tsezic: Tsez (Dido), Khvarshi, Hinukh, Bezht’a (K’ap’uch’a), Hunzib (these last
   two are sometimes regarded as co-dialects)

2. Lako-Dargic:

   Lakic: Lak

   Dargic: Dargwa (Dargi(n)) — some treat K’ubachi, Chiragh, and Megeb as full
   languages

3. Lezgic:

   Lezgi(an), Tabasaran, Rutul (Mukhad), Ts’akhur, Aghul, Udi, Archi, Budukh,
   Khinalugh, Kryts’

Some challenge the Lezgic status of Archi, Khinalugh, Budukh and Kryts’. Mutual
intelligence basically exists between Laz and Mingrelian, Abkhaz and Abaza, West
and East Circassian. Only Georgian has an ancient tradition of writing, but during the
Soviet period the languages in bold all enjoyed literary status. Publishing in
Mingrelian, Laz, Ts’akhur, Aghul, Rutul and Udi was tried in the 1930s but
discontinued, though there have been some post-Soviet attempts to publish more
widely (including Dido).

Phonetics and Phonology
All Caucasian languages have voiced vs voiceless aspirate vs voiceless ejective plosives, affricates and occasionally fricatives, to which some add a *fortis* series (voiceless unaspirated or geminate). N.W. Caucasian is characterised by large consonantal inventories coupled with minimal vowel-systems, consisting of at least the vertical opposition open /a/ vs close /ə/. Ubykh possessed 80 phonemes (83 if the plain velar plosives attested only in loans are admitted), with every point of articulation between lips and larynx utilised and displaying the secondary features of palatalisation, labialisation, and pharyngalisation — Daghestanian pharyngalisation is normally assigned to vowels:

**Consonantal Phonemes for Ubykh**

\[
\begin{array}{cccccc}
p & b & p' & m & w \\
p^c & b^c & p^c' & m^c & w^c \\
\end{array}
\]

\[
\begin{array}{cccccccccc}
f & v^c \\
\end{array}
\]

\[
\begin{array}{cccccc}
t & d & t' & n & r \\
t^w & d^w & t^w' \\
\end{array}
\]

\[
\begin{array}{cccccccccc}
f_ç & d_ç & f_ç' & s & z \\
f_ç & d_ç & f_ç' & s & z \\
f_ç^w & d_ç^w & f_ç^w & s^w & z^w \\
f_j & d_j & f_j' & j & ʒ \\
f_j^w & ʒ^w \\
f_ç & d_ç & f_ç' & ʒ & ʒ \\
\end{array}
\]

\[
\begin{array}{cccccccccc}
\iota & \iota' & \iota^w & \iota^w' \\
\end{array}
\]

\[
\begin{array}{cccccccccc}
(k) & (g) & (k') & \times & \gamma \\
(\mathit{k}) & \mathit{g} & (\mathit{k})' \\
(\mathit{k})^w & \mathit{g}^w & (\mathit{k})^w' \\
q & q' & \chi & e \\
q^c & q^c' & \chi^c & e^c \\
\end{array}
\]
Some recent analyses of Daghestanian languages have produced inventories rivalling those of the N.W. Caucasian, though no parallel minimality amongst the vowels is posited. One analysis of Archi assigns it 70 consonants:

<table>
<thead>
<tr>
<th>Consonantal System of Archi</th>
</tr>
</thead>
<tbody>
<tr>
<td>p  b  p'  ps             m  w</td>
</tr>
<tr>
<td>t  d  t'  ts             n  r</td>
</tr>
<tr>
<td>t'w  d'w</td>
</tr>
<tr>
<td>f's  f's'  ts'          s  z</td>
</tr>
<tr>
<td>ts'w  ts'w              s'w  z'w</td>
</tr>
<tr>
<td>tʃ  tʃ'  tʃː'  fʃː'  fʃː  j  jː  ʒ</td>
</tr>
<tr>
<td>fʃːw  fʃːw              j'w  j'w  z'w</td>
</tr>
<tr>
<td>lʃ  lʃː  lʃː  lʃːw  lʃːw  lʃːw</td>
</tr>
<tr>
<td>k  g  k'  kʃ            l  lʃ</td>
</tr>
<tr>
<td>k'w  g'w  k'w'  k'ʃw  l'w  l'ʃw</td>
</tr>
<tr>
<td>q  q'  qː  qa           x  xa</td>
</tr>
<tr>
<td>q'w  q'w'  qa'w  qa'w  x'w  x'w</td>
</tr>
<tr>
<td>ʔ  h  ʡ</td>
</tr>
</tbody>
</table>

