Khevsur and Tush and the status of unusual phenomena in corpora
Author(s): Thomas R. Wier
*Proceedings of the 37th Annual Meeting of the Berkeley Linguistics Society: Special Session on Languages of the Caucasus* (2013),
pp. 96-110
Editors: Chundra Cathcart, Shinae Kang, and Clare S. Sandy

# Khevsur and Tush and the Status of Unusual Phenomena in Corpora

THOMAS R. WIER
*University of Chicago*

## Introduction

Recent years have seen an increasing realization of the threat posed by language loss where, according to some estimates, upwards of ninety percent of all languages may go extinct within the next century (Nettle & Romaine 2002). What is less often realized, much less discussed, is the extent to which linguistic diversity that falls within the threshold of mutual intelligibility is also diminishing. This is especially true of regions where one particular language variety is both widely spoken and holds especially high prestige across many different social classes and communities. In this paper, we will examine two such 'dialects' of Georgian: Khevsur and Tush, and investigate what corpora-based dialectology can tell us about phylogenetic and typological rarities found in such language varieties.

## 1        Ethnolinguistic Background

Spoken high in the eastern Caucasus mountains along the border with Chechnya and Ingushetia inside the Russian Federation, for many centuries, Khevsur and Tush have been highly divergent dialects of Georgian, perhaps separate languages, bearing a relationship to literary Georgian not unlike that of Swiss German and *Hochdeutsch* (see map, from Hewitt 1995:vi).

Figure 1.  Distribution of Kartvelian dialects and languages.
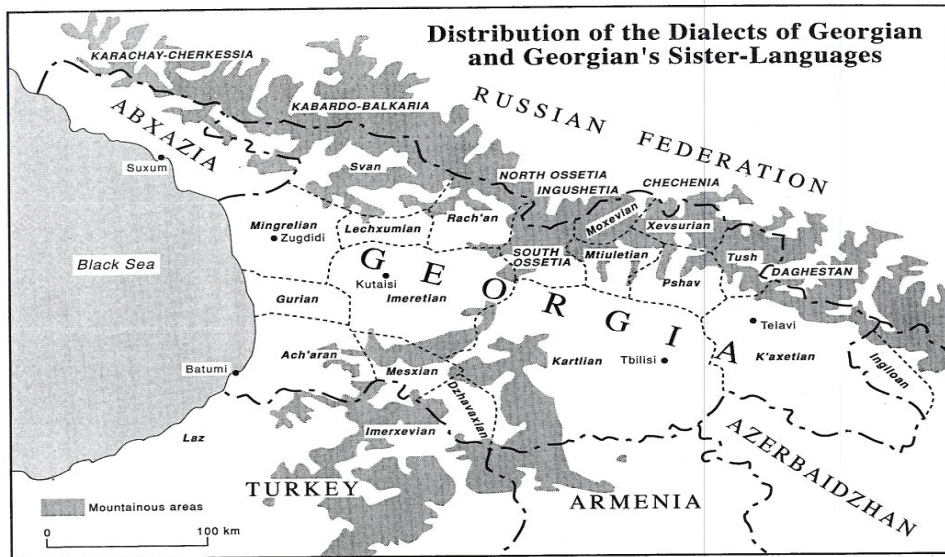


Among other reasons, for the better part of a thousand years they have had more intimate regular contact with Nakh-Daghestanian languages, and have borrowed numerous words from them: *balǧi* 'child' (cf. Lezgian *bal'a* 'child'), *bage* 'lip' (cf. Ingush and Chechen *baga*, Batsbi *bak* all 'mouth'), *ali* 'flame' (Ingush *ala*, Chechen *älu*, both 'flame'), *riq'e* 'stone' (cf. Botlikh *req'a* 'hill').  Independent of this, they have also developed a number of features not directly attributable to language contact, such as distinct lexical items (kood 'completely' instead of Standard sruliad, mtliad), differential suppletion patterns (*mi-ol* 'I'm going' vs. Standard *m-v-di-var* 'id.'), semantic shifts (Khevsur *xoq'ana* 'people' < *kveq'ana* 'land'; cf. standard *xalxi* 'people'), a different number system (*xut-oci* 'five-twenty' = 'hundred', vs. standard *asi*).

