

# Avar IME User Manual

---

## Welcome to Avar IME!

The **Avar IME (Input Method Editor)** allows you to type in the Avar language (авар маъл) using your standard QWERTY keyboard. This manual will guide you through installation, basic usage, and advanced features.

**This is the first public version of Avar IME, and it is still being refined.** The IME works for everyday typing, but rough edges remain -- some digraph variants may be missing, predictions may occasionally rank words oddly, and the candidate window can feel less polished than in mature commercial IMEs. Updates are released regularly with improvements to prediction quality, new digraph variants, expanded corpora, better personalization, and bug fixes.

**Please send feedback.** If a particular word doesn't predict the way you expect, if a digraph is missing from the candidate window, if something feels slow or awkward, or if you simply want to suggest a feature -- write to [kebed.zagidov@upf.edu](mailto:kebed.zagidov@upf.edu). Concrete examples ("I typed X, expected Y, got Z") are the single most useful kind of feedback and directly shape the next release.

---

## Table of Contents

1. What is Avar IME?
2. Installation
3. Getting Started
4. Basic Typing
5. Using the Candidate Window
6. Special Characters
7. Word Predictions
8. Keyboard Layout Reference
9. Tips and Tricks
10. Frequently Asked Questions

---

## What is Avar IME?

Avar IME is a keyboard input method that enables you to type in the **Avar language** using a standard QWERTY keyboard. The Avar language uses Cyrillic script with special characters including:

- **Palochka (|)** — A vertical bar character used in many Caucasian languages
- **Digraphs** — Two-character combinations like **гъ, гь, г|, къ, кь, к|, хъ, хь, х|, л|, т|, ц|, ч|**

## How It Works

1. You type on your QWERTY keyboard
2. Keys are automatically mapped to Russian Cyrillic letters (ЙЦУКЕН layout)
3. For letters with Avar variants, a **candidate window** appears showing all options
4. You press **1-9** to select the variant you want
5. The selected character is inserted into your document

## Why this design?

Avar has many more distinct sounds than there are keys on a QWERTY keyboard. There were several reasonable ways to handle the mismatch; this IME chose three deliberate trade-offs:

1. **Picking from candidates beats memorising key combinations.** Aside from the basic ЙЦУКЕН letters, an Avar keyboard has to fit 14+ digraphs (гъ, гь, гI, къ, кь, кI, лъ, лI, тI, хъ, хь, xI, цI, чI) plus geminates like КIкI and Хьхь. A "one key combination per digraph" design would mean learning a dozen Alt/Ctrl-based shortcuts you'd never use anywhere else. Letting you type the base consonant and pick the variant by number is faster to learn and lighter on muscle memory. Combined with the smart prediction bar (which usually offers the right form before you finish typing), the candidate window mostly fades into the background.
2. **There aren't enough keys for a separate Avar layout.** Even with a unique layout, the digraphs would still require some kind of compositional mechanism. The candidate-window approach generalises cleanly: when new letter variants get added in the future, they just appear as additional numbered options instead of demanding a re-learned layout.
3. **Same layout as Russian.** This is my personal choice, and I might well be wrong about it -- but the reasoning was this. Most Avar speakers today read, write, and type in Russian as part of daily life, so muscle memory for the ЙЦУКЕН layout is something we already have. Inventing a separate Cyrillic key arrangement just for Avar seemed to me to mean extra effort to learn the IME, with no obvious linguistic benefit in return. A similar pattern, as far as I can tell, holds in other cases too: people writing in many Latin-alphabet languages (English, Spanish, French, Italian, and so on) share essentially the same QWERTY layout, with only small per-language tweaks; even Chinese Pinyin input runs on top of plain English QWERTY because its users already know it. Avar IME follows the same logic: it reuses the standard Russian ЙЦУКЕН layout you already know; the only difference is that a handful of consonant keys open a small candidate window, and predictions help you bypass that most of the time. You can switch between Avar and Russian input with Win+Space without re-orienting your fingers. If you see this differently or have a better idea, please tell me -- I'm open to changing my mind.

My hope was an input system that respects what users already know (Russian ЙЦУКЕН + the muscle memory that goes with it), uses prediction and selection where the alphabet exceeds the keyboard, and stays out of the way during ordinary typing. Whether that hope was reached is for you to judge -- feedback is genuinely welcome.

