Stress and phrasal prosody in Udmurt: initial results

Lena Borise & Ekaterina Georgieva
Research Institute for Linguistics, Budapest
{lена.borise/ekaterina.georgieva}@nytud.hu

5th Workshop on the languages of Volga-Kama Sprachbund (online)
December 11-12, 2020
Aims and claims

Object of study: **prosodic realization of stress** in Udmurt (Uralic, Permic), in the contexts of minimal pairs consisting of:

(i) indicative verbs (PRS.3SG) and  
(ii) imperative verbs (IMP.2SG/PL).

- According to the literature:
  - indicative verbs are stressed on the **final** syllable,  
  - imperative verbs are stressed on the **initial** syllable

- Based on novel experimental evidence, we show that the realization of stress in Udmurt is cued by **vowel duration** and alignment **with** \( f_0 \) (pitch) **targets**.
• We also show that these results align with our earlier results for (iii) negated indicative verbs (PRS.PST.2/3PL), also stressed on the initial syllable.

• Additional support for the current analysis comes from the interaction of stress and the spread of high tone H, associated with the interrogative particle a.
Roadmap

• Background

• Previous research on Udmurt

• Current study

• Outlook
Background: stress placement in Uralic

The Uralic languages vary widely with respect to their default stress assignment. Stress in Uralic may be:

- fixed on the **initial** syllable (Balto-Finnic, Saami, Hungarian, Mansi (other than Tavdina Mansi));
- fixed on the **penult** (Hill Mari);
- fixed on the **ultima** (Udmurt, Eastern Mari dialects, Tavdina Mansi);
- may form various types of stress systems governed by **phonological and/or morphological factors** (Moksha Mordvin, Meadow Mari, Komi Permyak, Komi Yazva, Southern Khanty dialects);
- may be ‘**absolutely free**’ – “the same word may be pronounced as if it may carry stress on different syllables” (Erzya Mordvin, Komi Zyrian, Konda Khanty) (Lytkin 1964: 234; 1970).
Background: stress placement in Uralic

No consensus on the stress properties of Proto-Uralic:

- fixed initial stress (Itkonen 1955; Collinder 1960);
- mobile stress linked to vowel harmony processes (Setälä 1896; Szinnyei 1922);
- ‘absolutely free’ stress (Steinitz 1964; Lytkin 1970).
Background: stress in Udmurt

Stress targets the **final syllable** of a word (Yemelyanov 1927: 14; Lytkin & Tepliashina 1962: 47; Csúcs 1990: 29):

(1) *turná*

    mow.PRS.3SG

    ‘s/he mows’ (Csúcs 1990: 29)

➢ This might be the result of contact with Tatar (Tarakanov 1975), which, like most Turkic languages, has fixed final stress (Zakiev 1993: 98).
Exceptions to stress-finality

Initial stress with:


• Negative verbs: em čášete ‘we didn’t make noise’ (Lytkin & Tepliashina 1962: 47; Csúcs 1990: 29; Winkler 2011: 22; Edygarova 2015).

Exceptions to stress-finality

Variable stress placement:

• Some derived words, e.g. indefinite pronouns, váńmy ~ vańmý ‘we all’; “depending on utterance type” (Lytkin & Tepliashina 1962: 48; Csúcs 1990: 29; Winkler 2001: 11; Winkler 2011: 22)
  o Possible relevant factors: “emotional context of an utterance and/or logical emphasis” (Alatyrev 1983)

• Dialectal variation (Lytkin & Tepliashina 1962: 49; Winkler 2011: 23). Cf. Middle Cheptsa (Northern) and Beserman Udmurt (Yemelyanov 1927: 14; Tepliashina 1970; Karpova 2005: 48–51)
Acoustic correlates of stress

- **Duration**: stressed syllables/vowels may be greater in duration than unstressed ones
- **Intensity**: stressed vowels typically have greater intensity than unstressed ones
- **Pitch/\(f_0\)**: stressed vowels may have particular \(f_0\) properties (high or low)
- **Vowel quality**: there may be language-specific requirements for quality of stressed (or unstressed) vowels.

Most languages rely on more than one of these to cue stress.

