

Chapter 6: Languages and Scripts

Summary

South Asia is one of the most multilingual places on Earth. India's 2011 census counted 121 languages with at least 10,000 speakers each, and Pakistan, Nepal, Bangladesh, Sri Lanka, Afghanistan, Bhutan and the Maldives add many more. This chapter sorts those languages into five big families, explains the difference between a language and a script, and shows how most of the region's scripts grew from one ancestor, Brahmi. You'll learn how an abugida writes vowels, why Hindi and Urdu are so close in speech yet so different on paper, and what it means to grow up with a heritage language that you understand but don't always speak.

Concepts Covered

1. Language Family
2. Indo-Aryan Languages
3. Dravidian Languages
4. Tibeto-Burman Languages
5. Austroasiatic Languages
6. Iranian Languages
7. Eighth Schedule
8. Hindi-Urdu (Hindustani)
9. Bengali Language

10. Tamil Language
11. Brahmi Script
12. Abugida
13. Devanagari
14. Gurmukhi
15. Perso-Arabic Script
16. Code-Switching
17. Heritage Language

Prerequisites

- **Chapter 1: What Is South Asia?** for the eight countries and the map.
- **Chapter 2: Ancient Roots** helps with Ashoka and the Indus cities, but it isn't required.
- **Chapter 5: Faiths of South Asia** helps with the Sikh Gurus and with the languages of scripture.

So many ways to say hello



Hi, I'm Mora! Maybe you understand your grandparents perfectly but answer them in English, or maybe you can speak your family's language but can't read a single letter of it.

You're not alone, and this chapter is for you. By the end, you'll be able to look at a page of Bangla, Hindi or Tamil and see the family resemblance, even if you can't read it yet.

6.1 Three Names, Three Alphabets

It's the third meeting of the South Asian Student Association at Prairie Ridge High School, and the club is making name tags. Zara Hossain grabs a marker and writes her name in English, then adds the name of her family's language underneath in Bangla letters: বাংলা. Her friend Arjun Reddy, whose family speaks Telugu, writes తెలుగు on his. Simran Kaur Gill, the club treasurer, writes ਪੰਜਾਬੀ, the word for Punjabi, in the script her grandparents taught her.

Ben Lindqvist leans over the table and stares at the three name tags. Zara's letters hang from a straight line along the top. Arjun's look round, like little loops and curls. Simran's have a line on top too, but the shapes are different from Zara's. "Okay," Ben says. "Are these even related? They look like three totally different alphabets from three different planets."

Arjun laughs. "My grandpa says they're all cousins." Simran isn't so sure. Zara realizes she has never asked. She can read Bangla slowly, but she has no idea where its letters came from, or why Simran's look a little like hers but Arjun's don't. Are these scripts related, and if they are, why do they look so different?

6.2 A Language Is Not a Script

Start with something you already know. English, Spanish, French and Swedish are four different languages, but they are all written with the same 26-ish Latin letters (Swedish adds å, ä and ö). Ben's grandparents in Sweden could read the letters on a Spanish menu even if they couldn't understand the words. Now flip it around: Serbian is one language that is regularly written in two scripts, Cyrillic and Latin.

So there are two separate questions you can ask about any piece of writing:

- **What language is it?** A **language** is a system of words and grammar that people speak (or sign) to communicate.
- **What script is it written in?** A **script** is a set of written symbols used to write one or more languages.

South Asia is full of examples where the two don't line up neatly. Punjabi is usually written in the Gurmukhi script in India and in a Perso-Arabic script called Shahmukhi in Pakistan. Tamil and Telugu are completely different languages from Hindi, yet their scripts are distant cousins of the one used for Hindi. Keep the two questions apart, and much of this chapter becomes easy.

	Same language?	Same script?
English and Spanish	No	Yes (Latin)
Punjabi in India and Punjabi in Pakistan	Yes	No (Gurmukhi and Shahmukhi)
Hindi and Marathi	No	Yes (Devanagari)
Hindi and Urdu	Very close in everyday speech	No (Devanagari and Perso-Arabic)

Try this



Next time you see South Asian writing on a grocery bag, a temple or mosque sign, or a phone screen, ask yourself two questions: which script is this, and which language? I promise they aren't always the answer you'd guess.

6.3 The Five Big Language Families

Languages have family trees too. A **language family** is a group of languages that all descend from one older ancestor language. Linguists spot family members by comparing basic words (like numbers, body parts and "mother") and grammar. English and Swedish, for example, are both in the Germanic branch of the large Indo-European family, which is why Swedish *tre* and English *three* sound alike.

Five language families matter most in South Asia. India's 2011 census (the latest with language results as of 2026, since the next census counts people in early 2027) grouped its 121 counted languages by family, and the shares of people who speak each family as a mother tongue show how unevenly they're spread.

Family	Share of India's people, by mother tongue (2011 census)	Some languages in this family	Where in South Asia
Indo-Aryan (part of Indo-European)	78.05%	Hindi, Urdu, Bangla, Punjabi, Marathi, Gujarati, Odia, Nepali, Sindhi, Sinhala, Dhivehi	Across northern, central and western India, Pakistan, Bangladesh, Nepal, Sri Lanka, the Maldives
Dravidian	19.64%	Tamil, Telugu, Kannada,	Mostly southern India; also

Family	Share of India's people, by mother tongue (2011 census)	Some languages in this family	Where in South Asia
		Malayalam, Tulu, Gondi, Brahui	northern Sri Lanka; Brahui in Pakistan
Austroasiatic (Census: "Austro-Asiatic")	1.11%	Santali, Mundari, Ho, Khasi	Parts of eastern and northeastern India; also Bangladesh and Nepal
Tibeto-Burman (part of Sino-Tibetan)	1.01%	Manipuri (Meitei), Bodo, Mizo, Dzongkha, Newar, Sherpa, Tamang	The Himalaya and India's northeast; Bhutan, Nepal
Iranian (part of Indo-European)	tiny in India	Pashto, Dari (a form of Persian), Balochi	Afghanistan and western Pakistan

The census numbers above are for India only. In Afghanistan and Pakistan, Iranian languages are very large, and in Bhutan, Tibeto-Burman languages are the majority.