Noticeable here is the presence of 10 laterals, though some specialists recognise no more than 3/4.

Kartvelian occupies a mid-position with between 28 and 30 consonants (see *Georgian*). Georgian shares with Avar and Andi the simple 5-vowel triangle:

**Georgian-Avar-Andi Vowel-system**
Schwa is added to this in the other Kartvelian languages, whilst the various Svan dialects have length and/or umlaut, Upper Bal having the richest system:

**Svan's Upper Bal Vowel-system**

\[
i \quad \gamma \quad \gamma \quad u \quad \upsilon
\]
\[
\alpha
\]
\[
\epsilon \quad \epsilon \quad \epsilon \quad \epsilon \quad \epsilon
\]
\[
\eta \quad \eta \quad \eta \quad \eta \quad \eta
\]

Triangular or quadrilateral vowel-systems are attested in Nakh-Daghestanian:

**Bezht’a Basic Vowel-system**

\[
i \quad \gamma \quad u
\]
\[
\epsilon \quad \epsilon
\]
\[
\alpha
\]

[All but /γ, ə, ø/ possess long counterparts, and the nasalised vowels: /ɨ, ɨ, ɨ, ɨ, ɨ. ɨ, ɨ/ have also been recognised]

**Hunzib Basic Vowel-system**

\[
i \quad i \quad u
\]
\[
\epsilon \quad \epsilon
\]
\[
\alpha
\]

[All these Hunzib vowels have long counterparts, and fluctuating nasalisation on short vowels has been observed]

The simplest (near-)quadrilateral system is attested in Chiragh Dargwa, with four pairs distinguished by length:

**Chiragh Dargwa Vowel-system**

\[
i(\hat{\imath}) \quad u(\hat{\imath})
\]
\[
\epsilon(\hat{\imath})
\]
\[
\alpha(\hat{\imath})
\]

Udi has been analysed with:
Udi Vowel-system

\[ \begin{array}{ccc}
    \text{\textit{i}} & \text{\textit{u}} & \text{\textit{o}} \\
    \text{\textit{a}} & \text{\textit{o}} & \text{\textit{a}} \\
\end{array} \]

whilst Chechen presents the complicated system:

Chechen Vowel-system

\[ \begin{array}{ccc}
    \text{\textit{i}} & \text{\textit{u}} & \text{\textit{o}} \\
    \text{\textit{a}} & \text{\textit{o}} & \text{\textit{a}} \\
\end{array} \]

[Most, if not all, of these can be nasalised as a result of the weakening of a following \textit{/n/}]

Stress is sometimes distinctive (Abkhaz-Abaza) but usually not. Tonal distinctions have been proposed for some of the Daghestanian languages (Andi, Akhvakh, Ch’amalal, Khvarshi, Hinukh, Bezht’a, Tabasaran, Ts’akhur, Ingush and Budukh).

Morphology

N.W. Caucasian sememes are typically C(C)(V), and minimal case-systems combine with highly polysynthetic verbs, which may contain: up to four agreement-prefixes, locational preverbs, orientational preverbs and/or suffixes, interrogative and conjunctural elements, and markers of tense-modality, (non-)finiteness, causation, potentiality, involuntariness, polarity, reflexivity, and reciprocality (see Abkhaz).