- The *interplay between cues* can be complex: in Spanish, stress is perceived if cued by \(f_0\) and *duration* or \(f_0\) and *intensity*, but not any one cue alone (Llisterrri et al. 2003)
Languages with fixed stress: research challenges

• Languages with fixed stress have a **weaker acoustic expression of stress** (Rigault 1970; Cutler 2005); cf. also Fónagy (1966) on Hungarian, Janota (1967) on Czech, Jassem (1962) and Dogil (1999) on Polish.

• Speakers of languages with fixed stress have weaker intuitions about stress placement and have a hard time learning languages with variable stress placement – so called ‘**stress-deafness**’ (Dupoux & Peperkamp 2002; Dupoux, Peperkamp & Sebastián-Gallés 2001; Dupoux et al. 1997; Peperkamp & Dupoux 2002; Peperkamp, Vendelin & Dupoux 2010).

• Part of a more general pattern: **predictable** prosodic phenomena receive a **weaker** acoustic implementation (Aylett & Turk 2004; Gahl 2008; Watson, Arnold & Tanenhaus 2008; Turk 2010; Athanasopoulou, Vogel & Dolutian 2017; Turnbull 2017).
The phonetics of stress in Udmurt: earlier results

Impressionistic observations:
- A stressed [final] syllable is somewhat (ca. 1.5 times) longer than an unstressed one (Alatyrev 1983; Winkler 2001: 10; 2011: 22).

Preliminary experimental results:
- Lytkin & Tepliashina (1962): (1 speaker, 5 observations): stressed [final] syllables are marked by greater duration (ca. 1.5 times longer than the unstressed). Greater intensity and f₀ may also be used.
- Baitchura (1973): (4 speakers, no. of observations unknown): initial syllables are marked by greater intensity and f₀; final ones are 1.5-2 longer than the initial ones ⇒ interprets these findings as evidence for initial stress.
The phonetics of stress in Udmurt: earlier results

Instrumental study by Vakhrushev & Denisov (1992):

Set-up:

- di- and trisyllabic words
- minimal pairs:
  - 3SG indicative
  - 2SG/PL imperative verbs
- 2 speakers
The phonetics of stress in Udmurt: earlier results

**Duration** in minimal pairs:

- stressed syllables, whether non-final or final, have greater duration than their unstressed counterparts
  - BUT interfering factor: final lengthening
The phonetics of stress in Udmurt: earlier results

Solution to the problem of final lengthening: between-word as opposed to within-word comparison of syllable duration

Indicatives: \( \sigma_{\text{initial}} \) \( (\sigma) \) \( \sigma_{\text{final}} \)

\[ \updownarrow \]

Imperatives: \( \sigma_{\text{initial}} \) \( (\sigma) \) \( \sigma_{\text{final}} \)

➢ Also used in our study
The phonetics of stress in Udmurt: earlier results

\( f_0 (\text{pitch}) \):

- disyllables: mean \( f_0 \) on the second syllable lower than on the first syllable

- BUT: declarative intonation also typically has falling \( f_0 \)
The phonetics of stress in Udmurt: earlier results

**Intensity:**

- No results reported for di- and trisyllables.
- In minimal pairs, the fall in intensity is steeper when the initial syllable is stressed, flattens out when the final one is stressed:
- Intensity decreases from left to right within prosodic word (independent of stress).
The phonetics of stress in Udmurt: earlier results

Overall:

• Duration is a cue for stress in minimal pairs
• Final stressed syllables are associated with low $f_0$ values; picture unclear for initial stress
• A flattened (as opposed to falling) intensity curve throughout the word, in minimal pairs, is indicative of final stress in minimal pairs
**Instrumental study**

**Research questions:**

1. What acoustic means does the realization of Udmurt stress rely on?
2. What are the prosodic properties of cliticization in Udmurt?

**Tasks:**

Task 1: nouns, adjectives, postpositions → Q1
Task 2: minimal pairs of indicative and imperative verbs → Q1
Task 3: negative and affirmative verbs, different clitics, questions → Q2
Instrumental study

Stimuli:

- minimal pairs of verbs:
  - indicative (PRS.3SG): final stress
  - imperative (IMP.2SG/PL) initial stress
- di- and trisyllabic
- CV syllable shape
- low ([+low]) vs. mid/high ([−low]) vowels
  - NB: verbs with high vowels contained a mid vowel in the second syllable, for morphological reasons, e.g. bude ‘grow’
- Total: n=172, randomized

(Vakhrushev & Denisov 1992:27)
Instrumental study

- All items were collected from Kirillova’s (2008) dictionary and checked with a native speaker who did not participate in the experiment
- All items were embedded in **carrier phrases**
- Information structure: backgrounded vs. focused
Carrier phrases