Indo-Aryan languages

Indo-Aryan Languages are the South Asian branch of the Indo-European family, the same giant family that includes English, Persian, Russian and Greek. They descend from Old Indo-Aryan, the language of the Vedas (an early form of Sanskrit), through everyday languages called Prakrits. Hindi, Urdu, Bangla, Punjabi, Marathi, Gujarati, Odia, Assamese, Nepali and Sindhi are all Indo-Aryan. So are two island languages you might not expect: Sinhala in Sri Lanka and Dhivehi in the Maldives, which are each other's closest relatives.

Indo-Aryan languages are spoken by the largest number of people in the region. In India's 2011 census, the biggest single mother tongue was Hindi, with 528 million speakers (43.63% of India's people), followed by Bangla (8.03%) and Marathi (6.86%). In Pakistan's 2023 census, the largest mother tongue was Punjabi, at about 37% of the population, while Urdu, the national language, was the mother tongue of 9.25%. In Nepal's 2021 census, Nepali was the mother tongue of 44.86% of people, among 124 mother tongues recorded.

Iranian languages

Iranian Languages are another branch of Indo-European, closely related to Indo-Aryan but separate from it. Persian (Farsi in Iran, Dari in Afghanistan) is the best-known one. In South Asia the big Iranian languages are Pashto and Dari, the two most widely used languages of Afghanistan, plus Pashto and Balochi in Pakistan. Pashto is the second-largest mother tongue in Pakistan, after Punjabi.

Dravidian languages

Dravidian Languages form a completely separate family, not related to Indo-European at all. The four big ones are Tamil, Telugu, Kannada and Malayalam, spoken by about 20% of India's people, mostly in the south. Tamil

is also an official language of Sri Lanka, and Tamil-speaking communities live in Singapore, Malaysia and around the world. A Dravidian language even survives far to the northwest: **Brahui** is spoken mainly in Balochistan, Pakistan, and is usually written in Arabic script.

Dravidian languages have borrowed many words from Sanskrit over the centuries, especially for religion and learning, and Indo-Aryan languages have borrowed from Dravidian ones too. But their grammar and core vocabulary come from a different ancestor. That matters for one of the most common myths about South Asia, which we'll get to in Common Misconceptions.

Tibeto-Burman languages

Tibeto-Burman Languages are the South Asian members of the huge Sino-Tibetan family, which also includes Chinese and Burmese. They're spoken mostly in the Himalaya and in India's northeast. India's 2011 census counted 66 different Tibeto-Burman languages, more than any other family, even though together they have only about 1% of India's speakers. Dzongkha, the national language of Bhutan, belongs here, as do Sherpa and Tamang in Nepal, Newar (spoken in the Kathmandu Valley), and Manipuri (Meitei) and Bodo in India. Nima Sherpa, whose family comes from Solukhumbu in Nepal, speaks Nepali with friends from Nepal, but his grandparents speak Sherpa at home.

Austroasiatic languages

Austroasiatic Languages are a family spread from eastern India to Vietnam and Cambodia (Vietnamese and Khmer belong to it). In South Asia the best known are the Munda languages, such as Santali, Mundari and Ho, spoken mainly by Adivasi (Indigenous) communities in Jharkhand, Odisha, West Bengal and nearby states, plus Khasi in Meghalaya. Santali also has speakers

in northern Bangladesh and eastern Nepal, and in 2003 it was added to India's Eighth Schedule. Santali has its own script, Ol Chiki, created in 1925 by Raghunath Murmu. Unlike most Indian scripts, Ol Chiki is a true alphabet, with full letters for vowels.

6.4 One Language or Two? Hindi-Urdu

Ayesha Qureshi's family is from Lahore, Pakistan, and speaks Urdu at home. Arjun's mom grew up in Hyderabad, where she learned Hindi in school. When Ayesha and Arjun's mom chat at a club potluck, they understand each other almost perfectly. Ask for water and both will say *pani*. But when they write the word down, it looks completely different:

	Hindi	Urdu
"water"	पानी	پانی
Name of the language	हिन्दी	اردو
Script	Devanagari, left to right	Perso-Arabic (Nastaliq style), right to left

Hindi-Urdu (Hindustani) is the name linguists use for the shared spoken language underneath both. Hindi (हिन्दी, *HIN-dee*) and Urdu (اردو, *OOR-doo*) are two standard forms of it. Their everyday grammar and most common words are the same, which is why speakers understand each other in casual conversation, and why one Bollywood song can be enjoyed in both Lahore and Lucknow. The base of both comes from Khariboli, the speech of the region around Delhi.

The differences show up in three places:

1. **Script.** Hindi is written in Devanagari, Urdu in a Perso-Arabic script, usually in the flowing, slanted style called Nastaliq.
2. **Formal vocabulary.** For formal, literary or technical words, standard Hindi tends to draw on Sanskrit, and standard Urdu on Persian and Arabic. "Book" is usually *kitab* in both, but the formal word for "knowledge" might be *gyan* in Hindi and *ilm* in Urdu.
3. **Identity and history.** Urdu is Pakistan's national language and is also an official language in several Indian states, including Telangana, Bihar and Uttar Pradesh. Hindi is an official language of India's central government. For many families, which name they use is tied to religion, region and history, especially since Partition (see [Chapter 4](#)).

Whose Story?