On the other hand, both dialects preserve archaisms that have been lost in some or all of the other contemporary Georgian dialects. For example, both dialects preserve a contrast between an aspirated and a glottalized uvular stop /qh/ vs. /q'/ and between the glide /y/ and /i/; the former in each case has merged uniformly with the latter in the Standard.  Both dialects also preserve the permansive tense-aspect morphology and syntax from Old Georgian, a kind of gnomic aorist representing events that are always true; this has been lost in other dialects, including the standard.  All of these difference add up to a significant barrier in communication for most speakers of the Standard language.

## 2    Previous Corpora and the Current Study

Previous work on Khevsur and Tush[1] focused primarily on lexicography, phonological processes particular to these dialects, and idiosyncrasies of paradigm formation.  While like all basic documentation work this is unquestionably valuable, and much of this work has been of very high quality, a number of problems recur throughout these texts which stand in need of improvement. Firstly, many of these dialect texts were collected more than a century ago without the aid of modern recording devices and methods of elicitation.  Most such texts were transcribed by hand on site, while fragile wax-cylinders and records (to the extent they ever existed) suffered the vicissitudes of neglect and outright destruction during Georgia's complex history in the twentieth century. Furthermore, metadata about the consultants' age, sex, location and relationship to the wider community were rarely or only incompletely recorded, thus making our task of interpretation all the harder.  The corpora were, without exception, published in Georgian script with all commentary and linguistic analysis in literary Georgian, with the result that these dialects (or languages) were essentially inaccessible to all non-Kartvelologists.  Thus scholars working on unrelated but geographically close Nakh-Daghestanian, Abkhaz-Adyghean, Turkic, Indo-European and other languages were incapable of comparing how this small area interacted within the larger ethnolinguistic context.  Above and beyond these problems, however, because the dialects themselves have in all likelihood been greatly restructured in the direction of the standard language, or replaced by some sort of *Umgangsprache*, it is difficult to know whether recordings and elicitations made today are capturing the 'same' language form as that recorded a century ago. Given that any dialect of Georgian, whatever its form, also generally lies at one extreme of complexity in terms of morphosyntax among the world's languages, even specialists can have a hard time penetrating the labyrinthine relationships between paradigms, argument structure, and clausal architecture.

---

[1]    Shanidze (1984), Dolidze (1975), Chincharauli (1960)

Figure 2. Example Khevsur text:

Or-n          mama-švil-n          q'opil-an.        švil-is-ad[1]
Two-NOM.PL    father-child-NOM.PL    be.PERF-3PL    child-GEN-ADV

mama-s       col-mo-u-q'van-a=v[2].              i   kal-s        ksl-is
father-DAT   wife-PVB-PRIV-have.ANIM=QUOT this wife-DAT   warp-GEN

ks-ov-a              da-u-c'q'-a=v,                          ksel
weave-TH-MAS.NOM   PVB-PRV-begin-AOR3SG=QUOT          warp

da-u-ks-a=v
PVB-PRV-weave-AOR3SG=QUOT

---

ADDITIONAL REMARKS:
1. Double case. Generate list of other instances of: -GEN-ADV
2. Noun incorporation. Generate list of other instances of: N-V (type 1).

---

FREE TRANSLATION:
There were two children (with a father). It's said their father had a wife for his children. This woman/wife supposedly began to weave a warp (on the loom), it's said she wove the warp.

(This text taken from Shanidze 1984)

The current study seeks to correct some of these problems by making full use of modern technology and approaches through an online digital dialect corpus. This gateway, modelled in part on the Perseus Project at Tufts University and Jost Gippert's TITUS-Projekt in Frankfurt, when completed will gather and present glossed and translated dialect texts in Georgian and Latin script in which each word is hypertexted to a dialect dictionary allowing scholars to see the cloud of meanings a given lexical entry may have. Beyond this textual level however the corpus envisions both intratextual and intertextual metatextual annotations of how a given text relates to the language and other texts in the corpus. Thus, intratextually, constructions which vary from standard Tbilisi Georgian, or from typologically expected norms, will be flagged to allow scholars unused to the norms of Kartvelian to focus on and potentially explain such differences.

Intertextually, constructions and forms in a text which differ from other texts in the corpus, either by different speakers, recorded in different locations or from different time periods, will be marked as such. This dual approach will allow scholars to see how all a given form behaves across a variety of constructural contexts. Furthermore, the digital recordings (both audio and video, where available) from which these texts were made will be made available along with each text, so that users can actually isolate the constructions in context. The goal is to give corpus users the fullest possible understanding of language use from a variety of different perspectives.

