Existing hypotheses on lexical borrowing

1. Some concepts are more resistant to lexical borrowing than others especially basic vocabulary
cf. English hand, German Hand; English ballet, French ballet
but there is no scientific definition of the concept “basic vocabulary”
the Swadesh list purports to be a universal list of the human condition
but there are exceptions (e.g. ‘ice’, ‘horn’)
such a list of lexical concepts is a hypothesis that must be tested in a systematic and cross-linguistic manner
especially some semantic fields
but the identification of such semantic fields also needs systematic investigation
some semantic fields behave differently in different cultural environments
e.g. Austronesian languages have few loanwords connected to the sea, while (High) German has borrowed much of this vocabulary from Low German
especially more frequent words
but again there are almost no empirical investigations on the basis of a wide range of languages representing various kinds of language contact situations
especially some parts of speech
cf. the hypothesis that nouns are more borrowable than verbs, or that grammatical words are more resistant to borrowing than lexical words

2. Loanword Typology Project
(project leaders: Martin Haspelmath, Uri Tadmor)
worldwide sample of languages representing different language families, areas, and types of contact
in this case ca. 50 languages
list of words must be more extensive than Swadesh and include words from presumed non-basic vocabulary
in order to test Swadesh’s hypothesis
in this case a list of over 1300 words as used for the “Intercontinental Dictionary Series” project
each language is investigated by a specialist on the given language (and its history)
each author submits a data base, i.e. an electronic list of the lexical items, corresponding to the given 1300 concepts identifying loans and their immediate sources
each author also submits a chapter presenting empirical and theoretical generalizations
the project is made possible by recent work compiling dictionaries of various languages of the world, including comparative and etymological dictionaries

3. Results of the project
for each lexical concept, the percentage of languages with a loanword this will give a list of lexical concepts from the most to the least borrowable and facilitate the testing of hypotheses concerning the relevance of semantic field, frequency, part of speech, etc.
the comparison of different language contact situations
will facilitate the testing of hypotheses concerning the relevance of the social
environment for lexical borrowing

Bezhta

4. Bezhta (Kapuchi) [kap], ca 6461 speakers (2002 Russian census; informal
estimates give around 12,000) in west of Republic of Daghestan
main dialects: Bezhta Proper, Khasharkhota, Tladal
Bezhta Proper has some idiosyncratic properties:
shift of *r to y
shift of *h,*s to h, ? respectively with umlaut of neighboring vowels

5. Nakh–Daghestanian language family:

Nakh–Daghestanian (North–East Caucasian, East Caucasian)
Nakh languages
Daghestanian languages
Avar–Andi–Tsezic languages
Avar–Andic languages
Avar [ava]
Andic languages
Tsezic languages
West Tsezic languages
Khwarshi [khv] (incl. aberrant Inkhoqwari “dialect”)
Tsez (Dido) [ddo]
Hinuq [gin]
East Tsezic languages
Bezhta (Kapuchi)
Hunzib [huz]
Lak
Dargi
Lezgic languages
Khinalug

6. Major resources, mostly from recent years
a. dictionaries of Bezhta, Hunzib, Tsez, and Hinuq (but not Khwarshi)
b. comparative dictionary of North Caucasian (Nikolayev and Starostin 1994)

7. Major lacuna
no systematic work on inter–Tsezic loans
partly in absence of detailed comparative phonology

8. Inherited vocabulary
a. vocabulary inherited from Proto–East Tsezic (ca 1000 BP)
b. vocabulary inherited from Proto–Tsezic (ca 2000 BP)
c. vocabulary inherited from Proto–Nakh–Daghestanian (ca 8000 BP)
d. in principle, also vocabulary inherited from
   i) Proto–Avar–Andi–Tsezic
   ii) Proto–Daghestanian

