References

Anward, Jan. 1994. 'Från *potatisen kokas* till *koka potatisen*' In *Språkbruk, grammatik och språkförändring*. Lund: Institutionen för nordiska språk.

Goldberg, Eli, Richard Kittredge & Alain Polguère. 1988. 'Computer generation of marine weather forecasts'. *Journal of atmospheric and oceanic technology* 5:4, 472-83.

Kittredge, Richard et al. 1973. *TAUM-73*. Montréal: Université de Montréal.

Sigurd, Bengt, Mats Eeg-Olofsson, Caroline Willners & Christer Johansson. 1992. 'Deep comprehension, automatic translation and generation of weather reports (*Weathra*)'. *Proceedings of the Coling* 92, Nantes, Vol II, 749-55.

Sigurd, Bengt, Birgitta Lastow & Tomas Vävargård. 1996. 'Computer generation of weather overviews'. Paper presented at The European conference on artificial intelligence (ECAI-96), Budapest.

Bengt Sigurd <Bengt.Sigurd@ling.lu.se>
Birgitta Lastow <Birgitta.Lastow@ling.lu.se>
Hong Gao 高虹, University of Toronto <hgao@utm.utoronto.ca>
Mats Eeg-Olofsson <Mats.Eeg-Olofsson@ling.lu.se>

Lund University, Dept. of Linguistics
Working Papers 51 (2005), 199–207

Basic verb frequency in Megrelian

Revaz Tchantouria and Karina Vamling

Our aim is to investigate which verbs show the highest textual frequencies in the Kartvelian language Megrelian. The general assumption is that unmarked verbs represent lexical core concepts and that they will emerge among the verbs with the highest text frequencies, showing crosslinguistic similarities.

1 Background

1.1 Megrelian

Megrelian is a Kartvelian (South Caucasian) language spoken in Western Georgia. The number of speakers is estimated to approximately half a million (no official census data are available, as Megrelians are registered as Georgians). Apart from a short period around the 1930s, when some attempts were made to standardize the language (Vamling 2000), the language is not used in writing. Speakers are generally bilingual, using Georgian as their literary language for all administrative and educational purposes.

1.2 Basic verbs and text frequency[1]

Unmarked verbs are general and basic and are assumed to represent lexical core concepts and appear among the verbs with the highest text frequencies. When compared in various languages they also show patterns of cross-linguistic regularities. Typically, such verbs – *see, say, take, go, know* – form the nucleus of semantic fields: Perception, Verbal Communication, Possession, Motion, Cognition (Viberg 1994).

Such a concept of markedness is central to typological research on hierarchies and prototypes (following Greenberg 1966) and allows not only for binary relations but extends to several values along a hierarchy of relative markedness. An example is the hierarchy of sense modalities, which are ranked from unmarked to higher degree of markedness: *sight* > *hearing* >

[1]This study on verb frequency in Megrelian has been undertaken in connection with research on cross-linguistic lexicology headed by professor Åke Viberg at the Department of Linguistics, Lund Univerisity. An earlier version of the paper was presented at the Chikobava Conference in Tbilisi in 1998.

touch > *smell, taste*. For instance, a verb having the basic meaning *see* may extend its use to other positions lower in the hierarchy (Viberg 1984). In short, unmarked forms are characterized by a cluster of properties, where the most important ones are: (1) simple root structure, (2) phonologically simple form, (3) often include suppletive forms or appear as irregular verb, (4) high text frequency, (5) have several secondary meanings, (6) provide a basis for the development of grammatical markers, (7) function as syntactic prototypes, (8) be favoured during the first phase in first and second language acquisition (Viberg 1990:399).

By investigating the basic verb vocabulary of eleven European languages Viberg has identified several European areal and subareal features. Verbal concepts that were found among the 20 most common verbs in the languages under investigation are BE, CAN, GIVE, TAKE, SAY, SEE (Viberg 1993:347), which are called the nuclear verbs. Similar studies of non-European languages as Chinese and Arabic confirm the set. The verbs GO, MAKE and possibly also HIT, WANT, COME, KNOW are also proposed to belong to the set of nuclear verbs, as their frequency patterns are very close to the first set.

The Megrelian data is discussed in connection with these patterns. Statistical data on word frequencies in the Kartvelian languages are scarce, and therefore interesting from a comparative point of view.