Is Hindi-Urdu one language or two? Many linguists answer "one language with two standard forms," because ordinary speakers understand each other so easily. Many people in India describe Hindi as its own language rooted in Sanskrit, and many people in Pakistan describe Urdu as its own language shaped by Persian, Arabic and Muslim literary culture. From the 1860s on, arguments over whether courts and schools in British India should use Hindi in Devanagari or Urdu in Perso-Arabic script became one of the issues dividing Hindus and Muslims. After 1947, Pakistan chose Urdu as its national language and India gave Hindi official status. Poets, film songwriters and ordinary families often ignore the line completely and mix both vocabularies. All of these answers describe something real, which is why the question still sparks debate.

6.5 Two Languages Up Close: Bangla and Tamil

The Bengali language

The **Bengali Language**, which its speakers call Bangla (বাংলা, *BAANG-laa*), is an Indo-Aryan language. It's the official and national language of Bangladesh and an official language of the Indian states of West Bengal and Tripura and of Assam's Barak Valley. In India's 2011 census, about 97 million people spoke it as their mother tongue, second only to Hindi, and Bangladesh has about 176 million people (2025, World Bank). In October 2024, India's government declared Bangla one of its "classical languages," a title for languages with very old, original literary traditions.

Bangla has a special place in the story of language rights. After 1947, Bangladesh was East Pakistan. In 1948 Pakistan's leader, Muhammad Ali Jinnah, declared in Dhaka that Urdu alone would be the state language, even though most people in East Pakistan spoke Bangla. Students in Dhaka protested. On **21 February 1952**, police opened fire on student protesters in Dhaka, and several were killed. That day is remembered in Bangladesh as Language Martyrs' Day. In 1956 Bangla became an official language of Pakistan alongside Urdu, and in 1999, after a proposal from Bangladesh, UNESCO declared 21 February **International Mother Language Day**, and it has been observed around the world since 2000. Zara's Nanu remembers walking barefoot to the Shaheed Minar, the martyrs' monument, on that day as a girl.

The Tamil language

The **Tamil Language** (தமிழ், *THUH-mizh*) is a Dravidian language spoken by about 69 million people in India in 2011 (5.70% of the population), mostly in Tamil Nadu. It is also one of the two official and national languages of Sri Lanka, alongside Sinhala, and one of the official languages of Singapore. Tamil literature goes back roughly two thousand years, to a collection of poetry known as the Sangam literature.

In 2004 India's government created a new category of "classical languages," and on 12 October 2004 Tamil became the first language to receive the title. By October 2024 there were 11 classical languages: Tamil, Sanskrit, Telugu, Kannada, Malayalam, Odia, Marathi, Pali, Prakrit, Assamese and Bangla.

Tamil is a good reminder that "South Asian languages" don't all come from Sanskrit. Tamil has its own grammar, its own ancient vocabulary and its own literary tradition, even though it has also exchanged words with Sanskrit for centuries.

6.6 Official Languages and the Eighth Schedule

Every country in South Asia has to decide which languages its government, schools and courts will use. Here is how the eight countries handle it, in short:

Country	Official or national language(s)	Script(s)
Afghanistan	Pashto and Dari (the main official languages)	Perso-Arabic
Bangladesh	Bangla	Bengali-Assamese
Bhutan	Dzongkha (ཇོང་ཁ, <i>DZONG-kha</i> , "language of the fortress")	Tibetan
India	Hindi and English for the central government; 22 languages in the Eighth Schedule	Many, including Devanagari

Country	Official or national language(s)	Script(s)
Maldives	Dhivehi (ދިވެހި, <i>DHI-veh-hi</i> , "of the islanders")	Thaana
Nepal	Nepali	Devanagari
Pakistan	Urdu (national) and English (official)	Perso-Arabic; Latin for English
Sri Lanka	Sinhala (සිංහල, <i>SING-huh-luh</i>) and Tamil, with English as a "link language"	Sinhala and Tamil scripts

India's system needs a closer look. Article 343 of the Constitution of India says that the official language of the Union (the central government) shall be Hindi in the Devanagari script. A 1963 law, the Official Languages Act, lets English keep being used alongside Hindi for official work. Each state also chooses its own official language or languages, so Tamil Nadu works in Tamil, West Bengal in Bangla, and so on.

The **Eighth Schedule** is a list attached to the Constitution of India that names the languages the government officially recognizes and promises to develop. When the Constitution was adopted it listed 14 languages. Amendments have added more, and today there are **22**: Assamese, Bangla (Bengali), Bodo, Dogri, Gujarati, Hindi, Kannada, Kashmiri, Konkani, Maithili, Malayalam, Manipuri, Marathi, Nepali, Odia, Punjabi, Sanskrit, Santali, Sindhi, Tamil, Telugu and Urdu. Being on the list matters, because these languages get official support for their development, teaching and literature. In 2011, 96.71% of India's people had one of these 22 as their mother tongue.

Many communities want their languages added. India's Ministry of Home Affairs has for years listed demands to include 38 more languages, among them Bhojpuri, and no new language has been added since 2003.

Notice that English is not in the Eighth Schedule, even though it is used for official purposes. And notice what's missing from the Constitution entirely: the phrase "national language."

The national language trap



You'll hear people say "Hindi is India's national language."

India's Constitution doesn't name any national language.

Hindi and English are the official languages of the central government, and 22 languages are recognized in the Eighth

Schedule. Pakistan is different: its constitution does name Urdu as the national language.

6.7 One Ancestor for Many Scripts: Brahmi

Now back to Ben's question at the name-tag table. Bangla, Telugu and Gurmukhi letters look like they come from three planets. In fact, all three descend from one ancient script.

The **Brahmi Script** is the ancestor of most scripts used in South Asia today, and of many in Southeast Asia and Tibet. It was in use by the 3rd century BCE. Its best-known early examples are the edicts of the emperor Ashoka, messages about ethics and good government carved on rocks and pillars across his empire in the mid-3rd century BCE. Over the centuries, people forgot how to read Brahmi. In the 1830s, James Prinsep, secretary of the Asiatic Society of Bengal in Kolkata (then Calcutta), deciphered it, which let historians read Ashoka's words again. (The Indus Valley cities in [Chapter 2](#)

had their own signs much earlier, but that Indus script still hasn't been deciphered.)