9. Major non–Tsezic linguistic contacts of Bezhta:
direct contacts:
   Georgian [kat] (at least from early centuries CE)
   Avar (long–standing, heavy since 1860s)
   Russian [rus] (direct influence mainly from 1950s; some earlier influence
   through Georgian and especially Avar)
indirect contacts:
Arabic [arb] (at least from C18; assumed to be through Avar, with possible transmission to Avar via Turkic, esp. Azerbajiani [qzj] and Kumyk [kum])
Persian [prs] (assumed to be through Avar, with possible transmission to Avar via Turkic, esp. Azerbajiani and Kumyk)
Turkic, esp. Azerbajiani and Kumyk (assumed to be through Avar)

Sources of loans

10. Georgian
main studies Xalilov (1993; 2004)
loans into
Bezhta/Hunzib (ca 400 items)
Tsez/Hinuq (ca 129 items, of which 88 in Tsez or Tsez and Hinuq)
Kharshki (ca 10 items, all nouns — less borrowing? or secondary loss?)
some predate language-specific sound changes within Tsezic, so probably reflect Proto-Tsezic period loans, e.g. Geo gemo ‘taste’, Bezhta-Hunzib gemo, Tsez-Hinuq gimu; Geo okro ‘gold’, Bezhta-Hunzib okro (Bezhta Proper okko), Tsez-Hinuq ukru ‘money, silver’
some show typical Old or Middle Georgian forms, e.g. O/MiGeo toqi ‘hoe’, MoGeo toxî, Bezhta–Hunzib–Hinuq toqi, Tsez toxî’, though most are Modern Georgian
Bezhta–Hunzib (but not Tsez–Hinuq) have some Georgian loan verbs as parts of light verb constructions (LVCs), e.g. Bezhta dayup’a –owal ‘spoil’ < Georgian da–yup’v–a (PFV–spoil–INFIN), also in some dialects the conjunction magram ‘but’
loans mainly concrete nouns, reflecting Georgian cultural influence, e.g. Bezhta leybi ‘mattress’ (< leibi), topi ‘rifle’ (< topi)

11. Avar (esp. adjacent Antsukh (ánsuq) dialect)
traditional lingua franca of western Daghestan
at least fm 1860s, especially fm 1920s
also conduit for loans fm other languages
Avar origin vocabulary
nouns from wide range of semantic domains, e.g. beyten ‘wedding’, waccal ‘cousin’
includes many adjectives, e.g. bercinab ‘beautiful’ (in Avar –b is a class III [neuter] suffix; in Bezhta, the word is invariable)
verbs are borrowed only as parts of LVCs, in the infinitive form in –(i)zi, e.g. bak’arzi –owal ‘gather (tr.)’, bak’arzi –aqal ‘gather (itr.)’
some doublets, e.g. indigenous nuzo (archaic), Avar loan bet’erhan ‘husband’ phonetic adaptation r > y in older loans, e.g. beyten ‘wedding’ < Avar bertin but: bercinab, bak’arzi, bet’erhan – later loans

12. Russian
currently major source of loans into Bezhta (and other Tsezic languages)
including transmission of international vocabulary
even some words with a native counterpart, e.g. peč ‘stove’ (< Russian peć) — Bezhta has at least one indigenous word and two loans from Georgian for different kinds of stoves
also loan adjectives, adverbs, expressions
verbs only in LVCs, in form of infinitive (−t < Russian −t’)
a few loans have phonetic adaptation
   e.g. Russian fabrika ‘factory’ appears as Bezhta fabrika and fabrika
some early loans transmitted through Georgian have ejectives in Bezhta–Hunzib
   e.g. p’alt’o < Georgian p’alt’o < Russian pal’to ‘raincoat’
   but not Tsez–Hinuq paltu
13. Arabic
Islamicization from C18
Arabic loans largely restricted to nouns (incl. those used as parts of LVCs)
also some religious locutions, e.g. *wallah* ‘by God’
one conjunction, *amma* ‘but’, which is frequent in Bezhta
semantic domains:
  religion: *allah* ‘God’, *şäyt’an* ‘devil’
  abstract nouns: *häl* ‘state, condition’, *pikro* ‘thought’, *zaman* ‘time’, *saʔat* (also *şäʔat*) ‘hour’
  administration: *ahlo* ‘people’
  arts: *kalam* ‘word’, *surat* ‘picture’
  science, education: *dars* ‘lesson’ (often now replaced by Russian, e.g. older *müʔälim*, newer *ʊʃtel* ‘teacher’ < *učitel´*)
  names of weekdays, Islamic months, people’s names
stronger phonetic adaptation in Bezhta Proper
  e.g. *häl* < *ḥaːl*, *müʔälim* < *muʕaːlim*
  but without shift r > y, e.g. *dars* < *dars*