2 Investigating the Megrelian data

2.1 The data collection

Our study of verb frequency in Megrelian is based on a database that is being set up in the project *Reference grammar for Megrelian* (The Bank of Sweden Tercentenary Foundation). The database amounts to approximately 180.000 running words and includes various materials, mainly of ethnographical character (Kipshidze 1914, Khubua 1937, Samushia 2001) but also historical-political texts (Zhvania 1931) and field notes of personal narratives. The early unpublished materials were collected during fieldtrips in the 1930s and 1940s by Georgian linguists at the Chikobava Institute of Linguistics, Georgian Academy of Sciences. Later personal narratives have been recorded by the present authors during fieldtrips to Georgia. The published texts have been computerized by using scanner technique and OCR programmes. Concordances were generated using the programme *Conc*.

Table 1. Tense-aspect-mood (TAM) in Megrelian ('s/he writes it')

Present	*č'aruns*	Aorist	*(do)č'aru*
Future	*doč'aruns*	Optative	*(do)č'aras*
Future imperf.	*č'arundas iʔuapu / iʔii*	Conditional II	*(do)č'aruk'o(n)*
Imperfect	*č'arundu*	Evidential 1	*(du)uč'aru(n)*
Subj. pres.	*č'arundas*	Evidential 2	*(du)uč'arudu*
Subj. fut.	*doč'arundas*	Perf. subjunctive	*(du)uč'arudas*
Habitual	*doč'arundu*	Conditional III	*(du)uč'aruduk'o(n)*
Cond. pres.	*č'arunduk'o(n)*	Evidential 3	*noč'arue(n)*
Cond. fut.	*doč'arunduk'o(n)*	Evidential 4	*noč'aruedu*
Cond. imperf.	*č'arunduk'on iʔuapudu/iʔidu*	Subjunctive IV	*noč'aruedas*
		Conditional IV	*noč'arueduk'o(n)*

2.2 Megrelian verb forms

The verbal morphology of Megrelian is very rich, making calculations of verb frequency a complicated task. For instance, the list of forms assembled for the lexeme 'be' holds 150 different forms (types). The verbal conjugation, shown in Table 1 for the verb *č'arua* 'write', comprises over 20 tense/aspect/mood categories.

The forms include cross-reference markers showing person and number of subject, direct and indirect objects, spatial orientation (*ko-mortu* 's/he came here', *ke-mertu* 's/he went there'), causation (*kəmaɣal-apu* 's/he made her/him bring something') and other valency changing operations (for instance, benefactive *kəmom-i-ɣu* 's/he brought something for me'. For further information on Megrelian verbal morphology, cf. a short overview given in Vamling and Tchantouria 1993 (also available online) .

In order to obtain data on Megrelian verb frequency that would be as closely comparable as possible to data from different languages in earlier investigations, some Megrelian verbs were counted together in determining the frequency rankings. For instance, certain Megrelian verbs are sensitive to animacy in the selection of their objects. The verbs *mi-deʔonə* 's/he took him/her there' and *kə-miʔonə* 's/he took him/her here; brought' are used when the object is animate, as shown in examples (1) and (3), whereas *mi-deɣə* 'he took it (there)' and *kə-miɣə* 's/he took it (here), brought' select inanimte objects (2, 4).

1. muma-k skua (*diška) kalak-ša mi-deʔonə
   father-ERG son (*firewood) city-ALL OR-take.AOR
   'Father took his son (*firewood) to the city.'

2. muma-k diška (*skua) kalak-ša mi-deɣə
father-ERG firewood (*son) city-ALL OR-take.AOR
'Father took firewood (*his son) to the city.'

3. sk'ua-k ʔuča st'umari (*sačukar-i) kəmiʔonə
son-ERG home.ALL guest (*present-NOM) take.AOR
'Son brought home a guest (*present).'

4. sk'ua-k ʔuča sačukar-i (*st'umari) kəmiɣə
son-ERG home.ALL present-NOM (*guest) take.AOR
'The son brought home a present (*guest).'