Brahmi changed slowly as scribes in different regions wrote it with different tools, on stone, palm leaves, birch bark and paper. One common explanation for why many southern scripts look rounder is that scribes there wrote on palm leaves, where long straight lines could split the leaf. Over about two thousand years, regional styles drifted so far apart that today they look unrelated. A simplified version of the family looks like this:

- **Northern branch** (through the Gupta script): Devanagari and its close cousin Gujarati; Gurmukhi (through the Sharada script of Kashmir); Bengali-Assamese and Odia (through an eastern script called Gaudi); and Tibetan, which Bhutan uses for Dzongkha.
- **Southern branches:** Kannada and Telugu (through the Kadamba script), and Tamil, Malayalam and Sinhala (through the Pallava and Grantha scripts).

This is the surprising part: script families and language families are different trees. Telugu and Tamil are Dravidian languages, but their scripts are Brahmi's descendants, just like Hindi's. Sinhala is an Indo-Aryan language, but its rounded script grew up under strong influence from south Indian scripts. That's exactly the language-versus-script difference from Section 6.2.

6.8 How an Abugida Works

Brahmi and its descendants share one clever design. In English, a vowel is a separate letter: *k* plus *a* makes *ka*. Brahmi-family scripts work differently. Each consonant letter already includes a short "a" sound, and you change or remove that vowel by adding a mark.

An **abugida** is a writing system in which each consonant letter carries a built-in vowel (usually a short "a"), and other vowels are shown by marks attached to the consonant. The built-in vowel is called the **inherent vowel**. A small mark called a **virama** (in Hindi, *halant*) "kills" the inherent vowel so you get a bare consonant.

Here is the syllable "ka" and its family in six scripts. Every character below is the real Unicode character, not a picture.

Script	ka	ki	ku	k (no vowel)
Brahmi	𑀓	𑀔	𑀕	𑀖
Devanagari (Hindi, Nepali, Marathi)	क	कि	कु	क्
Bengali-Assamese (Bangla)	ক	কি	কু	ক্
Gurmukhi (Punjabi)	ਕ	ਕਿ	ਕੁ	ਕ੍
Tamil	க	கி	கு	க்
Sinhala	ක	කි	කු	ක්

Look down the "ka" column first: one letter, no extra mark. Then look across a row. The base letter stays the same, and a small mark appears for each vowel. Notice something odd in Devanagari, Bangla and Gurmukhi: the "i" mark is written *before* the consonant even though you say it after. Readers learn to take in the whole cluster at once. One more detail: Gurmukhi has a virama character, but everyday Punjabi spelling almost never uses it.

If your browser shows any of these as empty boxes, your device is missing a font for that script. The MicroSim below loads free Noto fonts so every script displays.

Devanagari (देवनागरी, *day-vuh-NAA-guh-ree*) is the abugida used for Hindi, Marathi, Nepali, Sanskrit, Maithili and several other languages. Its name joins the Sanskrit words *deva* ("god" or "divine") and *nagari* (from *nagara*, "city"). You can recognize it by the line that runs along the top of most letters, with the letters hanging below it like laundry on a line. It is the script named in Article 343 for Hindi, and in Nepal it is used for Nepali.

Gurmukhi (ਗੁਰਮੁਖੀ, *GUR-moo-khee*) is the abugida used to write Punjabi in India and the Sikh scripture, the Guru Granth Sahib. Its name is often explained as "from the mouth of the Guru." Its standard form took shape in the 16th century and is popularly credited to Guru Angad, the second Sikh Guru. Gurmukhi has 35 base letters, traditionally arranged in seven rows of five. Simran learned them at her gurdwara's Sunday school. In Pakistan, where most Punjabi speakers live, Punjabi is usually written in Shahmukhi, a Perso-Arabic script.

Learn one, unlock many



Because most South Asian scripts share the abugida design and even the same order of letters (k, kh, g, gh...), learning one makes the next one much easier. If you can read Devanagari, learning Gujarati or Bangla is more like learning a new handwriting than a whole new system.

6.9 The Other Root: Aramaic, Arabic and Thaana

Not every script in South Asia comes from Brahmi. A second family of scripts came from the west.

Aramaic was the script the Persian Empire used for government business. In the northwest of South Asia, in the region called Gandhara (today's northern Pakistan and eastern Afghanistan), it gave rise to **Kharoshthi**, which was written right to left. Some of Ashoka's edicts, at Shahbazgarhi and Mansehra in present-day Pakistan, were carved in Kharoshthi rather than Brahmi. Kharoshthi fell out of use around the 3rd to 5th centuries CE, but it shows that two script traditions lived side by side in Ashoka's empire.

Much later, Arabic script, which also descends from Aramaic through older Semitic scripts, came to South Asia with Persian- and Arabic-speaking traders, scholars and rulers. Persian writers had already adapted it by adding letters, and South Asian writers added still more for sounds like the retroflex "ṭ" and "ḍ" that Arabic lacks.