## 3 Typological Rara in the Corpus

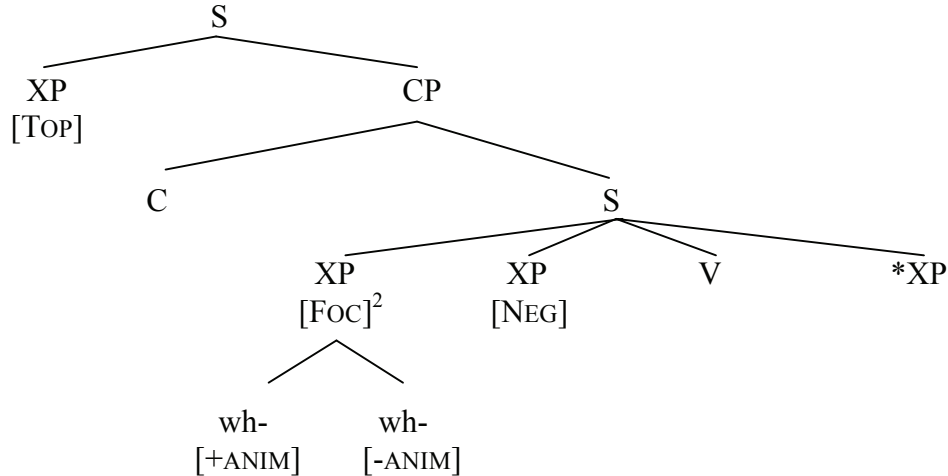### 3.1 Violations of Superiority Effects in Wh-Constructions

Such corpora tend to be 'messy' in the sense that they lack the idealization that accretes around studies based entirely on elicitation. Not only is this true of the current corpus of Khevsur and Tush, it reveals violations of typological norms not generally found in the already outré standard Georgian morphosyntactic system. So, for example, standard Georgian abides by the linguistic tendency that in constructions involving multiple wh-words an animacy restriction constrains otherwise rather free wordorder. In all varieties, such wh-words must surface preverbally (1a), and when both an animate and an inanimate wh-word are present, the animate wh-form must precede the latter (1b-c; Harris 1981:xx):

(1)  a. ra-s             a-k'et-eb-s              (*ras)                    (St. Geo.)
        what-DAT   PRV-do-TH-3SG
        'What is he doing?'
     b. vin             ra-s             a-k'et-eb-s
        who.NOM   what-DAT       PRV-do-TH-3SG
        'Who is doing what?'
     c. *ra-s           vin             a-k'et-eb-s
        what-DAT  who.NOM      PRV-do-TH-3SG
        'Who is doing what?'

Specialists who work on question constructions must often rely on elicitation because of the extreme rarity of multiple wh-constructions in corpora of natural languages. In the current corpus, however, not only are there numerous wh-constructions (or at least, more than expected from a corpus of considerably less than a million words), these multiple wh-constructions exhibit contrary tendencies in comparison with the standard dialect. As you can see in (2-3), in Khevsur, the wh-words still obligatorily surface before the verb complex (including negators). This is expected if all focal items surface immediately preverbally, as in Standard Georgian. However, other when you get two or more wh-words together, violations of superiority occur if one or more of the wh-words does not have a question interpretation, but rather is a homophonous indefinite pronoun (4-5).

(2)  ra-a(d)          v-k'l-av-t=av          ar=c             as=av            čven
     What-ADV    1-kill-TH-PL=QUOT   not=and        be.3Sg=QUOT  1PL
     dana-i=v,                        ra=ze           v-e-k'id-eb-i=ao=da
     knife-NOM=QUOT        what.DAT=on  1-PRV-hang-TH-PF-QUOT=and
     ga-u-šv-es.
     PVB-PRV-wipe-AOR3PL
     'Why are we killing it, it's not our knife, what will we hang it on and what will they wipe it with?'                                           (Kh)

(3)  peq<sup>h</sup>-t          ra-s              ča-v-i-c-om-d-i=v?
     foot-DAT.PL    what-DAT        PVB-1-PRV-fall-TH-IMPF-1/2IMPF=QUOT
     'How could I fall flat on my feet?'                                    (Kh)