---

## Installation

### System Requirements

- Windows 10 or Windows 11 (64-bit)

## Installation Steps

1. **Download** the installer: AvarIME\_Setup\_<version>.exe.
2. **Double-click the installer.** Windows will automatically prompt for administrator privileges (UAC) — click **Yes**. No "right-click → Run as administrator" step is needed.
3. Walk through the installation wizard (Next → Install → Finish). At the end, leave the "**Open the User Manual now**" checkbox checked.
4. **Sign out and sign back in** to Windows (required for the IME to register properly). A full restart works too.
5. **Add the IME to your language settings:**
  - Open **Settings** → **Time & Language** → **Language & Region**
  - Click on your language (e.g., "English" or "Russian")
  - Click **Language options**
  - Under "Keyboards," click **Add a keyboard**
  - Select "**Avar IME**" from the list
6. **Verify installation:**
  - You should now see "Avar IME" in your language bar (system tray)
  - You can switch to it using **Windows + Space** or by clicking the language indicator

---

## Getting Started

### Activating Avar IME

1. Click the **language indicator** in your system tray (bottom-right corner)
2. Select "**Avar IME**" from the list

**OR**

Press **Windows + Space** to cycle through your installed keyboards until you see "Avar IME"

---

## Basic Typing

### Standard Letters

Most letters work just like a Russian keyboard. Simply type the key and the corresponding Cyrillic letter appears:

QWERTY Key	Cyrillic Letter	Example
A	ф	Type a → get ф

S	ы	Type s → get <b>ы</b>
D	в	Type d → get <b>в</b>
F	а	Type f → get <b>а</b>

See the Keyboard Layout Reference section for the complete mapping.

## Uppercase Letters

Hold **Shift** while typing to get uppercase letters:

- Type f → **a** (lowercase)
- Type Shift + f → **A** (uppercase)

## Punctuation

Most punctuation keys work as expected:

- **Space**, **Enter**, **Backspace** work normally
- **Period (.)**, **Comma (,)**, and other punctuation insert their standard characters

## Using the Candidate Window

### What is the Candidate Window?

When you type certain letters that have Avar-specific variants, a **candidate window** appears showing all available options. This is the core feature of Avar IME!

### How to Use It

1. **Type a letter** that has variants (like `u` for `ɾ`)
2. **A window appears** showing numbered options:

- ```

1. ɾ
2. ɾʘ
3. ɾʘʘ
4. ɾɪ

```

3. **Press the number** (1-4) to select the variant you want
4. **The selected character is inserted** into your document

### Example: Typing "ɾɪ"

1. Type `u` (maps to `ɾ`)
2. Candidate window appears with 4 options
3. Press `4` to select `ɾɪ`
4. The text `ɾɪ` is inserted

## Letters with Candidates

The following letters show a candidate window:

| Key | Base Letter | Candidates Available    |
|-----|-------------|-------------------------|
| U   | г           | г, ГЪ, Гь, ГІ           |
| R   | к           | к, КЪ, Кь, КІ, Кк, КІКІ |
| [   | х           | х, ХЪ, Хь, ХІ, ХХ, ХьХь |
| K   | л           | л, ЛЪ, ЛІ, ЛЪЛЪ         |
| N   | т           | т, ТІ                   |
| W   | ц           | ц, ЦІ, ЦЦ, ЦІЦІ         |
| X   | ч           | ч, ЧІ, ЧЧ, ЧІЧІ         |
| C   | с           | с, СС                   |

## Uppercase Variants

When you hold **Shift**, the candidate window shows **uppercase** variants:

- Type `Shift + u` → Candidate window shows: **Г, ГЪ, Гь, ГІ**
- Press `4` → Get **ГІ**

---

## Special Characters

### Palochka (І)

The **Palochka** is a special character (І) used extensively in Avar. To type it:

**Press: Alt + ` (backtick)**

- The backtick key is usually located in the top-left corner of your keyboard, below **Esc**
- Hold **Alt** and press ` (backtick)
- You get: **І**

### Yo (ё/Ё)

To type the Russian letter **ё**:

- Press ` (backtick) → **ё** (lowercase)
- Press **Shift + `** → **Ё** (uppercase)

---

## Word Predictions

Avar IME has a **two-mode prediction engine** trained on real Avar corpora (~700K sentences) plus the full Avarlab project lexicon. It uses both the word you're typing AND the words that came before it.

