

**Интернет-ресурс
«Малые языки России»:
способы визуализации
языковых данных**

Лингвистический форум 2020:
Язык и искусственный интеллект



Лаборатория исследования и сохранения малых языков, Институт языкознания РАН



Ольга Анатольевна
Казакевич



Константин
Поливанов



Наталья
Сердобольская



Заира
Халилова



Елена
Будьянская



Анастасия
Евстигнеева



Карина
Мищенко



Дарья
Мордашова



Софья
Покровская



Евгения
Ренковская

Ресурс “Малые языки России”

- ❖ пилотная версия сайта: <http://minlang.iling-ran.ru>
- ❖ общедоступный информационный интернет-ресурс, содержащий материалы о структуре и функционировании миноритарных языков России и их локальных вариантов

Основные задачи:

- ❖ систематизация имеющихся сведений по конкретным языкам
- ❖ регулярное пополнение этих сведений актуальной информацией

Шаблон страницы языка: содержание

- ❖ генетическая принадлежность
- ❖ территория распространения
- ❖ языковые контакты
- ❖ сферы функционирования языка
- ❖ динамика развития языковой ситуации
- ❖ структура языка
- ❖ сведения об исследователях и исследовательских центрах
- ❖ список основных публикаций и электронных ресурсов
- ❖ информация о наличии административной и общественной поддержки

Шаблон страницы языка: техническая реализация

- ❖ Пользовательский интерфейс строится с использованием шаблонов разметки html5, каскадных таблиц стилей CSS3, а также js и jquery библиотек.
- ❖ В частности, используются библиотеки leaflet для создания интерактивных карт, а также JavaScript библиотека D3 для создания динамических интерактивных визуализаций и схем.

Шаблон страницы языка: проблемы визуализации

- ❖ представление генеалогии
- ❖ представление территории распространения языка
- ❖ представление сфер функционирования языка
- ❖ представление фонетических данных
- ❖ представление динамики развития языковой ситуации

Представление генеалогии

Основные задачи:

- ❖ визуализация полной картины генеалогии (от уровня диалекта до уровня языковой семьи) с возможностью перехода на конкретные фрагменты
- ❖ удобство использования
- ❖ адекватное отображение данных

Решение:

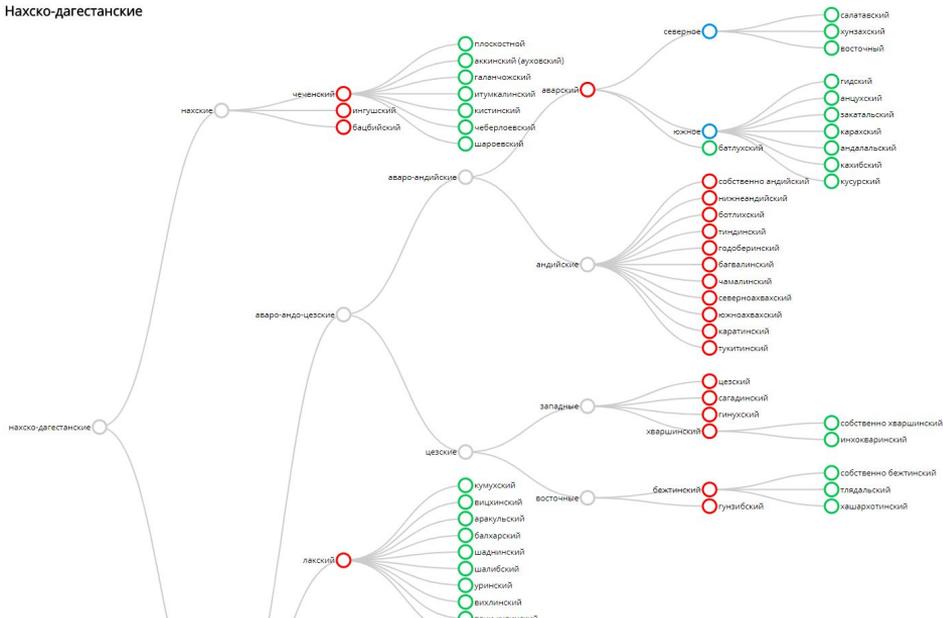
интерактивная схема

Представление генеалогии: решение

- ❖ Традиционное представление генетической принадлежности в древовидной форме.
- ❖ Имеется возможность сворачивать и разворачивать узлы дерева для фокусировки внимания на его отдельных фрагментах.