(4)  čem-tan-it      ro        c'a-xv-av,        rom      ra-s              vin
     1SG-with-INST if       PVB-roll.up-TH that      what-DAT        who.NOM
     vis                 s-tx-ov-d-as=av
     who.DAT        3-ask-TH-IMPF-3SG=QUOT
     'If you will roll it up for me so that whatever [lit. 'what'] anyone [lit.
     'who'] asks of anyone [lit. 'of whom']…'                              (Kh)


(5)  šen              dana            ra=ši              vis
     2SgPoss        knife.NOM      what=in          who.DAT
     š-č'ir-d-eb-od-a=v?
     3-need-INGR-TH-COND-COND3SG=QUOT
     'As for your knife, why would anyone [lit. 'who'] need?'              (Tu)


This is interesting in that the morphological signaling that is usually required to obviate the underlying question interpretation to produce an indefinite in standard Georgian and western languages is not present here. Thus the templatic constraint that is present in standard Georgian (6; see Wier, forthcoming), which like these dialects also has a nonconfigurational clause structure, is weakened, in that variation in wh-word ordering occurs which does not in the standard (7).

(6)   Standard Georgian clause structure

```
                        S
           ┌────────────┴────────────┐
          XP                        CP
         [TOP]          ┌────────────┴────────────┐
                        C                         S
                              ┌──────────┬────────┴──────┐
                             XP         XP       V      *XP
                            [FOC]²     [NEG]
                          ┌──┴──┐
                        wh-    wh-
                     [+ANIM]  [-ANIM]
```

(7)   Khevsur and Tush focal structure:

<div align="center">

XP
[FOC]

</div>

That is, in comparison with the standard, the dialectal forms simply lack any templatic specification for ordering generalizations of wh-words as long as the wh-forms are all grouped together.

## 3.2   Noun Incorporation

In standard Georgian, noun incorporation is at best a marginal morphological process, allowable only about to the same extent that it is in English with N-N compounds:

(8)   a.  tav-mo-q'var-e                                                  (St. Geo.)
            head-[AGENT-love-AGENT]
            'proud' (lit. 'head-lover')
      b.  c'qal-c'a-ğ-eb-ul-i
            water-PVB-take-TH-PART-NOM
            'a drowning person' (lit. taken by water)  (Shanidze 1953:162)

---

²       For purposes of this article, I will remain ambiguous as to whether this is a grammaticalized syntactic feature which both wh-words with question interpretation and indefinite pronouns bear and which thus triggers particular orderings in the syntax, or whether [FOC] is a semantic or discourse functional feature.  I believe it is probably the former, but this is really an empirical question testable by discourse analysis and beyond the scope of this article.

These constitute the first of the by now familiar four-way typology for noun incorporation posited by Mithun (1984, 1986):

(9)    a. Type 1, Lexical compounding: heads reduce their valence by one, as in N + N > N; V + V > V; N + V > V; etc.

        b. Type 2, Manipulation of case: syntactic heads not only reduce their valence by one, another argument moves in to take its place.

        c. Type 3, Manipulation of discourse structure: heads (usually verbs) incorporate their dependents (usually nouns) to background incidental information.

        d. Type 4, Classificatory NI: dependents incorporate into heads to act as classifiers of a more general free dependent.

In contrast to the standard language, in Khevsur, there are a variety of examples of NI, including at least one textual attestation of Mithun's Type 4 NI (aka 'syntactic' NI). In the form in (10), for example, the root *elam-* 'squint' has been incorporated into the verbal root *q'opil* 'be'[3]. A number of different criteria suggest that this form has truly been incorporated. First, this particular example comes from a story that was elicited by Č'inč'arauli who was a native speaker of Khevsur dialect. The fact that a native speaker intuitively sees them as a prosodic unit suggests (though does not prove) that they are also a morphological unit. More direct evidence of this is that the accent shifts to mark the noun as part of the verbal prosodic phrase: thus the nominative suffix *–i* receives accent in *elam-í-* instead of the initial syllable as in the free word: *élam-i*.

(10)  Type 1: noun compounding
      i    kal      elam-í-q'opil.
      that woman   squint-NOM-be.PERF
      'The woman had a squinty-eye.'