(1a) I ‘x!’ word said, but ‘y!’ word – not.  focused; imperative

(1b) I ‘x’ word said, but ‘y’ word – not.  focused; indicative

(2a) I ‘x!’ word quietly/slowly said, but loudly/quickly – not.  backgrounded imperative

(2b) I ‘x’ word quietly/slowly said, but loudly/quickly – not.  backgrounded indicative
Examples

vala ‘understand’, dysyllabic, [+low]

1a focused; imperative

Mon “vala!” kilez veraj, a “gaža!” kilez ej.
I understand.IMP.2SG word.ACC said but respect.IMP.2SG word.ACC didn’t
‘I said the word ‘vala!’, and not the word ‘gaža!’.’

1b focused; indicative

Mon “vala” kilez veraj, a “gaža” kilez ej.
I understand.PRS.3SG word.ACC said but respect.PRS.3SG word.ACC didn’t
‘I said the word ‘vala’, and not the word ‘gaža’.‘
Examples

2a backgrounded; imperative

Mon “vala!” kilez šip veraj, zol ej.
I understand.IMP.2SG word.ACC quietly said loudly didn’t
‘I said the word ‘vala!’ quietly, not loudly.’

2b backgrounded; indicative

Mon “vala” kilez šip veraj, zol ej.
I understand.IMP.2SG word.ACC quietly said loudly didn’t
‘I said the word ‘vala’ quietly, not loudly.’
Method

- 6 native speakers (5 f, 1 m; age range 20–40; Central and Northern dialects)
- Target sentences were displayed on the screen one at a time
- Experiments took place in June and October 2020 at RIL
- Recordings were made in a quiet room with a head-worn microphone
Processing

- The sound files were manually annotated in Praat
- Duration and intensity (not reported here) were measured for each vowel
- $f_0$ measurements were made at 10 fixed points per vowel
- Statistical analysis was performed in R
Results

- In indicative and imperative verbs, both di- and trisyllabic, the stressed syllable has greater duration than its counterpart in the verb of the other type.

  \[
  \begin{align*}
  \text{Indicatives:} & \quad \sigma_{\text{initial}} & \quad (\sigma) & \quad \dot{\sigma}_{\text{final}} \\
  \quad & \uparrow & \quad & \uparrow \\
  \text{Imperatives:} & \quad \dot{\sigma}_{\text{initial}} & \quad (\sigma) & \quad \sigma_{\text{final}}
  \end{align*}
  \]

- Imperative verbs are marked by high \( f_0 \) values on the stressed syllable (or the juncture between the stressed and post-tonic syllables).

- Indicative verbs may carry a high or a low \( f_0 \) target on the stressed syllable.

- Speakers may preferentially rely on \( f_0 \) or duration to mark stress.
## Averaged results

### Duration, [-low] vowels

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th>Backgrounded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
<td>Imperatives</td>
</tr>
<tr>
<td><strong>Disyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>68.08952 (16.85068)</td>
<td>116.7755 *** (27.9411)</td>
<td>61.87278 (17.31549)</td>
<td>112.6036 *** (29.11423)</td>
</tr>
<tr>
<td>Final</td>
<td>86.09696 (16.93234)</td>
<td>74.94284 *** (15.70377)</td>
<td>79.87921 (10.07818)</td>
<td>77.71154 (12.11575)</td>
</tr>
<tr>
<td><strong>Trisyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial</td>
<td>57.97592 (14.76401)</td>
<td>94.98495 *** (17.38961)</td>
<td>56.46343 (15.92255)</td>
<td>87.45692 *** (26.50306)</td>
</tr>
<tr>
<td>Final</td>
<td>70.81858 (11.48189)</td>
<td>72.0464 * (10.89306)</td>
<td>68.51495 (8.460949)</td>
<td>64.020114 (10.04928)</td>
</tr>
</tbody>
</table>

- duration measurements are made in ms
- in brackets: standard deviation
- significance values: LME model (Duration ~ Verb type + (1|Speaker) + (1|Item))
## Averaged results