Представление генеалогии: визуализация

Нахско-дагестанские



■ язык
■ наречие
■ диалект

Представление генеалогии: код

```
var nodeEnter = node.enter().append("g")
  .attr("class", "node")
  .attr("transform", function(d) { return "translate(" + source.y0 + "," + source.x0 + ")"; })
  .on("click", click);

nodeEnter.append("circle")
  .attr("r", 1e-6)
  .style("stroke", function(d) { return d.status; });

nodeEnter.append("circle")
  .attr("r", 1e-6)
  .style("fill", function(d) { return d._children ? "lightsteelblue" : "#fff"; });

nodeEnter.append("text")
  .attr("x", function(d) { return d.children || d._children ? -13 : 13; })
  .attr("dy", ".35em")
  .attr("text-anchor", function(d) { return d.children || d._children ? "end" : "start"; })
  .text(function(d) { return d.name; })
  .style("fill-opacity", 1e-6)
  .append("tspan")
  .attr("x", function(d) { return d.children || d._children ? -13 : 13; })
  .attr("dy", "1.5em")
  .attr("text-anchor", function(d) { return d.children || d._children ? "end" : "start"; })
  .text(function(d) { return d.description; })
  .style("fill-opacity", "1");
```

Представление территории распространения языка

Основные задачи:

- ❖ визуализация распространения в целом, а также отдельно по населенным пунктам
- ❖ удобство использования
- ❖ адекватное отображение данных

Решение:

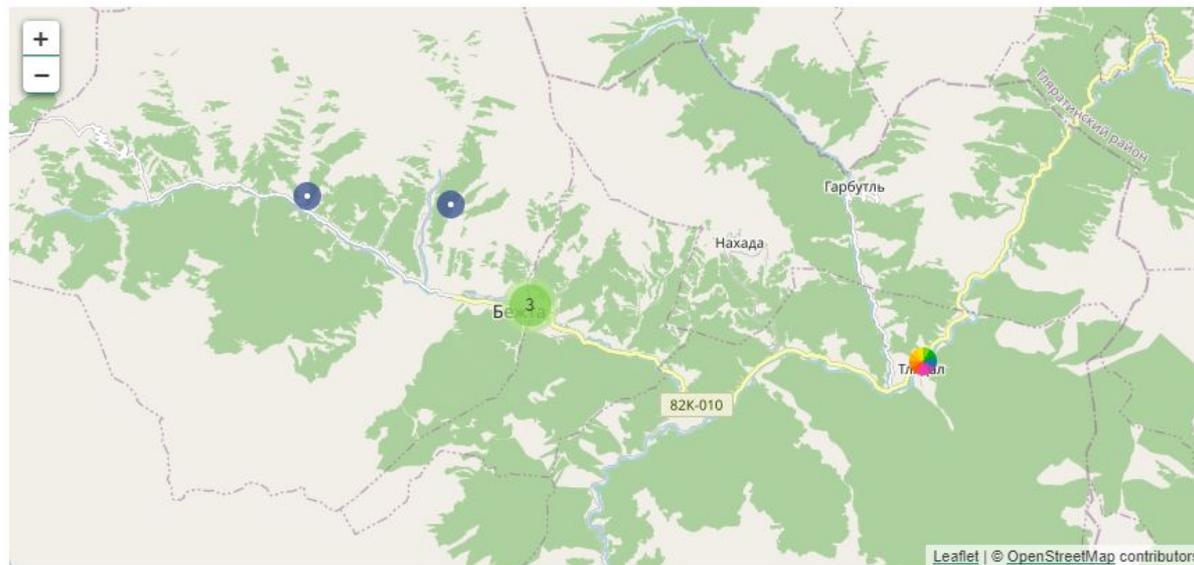
интерактивная карта

Представление территории распространения языка: решение

- ❖ Интерактивная карта, на которой в соответствующих локациях маркерами отмечены язык и диалекты.
- ❖ Для наглядности каждому диалекту присваивается свой цвет; значения используемых цветов даются в легенде под картой.
- ❖ Для населенных пунктов, где локализовано более одного диалекта, был введен особый маркер, состоящий из нескольких цветов.
- ❖ При клике на маркер во всплывающем окне возникает перечень диалектов.