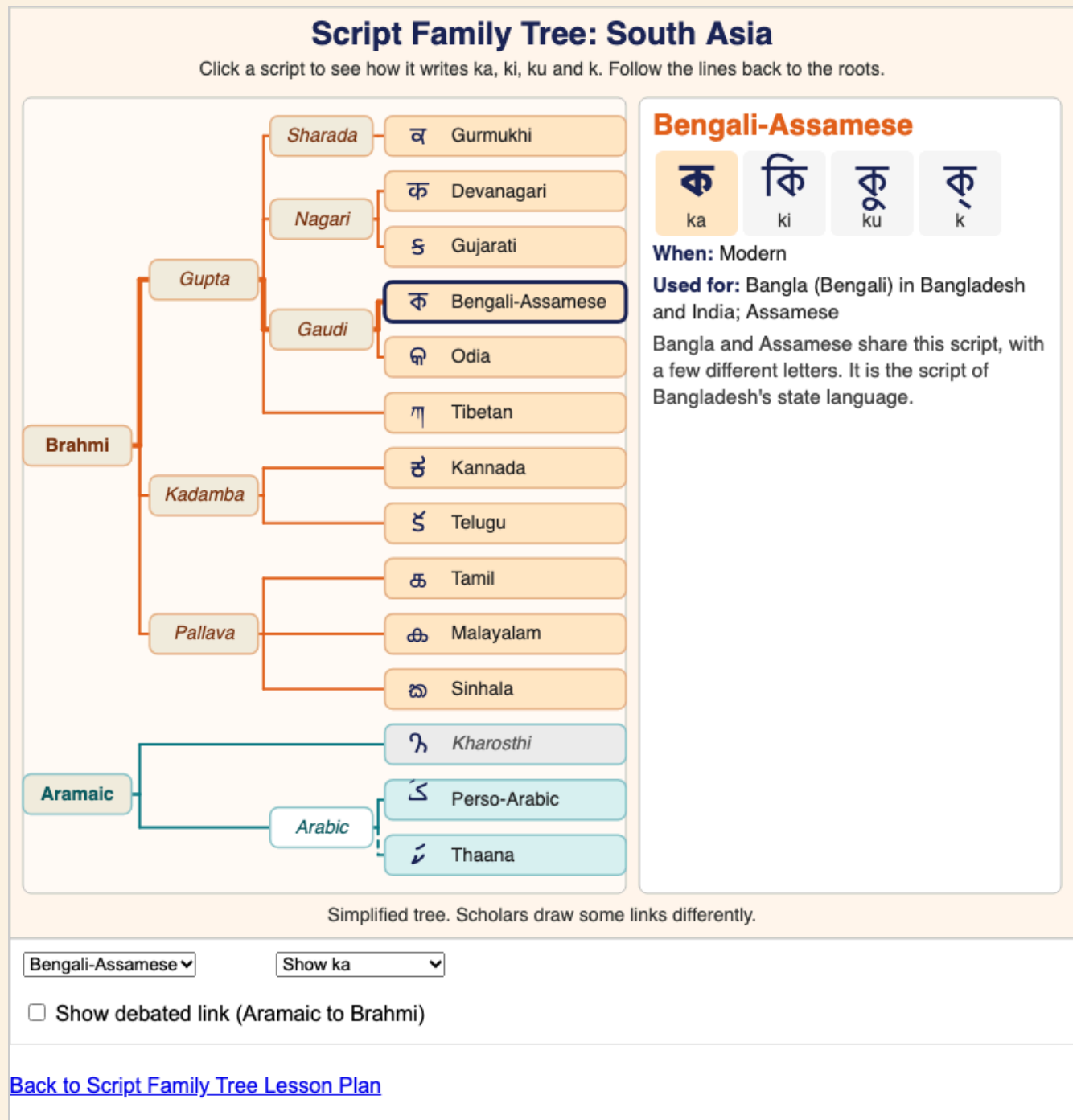
The **Perso-Arabic Script** is the Arabic script as adapted for Persian and then for many South Asian languages. Today it is used for Urdu, Pashto, Dari, Sindhi, Balochi, Kashmiri and Shahmukhi Punjabi, and for Brahui. It's written right to left. It works mostly as an **abjad**, a script that writes consonants and long vowels but usually leaves short vowels out. Readers fill them in from context, the way you can read "txt msgs" without vowels. Small optional marks can show the short vowels when needed, for example in children's books or the Qur'an. Urdu is usually written in Nastaliq, an elegant hanging style in which many words slope downward from right to left.

Thaana, the script of Dhivehi in the Maldives, has an unusual history. It is written right to left like Arabic, and its first nine letters were formed from Arabic numerals, while the next nine came from old Dhivehi numerals. Every vowel must be written, as a mark called a *fili* placed above or below the consonant, so Thaana sits somewhere between an abugida and a true alphabet. The oldest surviving Thaana writing, carved on the door posts of a

mosque in the northern Maldives, is dated 1599, and over the following centuries Thaana replaced the older Maldivian script, Dives Akuru.

So when people say "South Asian scripts," they mean at least two big families, plus newer inventions like Ol Chiki, and the Latin alphabet used for English and for typing on phones.

Diagram: Script Family Tree



Interactive simulation: <https://aaryavvatts-lab.github.io/south-asian-culture-textbook/sims/ch06-script-family-tree/main.html>

Open the Script Family Tree fullscreen

MicroSim specification: Script Family Tree

sim-id: ch06-script-family-tree

Library: p5.js

Status: Implemented

Bloom Level: Understand

Bloom Verb: explain

Learning Objective: The learner will explain that many South Asian scripts share Brahmi as a common ancestor, and explain how an abugida changes one consonant's inherent vowel with marks.

Canvas: responsive width; 700 px drawing area plus 80 px control area.

Controls: script menu, sound menu (ka, ki, ku, k with no vowel), "Show debated link" checkbox; click or tap any script on the tree.

Behavior: A simplified tree with two roots. Brahmi branches through Gupta (to Sharada, Nagari, Gaudi and Tibetan), Kadamba and Pallava to eleven modern scripts. Aramaic leads to Kharoshthi and, through Arabic, to the Perso-Arabic script, with a dashed line to Thaana. Each modern script shows the selected syllable in real Unicode text (Noto web fonts load so all scripts display). Selecting a script shows ka, ki, ku and k side by side, when and where it was used, and which languages use it. The checkbox adds a dashed "debated" link from Aramaic to Brahmi. A caption notes that the tree is simplified.