## Two prediction modes

**1. Prefix mode** — predicts the rest of the word you're currently typing.

Start typing a word. After at least one character, a prediction bar appears with completions of what you've typed. Press a number key (1-7) to insert the full word.

Example: type `v f u (маг)` → the bar shows `магIарул, магIарулъ, магIалъ, ...` in order of how likely they are *given the words you typed before*.

**2. Next-word mode** — predicts the *next* word after you finish one.

When you press **Space** or pick a prediction, the bar refreshes with predictions of what word usually comes next. Press 1-7 again to insert one without typing anything.

Example: after `гъабулеб` (with trailing space), the bar shows `буго, букIана, вуго, ...` because those are the words that most commonly follow `гъабулеб` in Avar text.

## Context-aware ranking

Predictions are not just based on word frequency. They're scored with **Stupid Backoff** (the standard technique used by Google Search autocomplete and Gboard) over three levels of context:

- **Trigram** — looks at the last *two* words you typed
- **Bigram** — looks at the last *one* word
- **Unigram** — overall word frequency in Avar

The engine prefers the strongest available signal. So:

- After `буго гъеб` the trigram tier kicks in and predicts words that historically followed `буго гъеб` specifically — not just any common Avar word.
- After a rare word (or at the start of a sentence), the engine quietly falls back to bigram, then unigram.

The same engine also re-ranks prefix predictions by context. So if you've typed `буго гъеб ц` and there are 30 Avar words starting with `ц`, the ones that historically followed `буго гъеб` will be at the top of the bar — not just the most common `ц...` words overall.

## Personal usage learning — the IME adapts to you

**The IME learns from your typing.** Every word you commit (by picking from the prediction bar, or by typing + space/enter) is added to a personal-usage tally. Words you've typed before get a ranking boost the next time they're candidates, so the suggestions you see become increasingly tuned to *your* vocabulary and writing style.

How it works in practice:

- The first time you type a rare word — say a personal name, a place, or specialized terminology — it may not appear at the top of the predictions.
- Pick it once (or finish typing it) and the IME remembers.

- The next time you type a prefix that matches, the word will rank higher because your personal-usage count contributes a positive bonus to the score.
- The bonus is **logarithmic**: typing the word once gives a small bump; typing it 5-10 times makes it competitive with mid-frequency corpus words; typing it dozens of times makes it dominant for any matching prefix. The personalization stays a *signal*, not a hammer — it never completely overrides the corpus.

The bonus combines with the context-aware (trigram/bigram) ranking, so a word you type often *in a particular context* will rank highest exactly where you'd want it.

### Current limitation: session-only learning (v2.0)

In this version, the IME's personal learning **resets when you sign out of Windows or restart your computer**. This is a known limitation -- the persistent personal-learning file is on the roadmap and will land in a future update (see roadmap section). While you're signed in and typing, learning works fully: the engine adapts within your session.

When persistent learning ships, no action will be needed on your part — your typing patterns will simply start accumulating across sessions.

### Picking candidates

- Press **1-7** to insert a numbered candidate.
- Or click the candidate with your mouse.
- After picking, the bar refreshes immediately with next-word predictions for the word you just picked. Picking another candidate inserts it without any typing.
- **Tab** inserts the first (top-ranked) prediction.
- **Esc** hides the bar without picking.

### Fuzzy digraph completion — type the easy way, get the right word

**This is one of the most powerful features of Avar IME — read carefully.**

You don't have to remember which digraph to use, or stop to pick it from the candidate window. **Just type the plain base letters and the predictions will offer all relevant digraph variants of what you typed.**

#### Real examples

**Example 1 — type кебед (plain к, no digraph picked):**

1 кебебедов    2 кебебед    3 кебебедовас    4 кебебедасул  
5 кебебедова    6 кебебедица    7 кебебедил

You typed к but the IME silently tried къ, къь, кI against the dictionary too. Since real Avar words start with кебебед (not кебед), those are what surfaces. Pick one with the number key — the IME inserts the *full* digraph form (including the ъ you never typed).

**Example 2 — type вачана:**

1 вачлана 2 вачланани 3 вачланаха

Same idea: even when no digraph is needed, the IME searches related word forms in parallel and ranks them by context, so the closest real Avar word still surfaces -- here `вачлана` (with the `л` you didn't type) outranks the literal `вачана` because it's the real word.