Another argument that these arguments are truly incorporated into the verb is that the focal elements, which in standard Georgian must usually immediately precede the verb complex (excluding negators), here precede the incorporated noun:

(11)  a-dg-a             da    c'a-ma-vid,        ra     met'
      PRV-stand-AOR3SG   and   PVB-vent-GO.AOR3  what  more
      gza-í=a-kv…
      way-NOM=PRV-have
      'He got up and went off, however much more he had to go…'      (Kh)

Perhaps most interestingly, there are even examples of Mithun's Type 4 NI in both Khevsur and Tush dialects. In the Khevsur example in (12), the incorporated

---

[3]     Cf. standard literary Georgian: 'is kali elami iq'o'.

nominal *mrudi* 'crooked' modifies the more general nominal *k'utxi* 'corner', classifying it.

Type 4: 'syntactic' NI

(12) ert      k'utx-i=v              mrud-í=akv=av
     one      corner-NOM=QUOT       crooked-NOM=have.INAN=QUOT
     '[The house is good but it] has one crooked corner'

In (13), on the other hand, we see evidence that the incorporated element need not be less general than the nonincorporated element. The incorporated nominal, *kali* 'woman' is modified by an external adjectival nominal *ukmro* 'husbandless, unmarried'.

(13) Im          saxl=ši           ert    u-kmr-o         kal-í-q'opil
     that.OBL    house=in    one    PRIV-husband-PRIV    woman-NOM-be.PERF
     axalgazda    kal-i,           i    kal-s            u-k'itx-a=v
     young.NOM  woman-NOM  that  woman-DAT  PRV-ask.AOR-AOR3SG=QUOT
     'In that house there was a certain unmarried woman, a young woman, and [the old man] asked her…'

What is more interesting, this is actually an example of subject incorporation – rather a rare phenomenon crosslinguistically (Baker 1988, Spencer 1995). Although such constructions are by no means unattested crosslinguistically, within Kartvelian they are asymptotically rare, so their relative productivity in these more conservative mountain dialects/languages reinforces the importance of the study of less prestigious varieties of 'exotic' languages along with standard or more widespread varieties.

## 3.3    Suffixaufnahme, or Double-Case?

Another unusual property of these dialects that distinguishes them from standard Tbilisi Georgian is the use of double case constructions which, however, do not necessarily take part in any system of agreement. One basic kind of construction involves the use of a genitive followed by a dative, which may be in agreement with another dative marked head noun in the same clause:

(14) ša-x-q'var-d-a          im            col-s        im-isa-s
     PVB-3-love-IMPF-3SG  that.DAT      wife-DAT      that-GEN-DAT
     'He fell in love with [the other man's] wife.'                    (Kh)

Here, the genitival possessor *imisas* 'his' involves both a genitival suffix and a dative suffix to indicate the grammatical function of the possessum, here a dative-marked direct object *cols* 'wife'. This represents a conservative retention of an

Old Georgian *Suffixaufnahme*, whereby all genitives had to agree in case and number with the possessum, as in (15) where the *samebisa-* 'of the trinity' takes an additional instrumental case *–yta* to indicate it modifies *šec'evn-ita* 'help-INST'. This construction still exists marginally in modern standard Georgian, but only in elliptical possessive constructions where the head becomes elided, as in (16).

(15)  šec'evn-ita      c'[mid]isa      sam-eb-isa-yta            (493 A.D.)
      help-INST        holy-GEN.SG      three-COLL-GEN.SG-INST.SG
      'With the help of the Holy Trinity…'  (Silogava 1994)

(16)  vis          saxl-s        e-Z-eb?        čem-i      amxanag-isa-s
      whose.DAT    house-DAT      PRV-seek-TH    my-GEN     comrade-GEN-DAT
      'Whose house are you looking for? My friend's.'  (Boeder 2003:46)

Although rare in the standard, such constructions are quite common in both Khevsur and Tush; among the current texts in the corpus, at least 22 *–isa-s* constructions occur in the Khevsur corpus (~10k words) and 6 times in the Tush corpus (~40k words).  Less expected however are double case constructions which do not take part in any kind of agreement with a nominal head. There are a variety of different kinds of double case, including genitive+dative *-isa-s* (where in its non-agreeing manifestation it usually functions as an adjunct), genitive+nominative *–is-i*, genitive+instrumental *–is-it*, genitive+adverbial *–is-ad*, and, exceptionally, double instrumental *–it-it*.  Although almost all of these make use of a genitival stem plus some oblique case, it is unclear that the genitive contributes any meaning to the form; rather it seems simply to serve as the building block onto which further case forms (themselves rarely bearing a consistent meaning) can attach.  For example, a double genitive+adverbial frequently reflects a thematic recipient of *verba dicendi* as in (17) and (18) or *verba sentiendi* as in (19) and (20), but sometimes merely the experiencer (19), and sometimes the thing being experienced (20).  Finally, sometimes the double-case form marks the recipient, as in (21).