The highly frequent verbs of possession are another example of the importance of the animacy distinction. The verb of animate possession is *ʔuns* 's/he has him/her' (5) and of inanimate possession *uɣu* 's/he has it' (6). These verbs have been counted separately at first and subsequently their frequencies were summed up for the frequency ranking (cf. Table 2).

5. cira-s žiri ǯima, k'at'u (*ušk'uri) ʔuns
girl-DAT two brother, cat (*apple) have.PRES
'The girl has two brothers, cats (*apples).'

6. cira-s žiri ušk'uri (*ǯima, *k'at'u) uɣu
girl-DAT two apple (*brother, *cat) have.PRES
'The girl has two apples (*brothers, *cats).'

Both proclitic and enclitic elements occur in Megrelian. Proclitics are the negating *va-*, *ve-* (*va-gičku* 'you don't know it', *ve-ečinə* 'he/she didn't know him/her'). Enclitic elements are the general subordinator *-n(i)* (*va-mortu-ni* 'that s/he didn't come'), the interrogative marker *-o* (*mortu-o*? 'Did he come?') and *-da* 'if' (*vamortu-da* 'if s/he didn't come'). Verb forms including these clitic elements have been counted together with the corresponding forms lacking such elements. For instance, out of all occurrences of WANT, approximately 15 % are negated forms, as in *vamok'o* 'I don't want it'.

3 Megrelian basic verb frequency

3.1 Frequency rankings

The frequencies for the most common verbs in our Megrelian database are given in Table 2. As Megrelian has no infinitives, finite past forms in the first person singular are shown in the table. (This particular form has been chosen as it provides important grammatical information. The *v-* (including the variants *b-*, *p-* and *p'-*) and *m*-markers separated by hyphens mark the first person subject. The *m*-forms indicate that these verbs occur in inversive constructions with dative marked subjects.)

Table 2. Frequency ranking of the 20 most common verbs in Megrelian

		frequency		
1	Be	3752	*v-ordi*	'I was'
2	Tell	2323	*v-uc'i*	'I told her/him something'
3	Go	2091	*mida-p-rti*	'I went'
4	Come	977	*komo-p-rti*	'I came'
5	Have	728	*m-ʔundu*	'I had it [anim]' (278)
			m-iɣudu	'I had it [inanim]' (450)
6	See	520	*ko-b-ʒiri*	'I saw it/her/him'
7	Give	450	*keme-p-či*	'I gave it to her/him'
8	Say	282	*ko-p-tkvi*	'I said it'
9	Ask	270	*ko-p-k'itxi*	'I asked her/him'
10	Want	262	*m-ok'odu*	'I wanted it'
11	Know	205	*m-ičkudu*	'I knew it'
12	Kill	185	*do-p'-ili*	'I killed it'
13	Take	176	*mide-p'-ʔoni*	'I took him/her there' (69)
			mole-p'-ʔoni	'I took him/her here' (44)
			mide-b-ɣi	'I took it there' (41)
			mole-b-ɣi	'I took it here' (22)
14	Call	146	*do-v-uʒaxi*	'I called him/her'
15	Begin	140	*di-b-č'q'i*	'I began it'
16	Die	136	*do-b-ɣuri*	'I died'
17	Do	133	*ko-p-kimini*	'I did it'
18	Collect	109	*do-p-šaq'ari*	'I collected it'
19	Speak	107	*v-ičiebudi*	'I spoke'
20	Catch	100	*o-p'-č'opi*	'I caught it/him/her'

Comments

1	Be	Suppletive forms occur based on the roots *-r-* and *-ʔop-*
2	Tell	This verb is always trivalent
3	Go	The prefix *mi-* shows spatial orientation 'away'
4	Come	The prefix *mo-* shows spatial orientation 'here'
11	Know	*m-ičkudu* 'I knew it' is used in the sense of 'having knowledge about something'. Cf. *v-ičinendi* 'I knew him/her, was aquanted with him/her'.
13	Take	The verb *ge-v-č'opi* (63) has the meaning 'I took something from somewhere'
17	Do	A synonymous verb *gavak'eti* 'I did it' (49) is also used

The data shown for Megrelian is based on a small database, whereas the data from the other languages rely on varied texts from large corpora. Another difference is that Megrelian does not have any written tradition, while

Table 3. Frequency ranking of the 20 most common verbs in Megrelian compared to basic verb frequencies in Swedish, English and Russian