14. Results for chapter 1: The physical world
Indigenous: 51
Avar loans: 9 of which 2 from Arabic, 1 from Turkic
Georgian loans: 2 of which 1 questionable
Russian: 3 of which 1 from Latin, 1 from Greek
No equivalent: 7
Total 72

Details:
Avar
  *raλ* ‘land’ < *raλ*
  *č’ink’ilHi* ‘island’ < *č’ink’iHi*
  *raʔal* ‘shore’ < *raʔal*
  *raʔad* ‘sea’ < *raʔad*
  *halagab* ‘rough (of sea)’ < *halagab*
  *kanHi* ‘the light’ < *kanHi*
  *dunaal (duniya)* ‘the world’ < *dunya* (duniya) (< Arabic *dunyā*)
  *hawa* ‘air’ < *hawa* (< Arabic *hawā*)
  *maydan* ‘valley’ < *maydan* ‘square, plain’ (< Turkic, cf. Kumyk *maydan* ‘square’); also indigenous *ko*, less common
Georgian
  *yaḍri* ‘embers’ < *yaḍari*
  *maʃola* ‘rainbow’ < *aʃuni* ‘belt’
  questionable; Tsez has both meanings ‘rainbow’, ‘belt’
Russian
  *picka* ‘match’ < *spička*
  *akean* ‘ocean’ < *ocean* (< Greek *oceanós*)
  *kantinent* ‘mainland’ < *kontinent* (< Latin *continentum*)

References
**MEANING**
- **LWT code:** 5
- **Suggested code:** 1.31
- **Missing word form:** [water]
- **Language name:** Swahili
- **Typical context:** The inhabitants get their water from the nuvi.
- **Semantic field:** PHYSICAL WORLD IN ITS LARGER ASPECTS
- **Semantic category:** Noun

**MEANING-Words/RELATIONSHIP**
- **Word to meaning relationship (W25):**

<table>
<thead>
<tr>
<th>Word Form (W25)</th>
<th>Meaning (W25)</th>
<th>Other (W25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[W25]</td>
<td>water</td>
<td></td>
</tr>
</tbody>
</table>

**FOR ALL WORDS**
- **Record #:** 10
- **Default lang.:** Swahili
- **Analyzability (W9):** Unanalyzable
- **Borrowed (W9):** No evidence for borrowing

**ONLY FOR LOAN WORDS, CALQUIES, AND CREATED ON LOAN BASIS**
- **Word form (W14):** [water]
- **Donor language (W15):** English
- **Meaning (W16):** [water]
- **Location (W18):** [water]
- **Reference (W21):** [water]

**List of words related to meaning: the water**

<table>
<thead>
<tr>
<th>Word Form (W25)</th>
<th>Meaning (W25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[W25]</td>
<td>water</td>
</tr>
</tbody>
</table>