Представление территории распространения языка: решение

Распространение



● собственно бежтинский диалект

● хашархотинский диалект

● тлядальский диалект

● более одного диалекта

Представление территории распространения языка: решение

```
var coords2 = [];  
var markerscluster2 = L.markerClusterGroup();  
for (var i = 0; i < langPoints2.length; i++) {  
  var a = langPoints2[i];  
  var title = a[2];  
  if(a[3] == '1') {  
    var langPoint2 = L.marker(new L.LatLng(a[0], a[1]), { icon: g1 });  
  }  
  if(a[3] == '2') {  
    var langPoint2 = L.marker(new L.LatLng(a[0], a[1]), { icon: g2 });  
  }  
}
```

Представление сфер функционирования языка

Основные задачи:

- ❖ визуализация полной картины сфер функционирования в рамках одного пользовательского экрана
- ❖ удобство и скорость получения информации

Решение:

интерактивная таблица

Представление сфер функционирования языка: решение

- ❖ Интерактивная таблица, в которой представлены основные сферы функционирования языка (16 сфер).
- ❖ В ячейках таблицы представлены названия сфер функционирования с отметкой, используется ли язык в соответствующей сфере.
- ❖ При клике на ячейку открывается всплывающее окно с развернутой информацией об особенностях функционирования языка в данной сфере.

Представление сфер функционирования языка: решение

Юридический статус	Письменность	Литературная норма	Сферы функционирования
Семейно-бытовое общение	Образование	СМИ	
Культура	Наука	Фольклор	
Литература	Религия	Законодательство	
Административная деятельность	Судопроизводство	Промышленное производство	
Сельскохозяйственное производство	Торговля и сфера обслуживания	Транспорт	
Интернет			

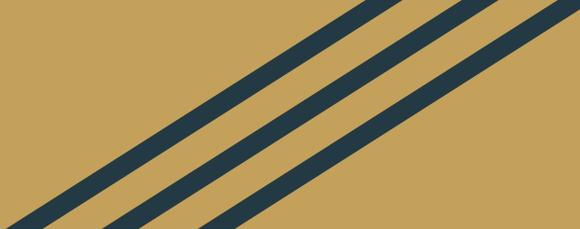
Представление сфер функционирования языка: решение

Функционирование языка

Юридический статус	Письменность	Литературная норма	Сферы функционирования
Семейно-бытовое общение	Образование	СМИ	
Культура	Наука	Фольклор	
Литература			

Религия

Некоторые обряды и песнопения исполняются на бежтинском языке, однако в целом используются арабский и аварский языки.



Представление фонетических данных

Основные задачи:

- ❖ визуализация фонетических и фонологических данных в единой и доступной форме
- ❖ удобство и скорость получения информации
- ❖ объективность и репрезентативность

Решение:

графические представления

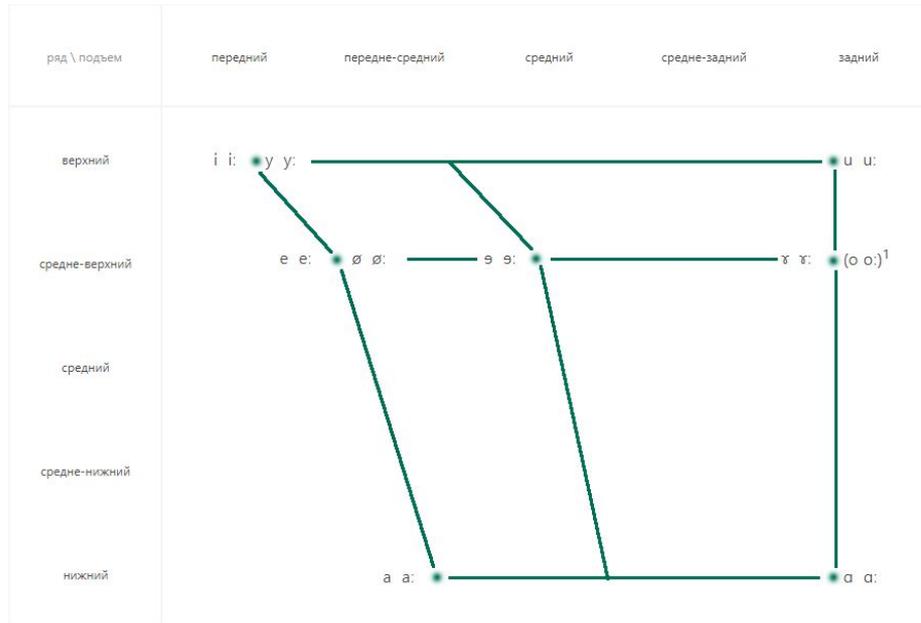


Представление фонетических данных: решение

- ❖ Графические представления, в которых одновременно отражаются данные фонетики и фонологии.
- ❖ Звуковой строй каждого языка записывается в системе МФА.
- ❖ Структуры представления строятся с учетом особенностей русскоязычной терминологии.
- ❖ Предпринимаются попытки стандартизировать представление с целью выработки оптимального варианта структуры, релевантного для любого языка России.
- ❖ Всё это призвано упростить сопоставление звуковых систем разных языков.

Представление фонетических данных: решение

Вокализм



Представление фонетических данных: решение

```
<div class="voc-data">
  i<span class="sl-3
  <span class="v-el-2">y<span class="sl-3
  <span class="v-el-3">u<span class="sl-3
  <span class="v-el-4">e<span class="sl-3
  <span class="v-el-5">ø<span class="sl-3
  <span class="v-el-6">ə<span class="sl-3
  <span class="v-el-7">ʌ<span class="sl-3
  <span class="v-el-8 with-index">(o<span
  <span class="v-el-9">a<span class="sl-3
  <span class="v-el-10">a<span class="sl-
</div>
```

Представление фонетических данных: решение

Консонантизм

		Место образования									
		губные		передненёбные				средненёбные		задненёбные	
				дентальные/ альвеолярные		постальвеолярные					
Способ образования	взрывные	p	b	t	d				tʃ	k	g
	носовые		m		n				ɲ		ŋ
	вибранты				r						
	аффрикаты								tʃ		
	фрикативные		β	s	z				ʃ		x
	боковые аппроксиманты								ɹ		

Цветовые обозначения:

глухость

аспирация

глоттализация

звонкость

Представление фонетических данных: решение

```
<tr>
  <td colspan="2" rowspan="2" class="top-sublabel">губные</td>
  <td colspan="6" class="top-sublabel">передненёбные</td>
  <td colspan="2" rowspan="2" class="top-sublabel">средненёбные</td>
  <td colspan="2" rowspan="2" class="top-sublabel">задненёбные</td>
</tr>
<tr>
  <td colspan="2" class="top-sublabel">дентальные/альвеолярные</td>
  <td colspan="2" class="top-sublabel">палатализованные</td>
  <td colspan="2" class="top-sublabel">ретрофлексные</td>
</tr>
<tr>
  <td rowspan="6" class="left-label" style="min-width:54px"><span>
  <td class="left-sublabel">взрывные</td>
  <td class="is feature-1">p</td><td class="is feature-4">b</td>
  <td class="is feature-1">t</td><td class="is feature-4">d</td>
  <td class="is feature-1">t<sup>j</sup></td><td class="is feature-4">
  <td class="is feature-1"></td><td class="is feature-4"></td>
  <td class="is feature-1"></td><td class="is feature-4"></td>
  <td class="is feature-1">k</td><td class="is feature-4">g</td>
</tr>
```



Представление динамики развития языковой ситуации

Основные задачи:

визуализация общей картины динамики, позволяющая отслеживать изменение во времени различных количественных показателей

Решение:

диаграмма



Представление динамики развития языковой ситуации: решение

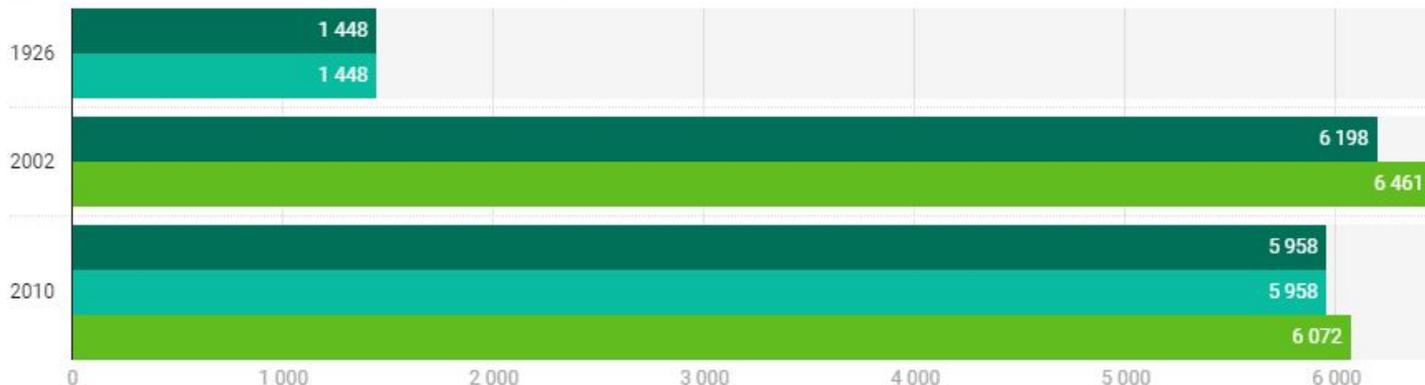
- ❖ Диаграмма, отображающая изменение величины этнической группы, численности членов группы, считающих этнический язык родным, и численности тех, кто сообщил о владении этим языком.
- ❖ Основой для построения диаграммы служат данные переписей населения разных лет, в перспективе - данные местных администраций, а также оценки исследователей.

Представление динамики развития языковой ситуации: решение

Бежтинский язык

Динамика развития языковой ситуации. Данные переписи 1926, 2002, 2010 годов.

■ величина этнической группы ■ признание родным ■ владение

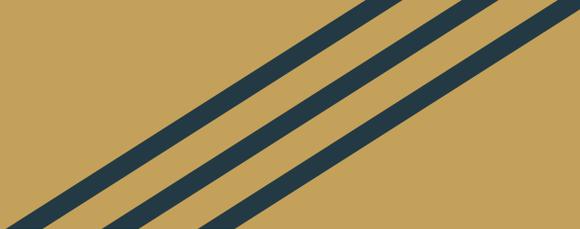


отсутствие параметра означает отсутствие данных

[Скачать данные](#) • Создано с помощью [Datawrapper](#)

Представление динамики развития языковой ситуации: решение

```
▶ <div class="dw-chart chart vis-height-fixed theme-datwrapper-data vis-d3-bars-grouped theme-undefined" id="__svelte-dw">...</div>
▶ <script>...</script>
  <script src="../../lib/load-polyfills.74159e63.js"></script>
  <script src="../../lib/vendor/jquery.min.0f23f82b.js"></script>
  <script src="../../lib/vendor/globalize.min.66aa7a35.js"></script>
  <script src="../../lib/vendor/globalize.culture.ru-RU.f1d7f1ee.js"></script>
  <script src="../../lib/vendor/underscore.min.be3cacb1.js"></script>
  <script src="../../lib/vendor/dw-2.0.min.b33eeecf.js"></script>
  <script src="../../lib/vis/d3-bars-grouped.369c7ba8.js"></script>
  <script>window.parent.postMessage('https://datwrapper.dwcdn.net/lib/vis/d3-bars-grouped.369c7ba8.js')
  </script>
  <script src="../../lib/main.d8d66830.js"></script>
  <script src="../../lib/blocks/subscriptions.chart-blocks.255d0b80.js"></script>
  
```



Демонстрация сайта



Спасибо за внимание!