(17)  "peq^h-t          ra-s              ča-v-i-c-om-d-i=v?"
      foot-DAT.PL       what-DAT          PVB-1-PRV-fall-TH-IMPF-1/2IMPF=QUOT
      u-tkv-am-is              memcxvar-is-ad
      PRV-say-TH-3SG          shepherd-GEN-ADV
      '"How could I fall flat on my feet?" he says to the shepherd.'          (Kh)

(18)  diac-is-ad                    u-ʒex-eb-a=v                "k'arg oq^hšam, k'arg
      peasant.woman-GEN-ADV    PRV-call-TH-3SG=QUOT    good    strap     good
      mo-gv-i-mzad-e=v                    me      da      čem st'umar-sa=v!"
      PVB-1PL-PRV-prepare-AOR1/2=QUOT 1SG    and     my guest-DAT=QUOT
      'He calls the peasant woman:  'Prepare a good strap – a good one! – for me and my guest!"                              (Kh)

(19) im           cxvar-s      k'ide  su     da-ckmat'ur-eb-ul-sa
      that.DAT     sheep-DAT   again  just   PVB-?-TH-PART-DAT
      <u>im-is-ad</u>      e-xed-v
      <u>that-GEN-ADV</u>  PRV-see.IMPF-TH
      'The sheep again made [unclear] had just caught sight of <u>him</u>.'       (Tu)

(20) u-t'ir-is=ad       memr gada-k'id-eb-ul-iq'v      <u>kmr-is-ad</u>
      PRV-cry-3SG=and  then   PVB-irritate-TH-PART-3SGPf  <u>husband-GEN-ADV</u>
      'She cried and bugged her <u>husband</u>…"          (Tu)

(21) memr  is        ak'avan gē-y-k'et-a=d         dē-y-c'v-in
      then   that.NOM cradle  PVB-PRV-do-AOR3SG=and  PVB-PRV-burn-CAUS
      is           balğ-i.       Še-mē-y-č'ir        še-mē-y-č'ir
      that.NOM      child-NOM   PVB-VENT-PRV-cut   PVB-VENT-PRV-cut
      im           ak'avan-s    Še-mē-y-č'ir        še-mē-y-č'ir
      that.OBL      cradle-DAT   PVB-VENT-PRV-cut   PVB-VENT-PRV-cut
      im           ak'avan-s    im           balğ-<u>is-ad</u>,
      that.OBL      cradle-DAT   that.OBL      child-<u>GEN-ADV</u>
      še-mē-y-č'ir          še-mē-y-č'ir-a=d         mē-y-k'l
      PVB-VENT-PRV-cut   PVB-VENT-PRV-cut-Aor3Sg=and   PVB-PRV-die
      'Then he made the cradle and burned the little child.  He dashed and
      smashed on the cradle, he dashed and smashed on the cradle for the child,
      and he dashed and smashed and he [the child] died.'      (Tu)

It's also worth pointing that while such double-case forms may encode obligatory arguments, as above, they can also encode optional adjuncts, as in (22) and (23) below.

(22) im-is          udumliv      mamamtl-is        <u>ksl-is-ad</u>
      this-GEN      stealthily    father.in.law-GEN   weave-<u>GEN-ADV</u>
      gara          ga-mo-u-c'vd-av-a=d       da-u-mal-av
      heddle.stick  PVB-VENT-PRV-reach-TH-EP=and   PVB-PRV-hide-th
      'She stealthily reaches for her father-in-law's heddle-stick <u>for weaving</u> and hides it from him.'       (Kh)

(23) im          beber[4]      diāyc-s              k'ide  qʰel
     that.OBL     old.man      old.woman-DAT        again  hand
     c'a-mē-y-vl-ū                    im       napot'-is-ada=d
     PVB-VENT-PRV-seize-AOR3SG        that     roach-GEN-ADV=and
     c'ē-y-ǧ-o=d                      k'idoban=ši   čē-y-d-v.
     PVB-PRV-take-AOR3SG=and   bin=in          PVB-PRV-put-AOR3SG
     'The old man took the woman by the hand and brought it to the roach and
     put it [the roach] in the bin.'                              (Tu)

The genitive+adverbial double-case construction predominate throughout the
texts. However, as noted above, most of the other logically possible combinations
of genitive plus another case are possible. In (24), we see a postposed genitival
modifier *k'ac-is-i* 'man-GEN-NOM' agreeing in case with the head noun *gon-i*
'thought-NOM'[5].