Start with Bengali-Assamese selected and read the four boxes in the panel: one base letter, four sounds. Then use the **Show** menu to switch the whole

tree from *ka* to *ki*. Watch every Brahmi-family script add a vowel mark to the same base letter, sometimes before it, sometimes after, above or below. Switch to **k (no vowel)** and find the one Brahmi-family script that shows "none" (the panel explains why).

Next, follow the lines. Click Tamil, then Devanagari, and trace each one back to Brahmi. Then click Perso-Arabic and Thaana and notice that they hang from a different root entirely. Finally, tick **Show debated link**. Most scholars think the people who created Brahmi borrowed the idea of writing from an Aramaic-family script, then redesigned it for South Asian sounds. Others argue Brahmi developed locally, and a few link it to the Indus signs, though the evidence for that is very thin. That's why the line is dashed.

6.10 Living Between Languages

Most people in South Asia, and many South Asian families in the United States, use more than one language every day, and that shapes how they talk.

Code-Switching is moving back and forth between two or more languages (or styles of speaking) in the same conversation, or even the same sentence. A teen in Mumbai might say, "Yaar, the exam was *bahut* hard," mixing Hindi and English. Mixes like this are sometimes nicknamed "Hinglish." Zara's family does the same in "Banglish," and Dilan Perera's cousins in Kandy slide between Sinhala and English. Code-switching isn't a sign that someone speaks either language badly. Linguists study it as a skill: speakers switch to fit the topic, show closeness, make a joke land, or use the word that simply fits best.

Your mix counts



If you answer your parents in English, mix languages mid-sentence, or understand far more than you can say, you're not "losing" your language. You're a heritage speaker, and that is a real foundation you can build on at any age.

A **Heritage Language** is a language that a person learned at home from family, while growing up in a place where a different language is the main one. Zara is a heritage speaker of Bangla: she understands her Nanu easily, but she answers in English, and reading Bangla is slow for her. That's a very common pattern. Heritage speakers often understand more than they speak, speak more than they read, and feel shy about their accent or grammar. Some US colleges offer classes designed for heritage learners of languages such as Hindi, Urdu and Bangla, and many temples, mosques, gurdwaras and cultural centers run weekend language classes, like the one where Simran learned Gurmukhi.

When heritage speakers text, they often write their family language in English letters: *kemon acho?* for Bangla "How are you?" or *kya haal hai?* in Hindi-Urdu. Writing one script's words in another script's letters is called transliteration, and there's a guide to it in this book's **Pronunciation Guide**.

Back at the name-tag table, Zara pulls up the Script Family Tree on her phone. She taps Bangla, then Telugu, then Gurmukhi, and follows each line back to Brahmi. "Cousins," she says, holding up her name tag next to Arjun's. Ben squints at বাংলা and తెలుగు. "So it's like Swedish and Spanish using the same alphabet," he says, "but backwards. Different alphabets, same great-great-grandparent." Simran nods. "And Ayesha's Urdu is from the other side of the family." Zara decides that tonight she'll ask Nanu to teach her one new Bangla word, and to write it down for her, too.

Ask Your Elders

Your interview for this chapter

Talk with a parent, grandparent, aunt, uncle, family friend or community elder about the languages they grew up with, at home, at school and in the street. If you don't have a South Asian elder to ask, interview a club advisor, a local temple, mosque, gurdwara or cultural center volunteer, or a classmate's family (with permission). Ten minutes is enough.

Before you start: Ask them to write a few words in their script on a piece of paper (their name, their town, or a favorite word), and ask if you can keep it and take notes.

Questions to ask

1. What language or languages did you speak at home when you were little? Was it different at school or with friends?
2. When did you learn to read and write, and in which script? Who taught you?
3. Is there a word in your language that has no good English translation? What does it mean, and when do you use it?
4. Do you mix languages when you talk? Which words do you say in English and which in your language?
5. How has the way our family uses its language changed since you were young, or since moving to a new country?
6. Do you worry about our language being lost? What is one thing about it you hope I keep?

Listen for: whether they grew up speaking a Heritage Language or a regional language at home and another at school; examples of Code-Switching; and

which script they write in, so you can find it on the Script Family Tree (is it Devanagari, Gurmukhi, a Perso-Arabic Script, or another descendant of the Brahmi Script?).

Save it: Write down one quote or story in their words, with their permission. If they agree, you can share it with the [Roots & Rivers Story Archive](#).

? Your turn: compare two answers

Find your elder's script on the Script Family Tree and trace it back to its root. Then compare with a friend's interview: do your two families' scripts share a parent, a grandparent, or no ancestor at all? Did either elder mention a word with no English translation, and could you guess its meaning from context?

"Every family does it a little differently, and there is no wrong answer here."

Common Misconceptions

Myth: Hindi is India's national language.

Reality: India's Constitution does not name a national language. Hindi in Devanagari is the official language of the central government, English is also used officially, and 22 languages are listed in the Eighth Schedule. States choose their own official languages.

Myth: Hindi and Urdu are unrelated languages.

Reality: Hindi and Urdu are two standard forms of one spoken language, Hindi-Urdu (Hindustani). Everyday speakers understand each other well. They differ in script (Devanagari versus Perso-Arabic) and in formal vocabulary (more Sanskrit words versus more Persian and Arabic words).

Myth: Sanskrit is the mother of all South Asian languages.

Reality: Sanskrit is an ancestor of many Indo-Aryan languages, like Hindi and

Bangla, but Tamil, Telugu, Kannada and Malayalam belong to the separate Dravidian family. Tibeto-Burman, Austroasiatic and Iranian languages have their own ancestors too. Languages have borrowed words from each other for centuries, but borrowing is not descent.

Myth: A script and a language are the same thing.

Reality: A language is spoken; a script is a way to write it. Punjabi is written in Gurmukhi in India and Shahmukhi in Pakistan, and Devanagari is used for Hindi, Marathi, Nepali and Sanskrit.

Myth: If you answer your parents in English, you've lost your family language.