### Duration, [+low] vowels

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th></th>
<th>Backgrounded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
<td>Imperatives</td>
<td></td>
</tr>
<tr>
<td><strong>Disyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>90.39888</td>
<td>121.8908***</td>
<td>87.09756</td>
<td>121.77***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.61061)</td>
<td>(32.42607)</td>
<td>(22.31177)</td>
<td>(39.06998)</td>
<td></td>
</tr>
<tr>
<td>final</td>
<td>113.159</td>
<td>92.79653***</td>
<td>107.3795</td>
<td>94.47824**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(47.97227)</td>
<td>(37.07969)</td>
<td>(46.88834)</td>
<td>(30.51985)</td>
<td></td>
</tr>
<tr>
<td><strong>Trisyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>74.88686</td>
<td>106.0115***</td>
<td>76.24305</td>
<td>112.3703***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(16.53813)</td>
<td>(30.80553)</td>
<td>(15.8824)</td>
<td>(25.91348)</td>
<td></td>
</tr>
<tr>
<td>final</td>
<td>94.83139</td>
<td>87.17771</td>
<td>101.3746</td>
<td>82.85396**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(39.03924)</td>
<td>(29.88872)</td>
<td>(41.51658)</td>
<td>(28.78845)</td>
<td></td>
</tr>
</tbody>
</table>
Averaged results: $f_0$, disyllables
Averaged results: $f_0$, trisyllables
Interspeaker variation

- Speakers in our sample varied with respect to which acoustic cue (duration or \( f_0 \)) they mainly used to cue stress.
  
  - three speakers relied mainly on duration
  - two speakers relied mainly on \( f_0 \)
  - one speaker utilized both
S6, illustration: indicative
S6, illustration: imperative
S6, $f_0$, disyllables
S6, $f_0$, trisyllables
### S6, Duration

[-low] vowels

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th>Backgrounded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
<td>Imperatives</td>
</tr>
<tr>
<td><strong>Disyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>72.2781 (22.97583)</td>
<td>81.31244 (14.59496)</td>
<td>68.28303 (19.20435)</td>
<td>89.98299 *** (15.26356)</td>
</tr>
<tr>
<td>final</td>
<td>81.90462 (20.30196)</td>
<td>77.21266 (17.47291)</td>
<td>82.84707 (16.69295)</td>
<td>80.32969 (19.00552)</td>
</tr>
<tr>
<td><strong>Trisyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>89.05209 (19.78594)</td>
<td>85.03692 (27.16686)</td>
<td>79.92946 (15.23259)</td>
<td>77.75648 (14.78279)</td>
</tr>
<tr>
<td>final</td>
<td>74.27141 (21.24324)</td>
<td>76.86357 (16.53693)</td>
<td>84.57082 (16.73069)</td>
<td>77.34087 (17.67608)</td>
</tr>
</tbody>
</table>
## S6, Duration

### [+low] vowels

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th>Backgrounded</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
</tr>
<tr>
<td>Disyllabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>85.773335 (24.71298)</td>
<td>84.7936 (16.46009)</td>
<td>78.27464 (16.99829)</td>
</tr>
<tr>
<td>final</td>
<td>74.24918 (18.02073)</td>
<td>69.14234 (17.52394)</td>
<td>76.40547 (14.29793)</td>
</tr>
<tr>
<td>Trisyllabic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>85.069 (12.76227)</td>
<td>79.10147 (12.91829)</td>
<td>76.99177 (13.13932)</td>
</tr>
<tr>
<td>final</td>
<td>76.79666 (32.40021)</td>
<td>83.19697 (19.50729)</td>
<td>79.28523 (12.23287)</td>
</tr>
</tbody>
</table>
S5, illustration: indicative

<table>
<thead>
<tr>
<th>Time (s)</th>
<th>vala</th>
<th>kilez</th>
<th>veraj</th>
<th>a</th>
<th>gazha</th>
<th>kilez</th>
<th>ej</th>
<th>I</th>
<th>get.IND</th>
<th>word</th>
<th>said</th>
<th>butrespect.IND</th>
<th>word</th>
<th>didn’t</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
S5, illustration: imperative
S5, $f_0$, disyllables

![Graph showing the relationship between $f_0$ and normalized time for different conditions and speakers.](image)

- $\sigma_1$ [-low]
- $\sigma_2$ [-low]
- $\sigma_1$ [+low]
- $\sigma_2$ [+low]