***-is-i* [GEN-NOM]**
 (24) gon-i        k'ac-is-i      m-kon-d-a,                        magre
      thought      man-GEN-NOM 1SG-have.INAN.IMPF-IMPF-3SG        but
      ʒaǧl-i       v-i-q'av-i
      dog-NOM      1-PRV-be.IMPF-1/2
      'I had the thought of a man, but I was a dog'                (Kh)

In (25), we see a genitival form *dǧ-is-ita* 'day-GEN-INST' which functions as an
adjunct of time:

***-is-it* [GEN-INST]**
 (25) c'a-vid-o=d                       is            k'ac-i
      PVB-go.AOR-OPT3SG=and       this.NOM        man-NOM
      da-i-c'q'-eb-d-a                      ʒal-s.        dǧ-is-ita=c
      PVB-PRV-begin-TH-IMPF-IMPF3SG       force-DAT       day-GEN-INST=too
      ša-i-ʒl-eb-od
      PVB-PRV-can-TH-COND
      'The man went, [and] really got a start.  By day he was able.'   (Kh)

Finally, we also find examples of multiple exponence of the same case, as in (26),
where one instrumental case suffix is followed by another identical case suffix

---

[4]      Interestingly, this argument lacks the expected narrative case suffix *–m* (standard
Georgian *–ma*).  It does however use the oblique form of the article used for the narrative case,
rather than the nominative case of the article *is*.  This is another example of a shift from head-
marking to dependent-marking in these dialects.
[5]      In standard such postposed modifiers are, rarely, possible in high literary style, but they
would be constructed differently in the nominative: *goni k'ac-isa*, where the suffix *–isa* is the
extended form of the genitive from Old Georgian used in relics like this.

apparently with no difference in meaning from the standard form *ik-it* [there-INST] 'over there':

***-it-it* [INST-INST]**
(26) im            dro-s        i      mamamtl-is        xalx
     that.OBL     time-DAT     that   father.in.law-GEN    people
     mo-svl-i=v             cxra-mta-s       <u>ik-it-it</u>:
     PVB-go.MAS-NOM=QUOT    nine-mountain-DAT   there-INST-INST
     'At that time, her father-in-law's people were going <u>up there</u> to Nine-
     Mountain.'                                      (Kh)

What is interesting about most of these double-case constructions is that they do not fit into standard discussions of the phenomenon of *Suffixaufnahme*, in which dependents must agree with heads in case and perhaps other features for the simple reason that in most of the above cases there is never any evidence of a head noun with which the nominal in question could stand in agreement. Even when the doubly-marked noun serves discourse-functionally as a modifier, as in (21), it does always do so syntactically, as when *balǧ-is-ad* [child-GEN-ADV] 'for the child' stands as a sentential adjunct and the extra case does not agree with that which it discourse-functionally modifies, *ak'avan-s* [cradle-DAT].

An immediate question thus arises: if these doubly marked constructions do not uniformly arise from some process of agreement, as they would have in Old Georgian, where do they come from? One intriguing possibility is that they result from language contact with Nakh languages to the immediate north of the Khevsur and Tush dialect regions in Georgia. As noted above, these dialects have for many centuries been in contact with Chechen, Ingush, Bats, and other North Caucasian languages, which have the unusual property that oblique cases do not attach directly to the nominal root, but rather to a stem formant exclusively used for obliques. These oblique stem formants vary both in form and distribution across Nakh-Daghestanian languages, but typically behave much like the Archi and Batsbi forms in (27) and (28) below:

(27) Archi (Diana Forker p.c.)
    a. gel
       cup.ABS.SG
    b. gel-li-s
       cup-OBL-DAT