Rank	*Megrelian*		*Swedish*	*English*	*Russian*
1 BE	*vordi*	'I was'	1 vara	1 be	1 byt'
2 Tell	*vuc'i*	'I told her/him it'	–	–	–
3 GO	*midaprti*	'I went'	12 gå	9 go	4 (po)jti
4 COME	*komoprti*	'I came'	7 komma	11 come	–
5 HAVE	*mʔundu, miyudu*	'I had it'	13 ha	14 have	–
6 SEE	*kobʒiri*	'I saw it/her/him'	14 se	12 see	7 videt'
7 GIVE	*kemepči*	'I gave it to her/him'	13 ge	14 give	8 dat'
8 SAY	*koptkvi*	'I said it'	11 säga	6 say	3 skazat'
9 Ask	*kopk'itxi*	'I asked her/him'	–	–	–
10 WANT	*mok'odu*	'I wanted it'	16 vilja	–	12 xotet'
11 KNOW	*mičkudu*	'I knew it'	–	13 know	6 znat'
12 Kill	*dop'ili*	'I killed it/her/him'	–	–	–
13 TAKE	*midep'ʔoni*	'I took him there'	10 ta	10 take	14 vzjat'
14 Begin	*dibč'q'i*	'I began it'	–	–	19 načat'
15 Call	*dovuʒaxi*	'I called him/her'	–	–	–
16 Die	*dobyuri*	'I died'	–	–	–
17 Do	*kopkimini*	'I did it'	8 göra	4 do	–
18 Collect	*dopšaq'are*	'I collected it'	–	–	–
19 Speak	*vičiebudi*	'I spoke'	–	–	5 govorit'
20 Catch	*op'č'opi*	'I caught it/him/her'	–	–	–

Sources: Swedish (Viberg 1990:394), English (Viberg 1994:178), Russian (Zasorina 1977).

the other languages investigated are standardized literary languages. Despite these differences, the frequency rankings for the most common verbs show great similarities. As an illustration, the frequency rankings of the Megrelian most common verbs are compared to the frequency rankings for Swedish, English and Russian corresponding verbs. As you may see in Table 3, all the nuclear verbs BE, CAN, GIVE, TAKE, SAY, SEE are present in the first group of languages, whereas CAN is lacking in the Megrelian list (cf. further comments below). The verbs in the second group – GO, MAKE, HIT, WANT, COME, KNOW – are also generally found among the top 20 verbs, with the exception of HIT and MAKE in Megrelian. Some verbs in the list are shown to appear with high frequency rankings only in Megrelian: 'tell', 'ask', 'kill', 'call', 'die', 'collect' and 'catch'.

3.2 European areal features

The very high frequency of BE and HAVE is noted as a European areal feature, that is not as widespread in other parts of the world (Viberg

1990:400). Megrelian is found to follow this pattern, although the frequency of HAVE is not at the very top.

Another widespread European feature is the high frequency of modal auxiliary verbs – 'shall', 'will', 'must', 'can'. As expected, this feature is not present in Megrelian, due to its rich verbal morphology. Possibility/ability is expressed by a main verb (7) or morphologically by the potential circumfix *a--e* (8).

7. šemilebə te davaleba gavak'ete(n)
S1SG.can.PRES this task S1SG.O3SG.do.OPT
'I can do this task'

8. te davaleba mak'etine
this task S1SG.POSS.do.PRES
'I can (am able to) do this task'

Verbs such as *få* 'get', *make* and *faire* that occur in periphrastic causative constructions are common in Standard European languages, but are not found among the top 20 verbs in Megrelian. Causatives in this language are primarily formed by morphological means with the suffix *-apu*. (9) shows the causative verb and (10) the corresponding non-causative form.

9. Otar-k Nodar-s berg-i dač'k'ad-apu
Otar-ERG Nodar-DAT hoe-NOM forge.AOR-CAUS
'Otar made Nodar forge a hoe'

10. Otar-k berg-i doč'k'adu
Otar-ERG hoe-NOM forge.AOR
'Otar forged a hoe'

When the figures for verbs in Megrelian with the causative suffix *-apu* were calculated they received a frequency ranking corresponding to position 8 (384), i.e. between SAY and GIVE.