Reality: Understanding a language you grew up hearing is real knowledge. Heritage speakers usually understand far more than they speak, and they often learn to speak, read and write faster than beginners when they get the chance.

Check Yourself

? 1. What is the difference between a language and a script? Give one South Asian example. ✓

A language is a system of spoken words and grammar, and a script is a set of written symbols. Punjabi is one language written in two scripts: Gurmukhi in India and Shahmukhi (Perso-Arabic) in Pakistan.

? 2. Name the language family that Tamil, Telugu, Kannada and Malayalam belong to, and one language from that family spoken outside India. ✓

They are Dravidian languages. Brahui, spoken mainly in Balochistan, Pakistan, is also Dravidian (and Tamil is spoken in Sri Lanka).

? 3. How many languages are in India's Eighth Schedule, and what does the Constitution say about a national language? ✓

The Eighth Schedule lists 22 languages. The Constitution names no national language. Article 343 makes Hindi in Devanagari the official language of the Union, with English still used officially.

? 4. In an abugida, what does the plain consonant letter क sound like, and what does the virama do? ✓

A plain consonant letter includes the inherent vowel "a," so क is "ka." The virama removes that vowel, so क् is a bare "k."

? 5. Why did UNESCO choose 21 February for International Mother Language Day? ✓

On 21 February 1952, students in Dhaka protesting for Bangla to be a national language of Pakistan were shot and killed by police. Bangladesh proposed the day, and UNESCO declared it in 1999.

Exercises

Remember

1. List the five language families of South Asia named in this chapter, with one language from each.

 Answer

Indo-Aryan (Hindi, Bangla, Sinhala), Iranian (Pashto, Dari), Dravidian (Tamil, Telugu), Tibeto-Burman (Dzongkha, Manipuri), Austroasiatic (Santali, Khasi). Any one correct example per family is fine.

2. Which scripts are used to write Hindi, Urdu, Punjabi in India, and Dhivehi?

 Answer

Hindi: Devanagari. Urdu: Perso-Arabic (Nastaliq style). Punjabi in India: Gurmukhi. Dhivehi: Thaana.

Understand

1. Explain in your own words why Telugu and Hindi use related scripts even though they are in different language families.

 Answer

Scripts and languages have separate family trees. Telugu is Dravidian and Hindi is Indo-Aryan, but both communities write with scripts that grew out of Brahmi, which spread across the region and changed over time in different places.

2. Explain why speakers of Hindi and Urdu can usually understand each other in conversation but may struggle with each other's newspapers.

Answer

Spoken Hindi and Urdu share grammar and everyday words, so conversation works. Newspapers use different scripts (Devanagari and Perso-Arabic) and more formal vocabulary drawn from Sanskrit in Hindi and from Persian and Arabic in Urdu.

Apply

1. Use the Script Family Tree. Switch the menu to *ki*. Pick three scripts and describe where the "i" mark sits (before, after, above or below the consonant) in each. *Hint: zoom in on the panel glyphs, which are larger.*
2. A classmate shows you a sign written right to left with flowing, slanted letters that seem to have few vowel marks. Using this chapter, name the most likely script family and two languages it might be. *Hint: which scripts on the tree are written right to left?*

Analyze

1. Compare the language-family table in Section 6.3 with the script tree in the MicroSim. Find two cases where languages in different families share a script family, and one case where one language uses scripts from two different families. *Hint: look at Dravidian languages and at Punjabi.*
2. Bangla is the state language of Bangladesh, but Urdu was the planned sole national language of Pakistan in 1948 to 1952. Analyze why a language decision could lead to protests and, later, to a new international day. *Hint: think about what it means to be educated and governed in a language you don't speak at home.*

Evaluate

1. Should India name a single national language? Using facts from this chapter (the Eighth Schedule, the census numbers and the different language families), argue for or against, and address the strongest point on the other side. *Hint: who would be helped and who might feel left out?*
2. Someone says code-switching is "lazy" and shows a person doesn't really know either language. Evaluate this claim using evidence from the chapter and from your own experience. *Hint: why might a speaker choose one word over another?*

Create

1. Design a name tag for your club that shows one person's name or one word in two scripts from different branches of the Script Family Tree, with a one-line caption explaining how the scripts are related. *Hint: verify the spelling with someone who reads that script, or with a dictionary.*
2. Write a short dialogue (8 to 12 lines) between a heritage speaker and a grandparent that includes realistic code-switching. Afterward, add two sentences explaining why each switch happens. *Hint: switches often happen for food, feelings, jokes and school words.*

Key Takeaways

- A language is spoken; a script is how it's written. One language can use two scripts, and one script can write many languages.
- South Asia's languages belong to five main families: Indo-Aryan and Iranian (both Indo-European), Dravidian, Tibeto-Burman and

Austroasiatic. In India's 2011 census, 78.05% of people spoke an Indo-Aryan mother tongue and 19.64% a Dravidian one.

- India's Constitution names no national language. Hindi in Devanagari is the Union's official language, English is also used officially, and the Eighth Schedule recognizes 22 languages.
- Hindi and Urdu are two standard forms of one spoken language, Hindi-Urdu (Hindustani), written in different scripts with different formal vocabulary.
- Most South Asian scripts, from Devanagari and Gurmukhi to Bangla, Tamil, Sinhala and Tibetan, descend from Brahmi, in use by the 3rd century BCE. The Perso-Arabic script and Thaana come from a different root.
- Brahmi-family scripts are abugidas: each consonant carries an inherent "a," and marks change or remove the vowel.
- Code-switching and heritage-language skills are normal and valuable, not signs of "losing" a language.

You can see the family resemblance!



You can now tell a language from a script, name South Asia's five language families, and read the "ka" family in several scripts. Next, in Chapter 7, we follow the calendar around the year to see why Diwali and Eid move on different schedules.

Quiz: Languages and Scripts

Test your understanding of Chapter 6. Choose the best answer, then click **Show Answer** to check.