(28)  Tsova-Tush (Holisky and Gagua 1994)

|        | 'bear'     | 'broom'      |
|--------|------------|--------------|
| NOM    | ča         | kož-o        |
| GEN    | ča-i-$^n$  | kož-ni-$^n$  |
| DAT    | ča-i-n     | kož-ni-n     |
| ERG/INST | ča-i-v   | kož-ni-v     |
| CON    | ča-i-x     | kož-ni-x     |
| ALL    | ča-i-go    | kož-ni-go    |
| ADV    | ča-i-ǧ     | kož-ni-ǧ     |
| COM    | ča-i-ci$^n$ | kož-ni-ci$^n$ |

Nominals having distinct oblique stems are not entirely rare; they occur e.g. in well-known Indo-European languages like Latin and Greek. The interest in these Nakh-Daghestanian forms lies in the fact that the oblique formant is separable but semantically vacuous from the point of view of case features. That is, although they sometimes bear other features like the singular/plural contrast in the Batsbi forms, the oblique formant serves only to provide a licit stem for the actual case suffix. It is this fact that brings us back to a discussion of Khevsur and Tush double-case, since the genitival suffix in most instances is semantically vacuous in exactly the same way, with the difference that that most of the oblique cases (i.e., those that are not nominative in the Kartvelian context) may also surface attached directly to the nominal root without any oblique stem intermediary. Thus, while the system is not identical to any Nakh-Daghestanian one, it is as if the Khevsur and Tush speakers, many of whom were also presumably bilingual in one or more other Nakh-Daghestanian language, were borrowing a constructional device from those languages but using their own indigenous resources, *Suffixaufnahme*, which had originally served a quite different function, to do so.

## 4    Conclusion

This survey of properties of Khevsur and Tush dialects of Georgian has shown the value of data-focused corpus studies for studies of linguistics and typology, since they have a tendency to confound traditional notions of how grammars are supposed to work. Khevsur and Tush show that even when a standard form of the language abides by supposed notions of superiority in wh-constructions, some dialects can and do violate these norms. They also show that noun incorporation may indeed vary, and even unusual forms of noun incorporation such as subject NI or syntactic NI may occur in one variety while in another variety NI is almost completely absent. Finally, most interestingly, even in languages noted for obscure construction types, such as Suffixaufnahme or double-case, nonstandard varieties of language may contain typologically unusual variants of those same construction types, rarities within rarities.

**References**

Baker, Mark. 1988. *Incorporation: a theory of grammatical-function changing*. Chicago: University of Chicago Press.

Ch'inch'arauli, Aleksandre. 1960. *Xevsurulis taviseburebani: t'ekst'ebita da indeksit*. [Peculiarities of Khevsur: with texts and an index]. Tbilisi: Press of the Georgian SSR Science Academy.

Dolidze, Givi. 1975. *Xevsuruli t'ekst'ebi: leksik'onebiturt* [Khevsur texts: along with lexica]. Tbilisi: TSU Press.

Harris, Alice C. 1981. *Georgian Syntax*. Cambridge: CUP.

Hewitt, S. George. 1995. *Georgian: a structural reference grammar*. Philadelphia: John Benjamins.

Holisky, Dee, and Rusudan Gagua. 'Tsova-Tush (Batsbi)' in *The Indigenous Languages of the Caucasus, 4: North East Caucasian, Part 2*, ed. by Rieks Smeets, 147-212. Delmar, N.Y.: Caravan Press.

Mithun, Marianne. 1984. 'The Evolution of Noun-Incorporation.' *Language* (60:4): 847-894.

Mithun, Marianne. 1986. 'On the Nature of Noun-Incorporation.' *Language* (62:1): 32-37.

Nettle, Daniel and Suzanne Romaine. 2000. *Vanishing-Voices: the Extinction of the World's Languages*. Oxford: OUP.

Shanidze, Ak'ak'i. 1953. *Kartuli gramat'ik'is sapuʒvlebi* [Foundations of Georgian Grammar]. Tbilisi: TSU Press.

Shanidze, Ak'ak'i. 1984. *Txzulebani tormet' t'omad, Vol. 1: kartuli k'iloebi mtaši*. [Compositions in twelve volumes: Georgian dialects in the mountains]. Tbilisi: TSU Press.

Dr. Thomas R. Wier
Dept. of Linguistics
1010 E. 59th St.
Chicago, IL 60615

trwier@uchicago.edu