### When this fires

- Any time you're typing a prefix that *could* contain digraphs of consonants in your text. The IME tries: `к` → `к`, `кЪ`, `кь`, `кI`, `г` → `г`, `гЪ`, `гь`, `гI`, `х` → `х`, `хЪ`, `хь`, `хI`, `т` → `т`, `тI`, `ц` → `ц`, `цI`, `ч` → `ч`, `чI`, `л` → `л`, `лЪ`, `лI`.
- The variants and the literal-prefix matches are scored together — so if `къебед` is more common than `кебед` in Avar (it is), `къебед` wins, even though you typed plain `к`.

### What this means for you

- **You almost never need to use the candidate window** for digraphs in everyday typing. Just type plain consonants and pick from the prediction bar.
- The candidate window is still there when you want to type a *specific* digraph form character-by-character (e.g. when typing a word the IME doesn't know).
- This is the difference between typing 5 keys and 1 number-pick to get `къебедов` (the easy way) vs. typing 8 keys with two candidate-window selections (the slow way).

**Tip:** Watch the prediction bar from the moment you type the first character. If your target word is already there, just pick it — don't keep typing.

### When predictions appear

- **Prefix mode:** after you type at least 1 character of a word.
- **Next-word mode:** immediately after you commit a word (by pressing Space, Enter, or picking from the bar).
- If no useful predictions exist (very rare context), the bar hides itself.

### Disabling predictions

If you prefer to type without predictions:

1. Navigate to the installation folder.
2. Open `config.ini` in a text editor.
3. Change `EnablePredictions=true` to `EnablePredictions=false`.
4. Save and restart your applications.

---

## Keyboard Layout Reference

### Complete Key Mapping

This table shows how QWERTY keys map to Cyrillic letters (based on the Russian ЙЦУКЕН layout):

## Letter Keys

| QWERTY | Cyrillic | QWERTY | Cyrillic | QWERTY | Cyrillic |
|--------|----------|--------|----------|--------|----------|
| Q      | й        | W      | ц        | E      | у        |
| R      | к        | T      | е        | Y      | н        |
| U      | г        | I      | ш        | O      | щ        |
| P      | з        | A      | ф        | S      | ы        |
| D      | в        | F      | а        | G      | п        |
| H      | р        | J      | о        | K      | л        |
| L      | д        | Z      | я        | X      | ч        |
| C      | с        | V      | м        | B      | и        |
| N      | т        | M      | ь        |        |          |

## Special Keys

| QWERTY    | Cyrillic | Notes                |
|-----------|----------|----------------------|
| ;         | ж        | Semicolon            |
| ,         | б        | Comma                |
| .         | ю        | Period               |
| [         | х        | Left bracket         |
| ]         | ъ        | Right bracket        |
| '         | э        | Apostrophe           |
| `         | ё        | Backtick (lowercase) |
| Shift + ` | Ё        | Backtick (uppercase) |
| Alt + `   | І        | Palochka             |

## Visual Layout Diagram

|       |   |   |   |   |   |   |   |   |   |   |   |   |       |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|-------|
| ё/І   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | - | = | Back  |
| Tab   | й | ц | у | к | е | н | г | ш | щ | з |   |   | Enter |
| Caps  | ф | ы | в | а | п | р | о | л | д | ж | э | х |       |
| Shift | я | ч | с | м | и | т | ь | б | ю |   |   |   | Shift |

**Note:** Letters in **bold** in the candidate tables (U, R, , K, N, W, X, C) will show a candidate window with

Avar variants.

## Tips and Tricks

### 1. Learn the Common Candidates

The most frequently used Avar-specific characters are:

- **ɢ** (ejective g) — Type `u` then `4`
- **кɢ** (ejective k) — Type `r` then `4`
- **хɢ** (ejective kh) — Type `l` then `4`
- **лɢ** (ejective l) — Type `k` then `2`
- **тɢ** (ejective t) — Type `n` then `2`
- **цɢ** (ejective ts) — Type `w` then `2`
- **чɢ** (ejective ch) — Type `x` then `2`

### 2. Use Word Predictions — and don't worry about digraphs while typing

Don't type every letter! After 1-3 letters, check if the word you want appears in predictions. This can save significant time.