3.3 Other properties of unmarked verbforms

As noted above, unmarked forms are characterized by a cluster of properties, including high text frequency. In this concluding section we would like to point at two other features of basic verbs in Megrelian.

The simplicity of verb roots is one such feature. The root structure of the verbs in the list is generally very simple: SAY, *ko-p-tkvi* 'I said' – root: *tk*, KNOW, *m-ičkudu* 'I knew' – root: *-čk*. Even some mono-consonantal roots are found among the most frequent verbs at the top of the list: BE, *v-ordi* 'I

was' – root: *r/ʔ*; TELL *v-uc'i* 'I told him something' – root: *c'*; *GIVE, keme-p-či* 'I gave it to her/him' – root: *-č*.

Finally, we find a verb with top frequency including suppletive forms: BE. The root element *-r-* occurs in for instance the present *v-orek*, imperfect, optative *ko-v-orda* and the *-ʔ-* root in the future *v-iʔik*, evidential 1 *v-ʔopek* and evidential 3 *no-v-ʔopuek*.

4 Summary

The frequency rankings for basic verbs in Megrelian show great similarities to comparable data from other languages, despite the fact that Megrelian has no written standard and the study is based on a rather small database (about 180,000 running words).

A European areal feature that is found in Megrelian is the very high frequency of BE and HAVE, although the frequency of HAVE is not among the very highest. Basic verbs not found in the Megrelian top 20 list are CAN, MAKE and HIT, which partly may be explained with reference to features of Megrelian verbal morphology: the presence of morphological causatives, potential and rich modal forms. Verbs that show high frequency rankings in Megrelian but not in the other languages investigated are: 'tell', 'ask', 'kill', 'call', 'die'.

References

Greenberg, Joseph. 1966. *Language universals, with special reference to feature hierarchies* (Janua linguarum, series minor 59). The Hague: Mouton.

Khubua (Xubua), Mak'ar. 1937. *megruli t'ekst'ebi*. Tbilisi.

Kipshidze (Q'ipšiʒe), Iosip. 1914. *rčeuli txzulebani*. Tiflis.

Samushia (Samušia), K'alist'rat'e. 2001. *ʒveli k'olxuri (megruli) leksimyerebi*. Tbilisi: egrisis macne.

Vamling, Karina. 2000. 'Language use and attitudes among Megrelians in Georgia'. *Analysis of current events* (Slavic & East European Studies, Baylor University) 12:5-6, 9-11.

Vamling, Karina & Revaz Tchantouria. 1993. 'On subordinate clauses in Megrelian'. In Kees Hengeveld (ed.), *The internal structure of adverbial clauses, Eurotyp Working Papers* V, 67-86. (Online version: http://www.bonetweb.com/caucasus/KV/CAUCDB/Publications.html).

Viberg, Åke. 1984. 'The verbs of perception: a typological study'. *Linguistics* 21:1, 123-62.

Viberg, Åke. 1990. 'Svenskans lexikala profil'. In Erik Andersson & Marketta Sundman (eds.), *Svenskans beskrivning* 17, 391-408. Åbo: Åbo Akademis förlag.

Viberg, Åke. 1993. 'Crosslinguistic perspectives on lexical organization and lexical progression'. In Kenneth Hyltenstam & Åke Viberg (eds.), *Progression & regression in language: sociocultural, neuropsychological and linguistic perspectives*, 340-385. Cambridge: Cambridge University Press.

Viberg, Åke. 1994. 'Vocabularies'. In Inger Ahlgren & Kenneth Hyltenstam (eds.), *Bilingualism in deaf education* (*International studies on sign language and communication of the deaf* 27), 169-99. Hamburg: Signum.

Zasorina, L. N. (ed.). 1977. *Častotnyj slovar' russkogo jazyka*. Moskva: Russkij jazyk.

Zhvania (Žvania), Isak'i. 1931. *muč'o lemendə samargaloši moxande q'azaq'oba sabč'oepiši xešuulobašeni*. Zugdidi.

Revaz Tchantouria, School of International Migration and Ethnic Relations, Malmoe University <Revaz.Tchantouria@imer.mah.se>

Karina Vamling, School of International Migration and Ethnic Relations, Malmoe University <Karina.Vamling@imer.mah.se>