**List of meanings related to word: the water**

<table>
<thead>
<tr>
<th>Word Form (W25)</th>
<th>Meaning (W25)</th>
<th>Other (W25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[W25]</td>
<td>water</td>
<td></td>
</tr>
<tr>
<td><strong>MEANING</strong></td>
<td>LWT code</td>
<td>Suggested code</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td>LWT label</td>
<td>(W2)</td>
<td></td>
</tr>
<tr>
<td>LWT label</td>
<td>(W3)</td>
<td></td>
</tr>
<tr>
<td>Meaning Description (M3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical Domain (M4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic category (M5)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>MEANING-WORD RELATIONSHIP</strong></th>
<th>Word to meaning relationship (W25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comments on meaning-word relationship (W5A)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>FOR ALL WORDS</strong></th>
<th>Record #</th>
<th>Default long</th>
<th>Suggestion</th>
<th>remove word</th>
<th>set to an existing word</th>
<th>Original Source (W9)</th>
<th>Comments On Word Form (W5B)</th>
<th>Comments On Word Form (W5A)</th>
<th>Comments On Word Form (W5C)</th>
<th>Comments On Word Form (W5D)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Form (W2)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meaning (W6)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Analyzability (W7)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Borrowed (W9)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>populate age table</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS</strong></th>
<th>No source word (neither W14 nor W14A is available)</th>
<th>Immediate source word</th>
<th>Earliest known source word</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word form (W14)</strong></td>
<td>Avor (W16A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Donor Language (W16)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Meaning (W16)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effect (W18)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Environmental Balance (W22)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Reference (W21)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other Comments (W22)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>List of words related to meaning</strong></th>
<th>The sea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Form (W5)</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>List of meanings related to word: redad</strong></th>
<th>The sea</th>
</tr>
</thead>
</table>
### Meaning-Word Relationship

**Word to meaning relationship (W25)**

<table>
<thead>
<tr>
<th>Word Form (W2)</th>
<th>Meaning (W8)</th>
<th>Comments (on Word Form (W38))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunnal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Word to meaning relationship (W25)**

<table>
<thead>
<tr>
<th>Word Form (W2)</th>
<th>Meaning (W8)</th>
<th>Comments (on Word Form (W38))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunnal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Only for Loanwords, Calques, and Created on Loan Basis

<table>
<thead>
<tr>
<th>Word Form (W14A)</th>
<th>Word Form (W14)</th>
<th>Source Word (W17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunnal</td>
<td>Dunnal</td>
<td></td>
</tr>
</tbody>
</table>

**Donor Language (W15) | Arabic**

**Meaning (W16) | world**

**Reference (W17) | Highly integrated**

**Other Comments (W22) |**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>pliska</td>
<td>match</td>
<td>Unanalyzable</td>
<td>Borrowed (W9)</td>
<td>4. Clearly borrowed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**MEANING**

- **LWT code:** 1.59
- **Suggested code:**
- **remove meaning:**
- **set to an existing meaning:**

**Word:** the rainbow

**Morphemes:**
- **Bozlvla** (Bozlvla)

**Meaning:**
- **Bozlvla:** (Bozlvla)

**Typical Context:**
- **When the storm, a rainbow appeared in the sky.**

**Comments:**
- **language name:** Bozlvla
- **language:**
- **manner:**
- **for missing word form:**
- **comments:** (on missing word form)

**MEANING-WORD RELATIONSHIP**

**Word to meaning relationship (W25)**

### FOR ALL WORDS

- **Record #** 31
- **Default lang:**
- **remove word:**
- **set to an existing word:**

**Word Form (W2):** milola

**Meaning (W6):**
- **rainbow**

**Analyzeability (W7):** Unanalyzable

**Morpheme By Morpheme Gloss (W2):**
- **2:** Perhaps borrowed

**Record Status:** INCOMPLETE

### ONLY FOR LOANWORDS, CALQUES, AND CREATED ON LOAN BASIS

<table>
<thead>
<tr>
<th>Immediate source word</th>
<th>Earliest known source word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word form (W14A)</td>
<td>Word form (W14A)</td>
</tr>
<tr>
<td>Donor Language (W15A)</td>
<td>Donor Language (W15A)</td>
</tr>
<tr>
<td>Meaning (W16A)</td>
<td>Meaning (W16A)</td>
</tr>
<tr>
<td>Effort (W18)</td>
<td>Effort (W18)</td>
</tr>
<tr>
<td>Environmental Science (W20)</td>
<td>Environmental Science (W20)</td>
</tr>
<tr>
<td>Reference (W21)</td>
<td>Reference (W21)</td>
</tr>
<tr>
<td>Other Comments (W22)</td>
<td>Other Comments (W22)</td>
</tr>
</tbody>
</table>

**The Tsez cognate has the meanings `bell' and `rainbow'.**

---

**List of words related to meaning: the rainbow**

- **Word Form (W2):** milola

**List of meanings related to word: the rainbow**

- **LWT label (W2):**

---

10