Legend:
- imp, F
- ind, F
- imp, non-F
- ind, non-F
S5, $f_0$, trisyllables

\[
\begin{align*}
\sigma_1 [-\text{low}] & \quad \sigma_2 [-\text{low}] & \quad \sigma_3 [-\text{low}] \\
\sigma_1 [+\text{low}] & \quad \sigma_2 [+\text{low}] & \quad \sigma_3 [+\text{low}] \\
\end{align*}
\]
## S5, Duration

### [-low] vowels

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th>Backgrounded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
<td>Imperatives</td>
</tr>
<tr>
<td>Disyllabic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>86.50864</td>
<td>128.494 ***</td>
<td>88.89531</td>
<td>143.6463 ***</td>
</tr>
<tr>
<td></td>
<td>(15.53248)</td>
<td>(27.54516)</td>
<td>(30.36566)</td>
<td>(26.7408)</td>
</tr>
<tr>
<td>final</td>
<td>132.0594</td>
<td>107.7936</td>
<td>133.5819</td>
<td>130.8547</td>
</tr>
<tr>
<td></td>
<td>(37.89683)</td>
<td>(37.58826)</td>
<td>(44.33091)</td>
<td>(31.23878)</td>
</tr>
<tr>
<td>Trisyllabic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>84.67487</td>
<td>109.1967 **</td>
<td>78.36445</td>
<td>116.7359 ***</td>
</tr>
<tr>
<td></td>
<td>(20.44461)</td>
<td>(28.17197)</td>
<td>(11.52493)</td>
<td>(20.88054)</td>
</tr>
<tr>
<td>final</td>
<td>123.9002</td>
<td>90.84126</td>
<td>111.2852</td>
<td>101.9282</td>
</tr>
<tr>
<td></td>
<td>(49.5809)</td>
<td>(31.39045)</td>
<td>(41.88558)</td>
<td>(42.13521)</td>
</tr>
</tbody>
</table>
### S5, Duration

**[+low] vowels**

<table>
<thead>
<tr>
<th></th>
<th>Focused</th>
<th></th>
<th>Backgrounded</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Indicatives</td>
<td>Imperatives</td>
<td>Indicatives</td>
<td>Imperatives</td>
</tr>
<tr>
<td><strong>Disyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>100.9941 (16.62947)</td>
<td>158.6109 *** (14.43941)</td>
<td>100.5229 (16.68486)</td>
<td>160.1965 *** (24.58089)</td>
</tr>
<tr>
<td>final</td>
<td>149.8482 (47.64543)</td>
<td>83.67448 *** (20.01257)</td>
<td>114.0783 (17.64767)</td>
<td>88.96737 * (24.71387)</td>
</tr>
<tr>
<td><strong>Trisyllabic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>initial</td>
<td>82.35457 (12.97789)</td>
<td>135.7533 * (49.85971)</td>
<td>89.21813 (16.60879)</td>
<td>126.8499 *** (22.88382)</td>
</tr>
<tr>
<td>final</td>
<td>112.2145 (51.95583)</td>
<td>74.24305 * (22.28727)</td>
<td>119.2101 (50.53844)</td>
<td>63.8078 * (7.64331)</td>
</tr>
</tbody>
</table>
Conclusions

- In indicative and imperative verbs, the stressed syllable has greater duration than its counterpart in the verb of the other type
- Imperative verbs are marked by high f₀ values on the stressed syllable (or the juncture between the stressed and post-tonic syllables)
- Indicative verbs may carry a high or a low f₀ target on the stressed syllable
- Speakers may preferentially rely on f₀ or duration to mark stress

➢ Implications for perception of stress: different acoustic cues can have the same phonological interpretation
Outlook

- Task 1: nouns, adjectives, postpositions → Q1
- Task 3: negative and affirmative verbs, different clitics, questions → Q2

Other acoustic cues to consider:

- intensity
- vowel quality/formant structure
Thank you for your attention!
Тау кылзййськемды понна!

We are deeply indebted to Udmurt consultants (Yulia Speshilova, Elena Rodionova, Valeria Fedorova, Anna Kadrova, Lukeria Shikhova, Vladislav Volkov, and one anonymous informant), as well as the research assistants who annotated the recordings. This research has been supported by the research projects “Nominal Structures in Uralic Languages” (NKFIH FK 125206), “Implications of endangered Uralic languages for syntactic theory and the history of Hungarian” (NKFIH KKP 129921), and “How prosody shapes word order: an integrated interface-based approach to the post-verbal domain in OV languages” (NKFIH K 135958).
References