Question 1

How many languages are listed in the Eighth Schedule of the Constitution of India?

- A. 14
- B. 22
- C. 121
- D. 270

? Show Answer



The correct answer is **B**. The Eighth Schedule lists 22 languages today. It started with 14 when the Constitution was adopted, and 121 is the number of languages India's 2011 census counted with at least 10,000 speakers, not the number on the official list.

Concept Tested: Eighth Schedule **Bloom's Level:** Remember

Question 2

Which language family do Tamil, Telugu, Kannada and Malayalam belong to?

- A. Indo-Aryan
- B. Tibeto-Burman
- C. Austroasiatic
- D. Dravidian

 Show Answer

The correct answer is **D**. These four are the largest Dravidian languages. Many people assume they are Indo-Aryan because they borrowed many Sanskrit words, but their grammar and core vocabulary come from a separate ancestor.

Concept Tested: Dravidian Languages **Bloom's Level:** Remember

Question 3

Which script is the common ancestor of Devanagari, Bengali-Assamese, Gurmukhi, Tamil and Sinhala?

- A. Brahmi
- B. Kharoshthi
- C. Perso-Arabic
- D. Thaana

 Show Answer

The correct answer is **A**. All five descend from Brahmi, which was in use by the 3rd century BCE. Kharoshthi was a separate script of the northwest that came from Aramaic and died out; it is not their ancestor.

Concept Tested: Brahmi Script **Bloom's Level:** Remember

Question 4

What does it mean to say that Devanagari is an abugida?

- A. It writes only consonants and leaves all vowels out.
- B. Each letter stands for a whole word.
- C. Each consonant letter carries a built-in "a," and marks change or remove the vowel.
- D. Every vowel is a separate full letter, as in English.

? Show Answer



The correct answer is **C**. In an abugida, क already reads "ka"; adding a mark makes कि ("ki") or कु ("ku"), and the virama makes क् ("k"). Option A describes an abjad, like the Perso-Arabic script, which usually leaves short vowels out.

Concept Tested: Abugida **Bloom's Level:** Understand

Question 5

Why can Hindi and Urdu speakers usually chat easily but have trouble reading each other's newspapers?

- A. They share everyday grammar and words but use different scripts and different formal vocabulary.
- B. They are unrelated languages that happen to share a few words.
- C. Urdu is written left to right and Hindi right to left, so the letters are reversed.
- D. Hindi has no written form, so newspapers use English.

 Show Answer

The correct answer is **A**. Hindi and Urdu are two standard forms of Hindi-Urdu (Hindustani). Speech is shared, but Hindi uses Devanagari and more Sanskrit-based formal words, while Urdu uses Perso-Arabic script, written right to left, and more Persian and Arabic formal words. Option C has the directions backward.

Concept Tested: Hindi-Urdu (Hindustani) **Bloom's Level:** Understand

Question 6

Which statement about India's languages is accurate?

- A. Hindi is India's national language under the Constitution.
- B. English is one of the 22 languages in the Eighth Schedule.
- C. Every Indian state must use Hindi as its official language.
- D. The Constitution names no national language; Hindi in Devanagari is the official language of the Union.

 Show Answer

The correct answer is **D**. Article 343 makes Hindi in Devanagari the official language of the central government, English is still used officially, and the Constitution names no national language. Option A is the common "national language" myth, and states choose their own official languages.

Concept Tested: Eighth Schedule **Bloom's Level:** Understand

Question 7

Ben finds an old family letter from a classmate's grandfather. It's written right to left, in flowing letters, with almost no vowel marks. Which language could it most likely be in?

- A. Tamil
- B. Urdu
- C. Bangla
- D. Gujarati

 Show Answer

The correct answer is **B**. Urdu is written in the Perso-Arabic script, right to left, usually without short vowel marks. Tamil, Bangla and Gujarati are written left to right in Brahmi-family abugidas that mark their vowels.

Concept Tested: Perso-Arabic Script **Bloom's Level:** Apply

Question 8

At dinner, Zara's cousin says, "Ammu, the *biryani* is so good, *aro dao* please." Which idea from the chapter best describes what she is doing?

- A. Transliteration
- B. Abugida
- C. Code-switching
- D. Classical language status

 Show Answer

The correct answer is **C**. She moves between English and Bangla in one sentence, which is code-switching. Transliteration is writing one language's words in another script's letters, which is about writing, not about mixing languages while speaking.

Concept Tested: Code-Switching **Bloom's Level:** Apply

Question 9

Telugu is a Dravidian language and Hindi is an Indo-Aryan language, yet the Telugu and Devanagari scripts are both descendants of Brahmi. What does this best show?

- A. Telugu must really be an Indo-Aryan language.
- B. Language families and script families are separate trees, so unrelated languages can share a script ancestor.
- C. Brahmi was invented for Telugu and later borrowed by Hindi speakers.
- D. All South Asian languages come from Sanskrit.

 Show Answer

The correct answer is **B**. A script can spread to communities speaking unrelated languages. Telugu's grammar and core words are Dravidian, while its script descends from Brahmi. Option D is the "Sanskrit is the mother of all languages" myth.

Concept Tested: Language Family **Bloom's Level:** Analyze

Question 10

Zara understands her Nanu's Bangla easily but answers in English and reads Bangla slowly. Which conclusion is best supported by the chapter?

- A. She is a heritage speaker, a common pattern in which understanding runs ahead of speaking and reading.
- B. She has lost Bangla and cannot relearn it.
- C. She is code-switching because she doesn't know English well.
- D. Bangla is not her family's language.

Show Answer

The correct answer is **A**. A heritage language is learned at home while growing up where another language dominates, and heritage speakers often understand more than they speak or read. Option B is the misconception the chapter corrects: understanding is real knowledge to build on.

Concept Tested: Heritage Language **Bloom's Level:** Analyze