**You can skip digraphs entirely.** Type plain `к, ɢ, х, т, ц, ч, л` even when the real word has `кɢ, ɢɢ, хɢ, тɢ, цɢ, чɢ, лɢ` — the IME tries all digraph variants of what you typed and surfaces real Avar words. Type `кебед` → pick `кɢебедов` from the bar. You never touched `ɢ`.

See [Fuzzy digraph completion for the full explanation with examples.

### 3. Practice the Palochka

The Palochka (l) is used very frequently in Avar. Practice the **Alt + `** combination until it becomes second nature.

### 4. Switch Keyboards Quickly

Use **Windows + Space** to quickly switch between Avar IME and your regular keyboard. This is useful when typing mixed Avar and English text.

### 5. Candidate Window Shortcuts

- Press **Esc** to close the candidate window without selecting anything
- The candidate window automatically closes if you press **Space** or **Enter**

### 6. Check Your Typing

If you're not seeing the characters you expect:

1. Make sure **Avar IME is active** (check the language indicator in your system tray)
2. Remember that some letters show a **candidate window** — you must select a number
3. Check if **Caps Lock** is on (it affects capitalization)

---

## Frequently Asked Questions

### Q: Can I use Avar IME in all applications?

**A:** Avar IME works in most modern Windows applications that support TSF (Text Services Framework), including:

- Notepad, WordPad
- Microsoft Word, Excel, PowerPoint
- Web browsers (Chrome, Edge, Firefox)
- Visual Studio Code
- Most email clients

Some older applications may not support TSF input methods.

### Q: Can I customize the candidate lists?

**A:** Yes! Advanced users can edit the `config.ini` file in the installation folder to customize which candidates appear for each letter.

### Q: How do I type numbers and punctuation?

**A:** Numbers (0-9) and most punctuation work exactly as on a standard keyboard. The only exceptions are:

- , (comma) → **б**
- . (period) → **ю**
- ; (semicolon) → **ж**
- ' (apostrophe) → **э**
- ` (backtick) → **ë** (or **l** with Alt)

### Q: Can I use Avar IME on other operating systems?

**A:** Currently, Avar IME is only available for Windows 10 and Windows 11. Versions for macOS and Linux are not available at this time.

### Q: How do I uninstall Avar IME?

**A:**

1. Open Windows Settings → Time & Language → Language & Region
2. Click on your language → Language options
3. Under "Keyboards," find "Avar IME" and click **Remove**
4. Run `unregister.ps1` from the installation folder as Administrator
5. Restart your computer

### Q: Is my typing data collected?

**A:** No. Avar IME runs entirely locally on your computer. No typing data is sent to any server or

collected in any way. The only file created is a local log file (`avar_ime.log`) for debugging purposes, which you can delete at any time.

## Q: Can I add my own words to the dictionary?

A: Yes! The user dictionary is stored at:

```
%APPDATA%\Avar IME\user_dictionary.txt
```

You can edit this file to add your own words. Each line should contain one word.

## Q: How smart are the predictions, exactly?

A: The IME's prediction engine is trained on:

- ~287K sentences of plaintext Avar corpus
- 277K sentences of monolingual POS-tagged corpus
- 111K sentences of combined Russian-Avar corpus
- 21K sentences of trilingual (Russian-Avar-English) corpus
- ~1.03M dictionary lexemes (headwords + inflected forms + phraseology) from Avarlab project

Total: about **600,000 word pairs** (bigrams) and **1.2 million word triples** (trigrams) that the engine considers when ranking. Predictions are scored with **Stupid Backoff** — the same technique used by Google Search and major mobile keyboards — which combines all three signals (single-word frequency, two-word context, three-word context) into one coherent ranking.

## Q: Why do I need to select candidates? Can't it just insert the character?

A: The candidate system is necessary because multiple Avar characters share the same base letter. For example, 'r' has four variants (r, r̄, r̅, r̆), and the IME can't know which one you want without asking. This is similar to how Chinese and Japanese IMEs work.

## Q: The candidate window appears in the wrong position

A: The candidate window tries to appear near your cursor. If it appears in the wrong location, this is usually due to the application not properly reporting cursor position. Try:

1. Clicking in the text field first
2. Using a different application
3. Restarting the application

---

## Thank you

Thank you for using Avar IME.

This input method editor was designed and developed by **Kebed Zagidov** as part of the Avarlab project.

**Contact:** kebed.zagidov@upf.edu

