

Scripture 101

Transliteration and the fold

🕒 2 min read · Updated 27 September 2026 · Verified against [6332e94](#) on 2026-09-25

The Roman reading aid beside every line, what it is and is not, and the phonetic fold that lets a seeker type a Gurmukhi word the way they hear it and still find it.

<https://docs.gurbanisoul.com/scripture/transliteration-and-the-fold/>

Explanation, not scripture — scholar review pending. These pages explain the scripture to engineers and students. Every quoted line is verbatim from the pinned database and cited by Ang; everything else is explanation, written by engineers and awaiting a Granthi's review. Where this page and the scripture differ, the scripture is right.

Beside every line the database keeps a **transliteration** — the line in Roman letters, built by the pipeline from the Gurmukhi:

ੴ ਸਤਿ ਨਾਮੁ ਕਰਤਾ ਪੁਰਖੁ ਨਿਰਭਉ ਨਿਰਵੈਰੁ ਅਕਾਲ ਮੂਰਤਿ ਅਜੂਨੀ ਸੈਭੰ ਗੁਰ ਪ੍ਰਸਾਦਿ ॥

*ik oankaar sat naam karataa purakh nirabhau niravair akaal moorat ajoonee saibhn
gur prasaad*

— Sri Guru Granth Sahib Ji · Ang 1

It is a reading aid for people who do not read Gurmukhi, and a search key. It is **not** scripture: it is never cited, never verified, never shown alone. Several Roman schemes exist and none is canonical; the pipeline's is consistent, and that consistency is what search needs.

What the fold is for

Seekers type a Gurmukhi word the way they hear it: *waheguru*, *vaahiguroo*, *wahegurooh*. A search that demanded the pipeline's exact spelling would miss most of them. The **fold** — `roman_norm` — reduces every spelling of a word to one key by five rules (letter substitutions, digraphs, voicing pairs, the y glide, vowels dropped and doubles collapsed), so *waheguru* and *vaahiguroo* both become `vhgr`. The database stores the folded form of every line's transliteration (`translit_norm`), and the query is folded the same way at search time.

The fold is deliberately lossy: it trades precision for recall, which is why the search waterfall tries it only after the exact tiers, and why the verification engine uses it to *find candidates* but scores them with the full transliteration before deciding a verdict.

One function, three homes

The same fold must run wherever a key is made: in sggs-data when the database is built, in the platform at query time, in the app on the device. Three copies, held byte-identical by 24,719 golden vectors — the whole story, with the function read from the source, is The Roman fold.

What to remember

- Show the transliteration *under* the Gurmukhi and in a lighter style; never in place of it.
- Never cite a transliteration; cite the Gurmukhi line by Ang.
- Never “improve” the transliteration scheme in one place: it is an index, and a change is a coordinated dataset release.
- A seeker’s misspelling is the fold’s job, not the reader’s: the first-letters and skeleton tiers exist for the same reason.

Read next: The Answer Protocol for engineers.