Kartvelian balances a moderate total of cases with reasonably complex verbs, which may contain: agreement with two or three (rarely four) arguments via two sets of agreement-affixes, directional/perfectivising preverbs (the large total in Mingrelian-Laz suggests N.W. Caucasian influence), and markers of tense-aspect-modality, causation, potentiality, version (vocalic prefixes indicating certain relations between arguments), and voice — Kartvelian is the only family to have a full active-passive diathetic opposition. Nakh-Daghestanian has complex nominal systems with both grammatical and sometimes large numbers of locative cases; Lezgi(an), Aghul and Udi apart, nouns fall into one of between two and (depending on the analysis) five or eight (largely covert) classes. Verbs are correspondingly simple: agreement is totally absent from Lezgi(an) and Aghul; elsewhere, verbs with an agreement-slot typically
allow only class-agreement (Andic), though some languages (Bats, Lak-Dargwa, Tabasaran, Akhvakh, Arichi, Hunzib, Avar dialects) have added perhaps rudimentary person-agreement, whilst Udi has person-agreement only. Some languages have a small selection of preverbs. Some distinguish perfective from imperfective roots. Some North Caucasian verbs can be construed transitively or intransitively (?passively), depending on the clausal structure; anti-passives are also attested.

Avar illustrates a typical system of locative-cases:

**Avar Locative Case-endings**

<table>
<thead>
<tr>
<th>Series</th>
<th>Essive</th>
<th>Allative</th>
<th>Ablative</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 'on'</td>
<td>-d((.)a</td>
<td>-d.ø</td>
<td>-d((.)a.øa</td>
</tr>
<tr>
<td>2. 'near'</td>
<td>-f</td>
<td>-f.ø</td>
<td>-f.øa</td>
</tr>
<tr>
<td>3. 'under'</td>
<td>-f(i)(i)</td>
<td>-f(i)(i).ø</td>
<td>-f(i)(i).øa</td>
</tr>
<tr>
<td>4. 'in (mass)'</td>
<td>-f(i)</td>
<td>-f(i).ø</td>
<td>-f(i).øa</td>
</tr>
<tr>
<td>5. 'in (space)'</td>
<td>-D (= class-marker)</td>
<td>-D.ø</td>
<td>-øa</td>
</tr>
</tbody>
</table>

Ergativity and some other oblique case-function are often merged in a single morph.

Deictic systems range from two-term (Mingrelian, Ubykh, Kryts’), through three-term (Georgian, Abkhaz, Circassian), to five-term in a swathe of Daghestanian, and even six-term (Lezgi(an), Godoberi).

Counting-systems are predominantly vigesimal, at least up to '99' (though Bats is vigesimal throughout), but some systems are decimal.

**Syntax**

Word-orders are: Kartvelian and Nakh-Daghestanian AN, GN, N-Postposition, SOV, though Old Georgian was rather NA and NG; N.W. Caucasian GN, predominantly NA, N-Postposition, SOV. Some degree of ergativity characterises all the languages but in Mingrelian, where the system was orginally as illustrated for Georgian (q.v.), the Ergative case-marker was extended vertically to replace the original Nominative for intransitive (including Indirect) verbs in Series II (Aorist Indicative and Subjunctive), where it functions as a Series II Nominative allomorph, the original
Nominative effectively becoming an Accusative just for Series II — Laz has extended the case-marker horizontally across its three Series for all transitive subjects. Active-Inactive alignment plays a role in some languages (Bats).

A Nominative/Absolutive argument is the obligatory minimum in a clause, and where verbs have class-agreement, this is the determiner for the class-marker (which in some languages also appears on adverbs and as part of a locative case-exponent); the determiner for person-agreement in languages with class-agreement might be this same or a different argument (e.g. the logical subject), depending on a variety of factors.

Verbs such as 'want', 'have', 'hear' are construed indirectly with the logical subject in an oblique case, but, if Kartvelian and N.W. Caucasian employ just the Dative/general Oblique case for this argument, greater distinctions can apply in Nakh-Daghestanian: Avar employs its Dative case with verbs of emotion ('love'), a locative (Series I Essive) with verbs of perception ('see'), and the Genitive for the possessor in conjunction with the copula.

Only Kartvelian has the category of subordinating conjunctions, naturally associated with full clauses containing indicative or subjunctive finite verbs. Such structures are rare in North Caucasian, where one finds a variety of non-finite (nominalised) verb-forms fulfilling the subordinate role.

Examples:

\texttt{mother-Erg meat.Absol3 3-fry-Past}

'Mother fried the meat' vs 'The meat (was) fried' [Andi]

\texttt{brother-Erg water.Absol boil-Pres}

'Brother is boiling the water' [Bezht’a]

\texttt{is-ti si valueOf-ti}

'Brother.Absol water-Instr boil-AntiPass-Pres'

'Brother is regularly engaged in boiling water' [Bezht’a]
The man saw the girl' vs 'The girl died'  [Mingrelian]

'The man could see the girl'  [Mingrelian]

'Father loves (his) daughter'  [Avar]

'Father sees (his) sons'  [Avar]

'Father has a horse'  [Avar]

'The north wind admitted that the sun was apparently the stronger of them'  [Lower Bal Svan]
'It became impossible for the north wind not to admit how/that the sun is stronger
than it' [Temirgoi West Circassian]

Kinship

Kartvelian is unrelated to any known language(-family), but the debate continues
concerning the relationship between the northern families. Linkage to Hattic is
postulated for N.W. Caucasian and to Hurrian for Nakh-Daghestanian. Udi has
recently been conclusively demonstrated to descend from Caucasian Albanian.

Bibliography

introduction to Dargi grammar*. Leiden: Research School CNWS.
of *Lingua*, 147-190.
Press.
Phonological System and Parts of its Lexicon and Morphology*. Leiden: CNWS.
Markert und Petters.


<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absol</td>
<td>Absolutive</td>
</tr>
<tr>
<td>Acc</td>
<td>Accusative</td>
</tr>
<tr>
<td>AntiPass</td>
<td>Anti-passive</td>
</tr>
<tr>
<td>Aor</td>
<td>Aorist</td>
</tr>
<tr>
<td>Caus</td>
<td>Causative</td>
</tr>
<tr>
<td>Comp</td>
<td>Comparative</td>
</tr>
<tr>
<td>Dat</td>
<td>Dative</td>
</tr>
<tr>
<td>Dyn</td>
<td>Dynamic</td>
</tr>
<tr>
<td>Erg</td>
<td>Ergative</td>
</tr>
<tr>
<td>Fin</td>
<td>Finite</td>
</tr>
<tr>
<td>Fut</td>
<td>Future</td>
</tr>
<tr>
<td>Gen</td>
<td>Genitive</td>
</tr>
<tr>
<td>Instr</td>
<td>Instrumental</td>
</tr>
<tr>
<td>Loc</td>
<td>Locative</td>
</tr>
<tr>
<td>N/F</td>
<td>Non-finite</td>
</tr>
<tr>
<td>Nom</td>
<td>Nominative</td>
</tr>
<tr>
<td>Obl</td>
<td>Oblique</td>
</tr>
<tr>
<td>P/I</td>
<td>Past Indefinite</td>
</tr>
<tr>
<td>Perf</td>
<td>Perfect</td>
</tr>
<tr>
<td>Pl</td>
<td>Plural</td>
</tr>
<tr>
<td>Pot</td>
<td>Potential</td>
</tr>
<tr>
<td>Pref</td>
<td>Prefix</td>
</tr>
<tr>
<td>Pres</td>
<td>Present</td>
</tr>
<tr>
<td>Prev</td>
<td>Preverb</td>
</tr>
<tr>
<td>Res</td>
<td>Resultative</td>
</tr>
<tr>
<td>Stat</td>
<td>Stative</td>
</tr>
<tr>
<td>Suff</td>
<td>Suffix</td>
</tr>
<tr>
<td>SV</td>
<td>Subjective Version</td>
</tr>
<tr>
<td>TS</td>
<td>Thematic Suffix</td>
</tr>
<tr>
<td>TV</td>
<td>Thematic Vowel</td>
</tr>
</tbody>
</table>