

LinguaX Arc: A Corpus Linguistics Toolkit in Your Pocket

By Dr. Adnan Tahir, Associate Professor & Chairman, Department of English, Emerson University, Multan, Pakistan

Disclosure: the app's developer is the author's nephew, and the app carries a dedication to the author. The description below is based on the app's Google Play listing and the developer's release announcement as of October 2026, together with the author's own assessment.

Summary. Corpus linguistics has long depended on desktop software, which keeps it out of reach in many classrooms where the smartphone is the only computer students own. LinguaX Arc, a free Android app released in 2026, packs frequency analysis, concordancing, collocation, keyness and other corpus tools into a phone that works entirely offline. This article describes what the app offers, suggests five classroom activities built around it, and sets out the cautions researchers and teachers should keep in mind when using a new tool.

Introduction: the access problem

Since the early days of concordancing, corpus evidence has changed how languages are described and taught. Teachers can check whether a phrase is idiomatic, learners can discover patterns for themselves, and researchers can test claims against large bodies of authentic text. Tim Johns's idea of data-driven learning, in which the learner acts as a researcher, remains one of the most persuasive arguments for bringing corpora into the classroom.

The practical obstacle has been the tools. The best-known corpus programs were designed for laptops and desktops, and they assume a computer, a compatible operating system and often a computer lab. In many universities and colleges across South Asia and elsewhere, a large share of students own a smartphone but not a laptop. For them, corpus linguistics can remain something they read about rather than something they do.

LinguaX Arc is an attempt to close that gap.

What LinguaX Arc is

LinguaX Arc is an Android app in the Education category of Google Play, published under the name Dailwind Studios by the developer Arslan Tahir, a PhD linguistics scholar based in Pakistan. The listing describes it as a toolkit for text mining, natural language processing and corpus linguistics, created for students, teachers and language researchers, who can analyse text documents directly on their phones.

According to the developer's announcement, every feature is free and unlimited, and the developer describes the app as the first Android-based corpus linguistic analysis app on Google Play. That last claim is the developer's own and has not been independently checked. At the time of writing, the listing showed version 5.0 (updated 4 September 2026), a 4.9 rating from 36 reviews and more than 100 downloads, so the app is best seen as new and growing rather than established.

Offline by design

The listing's central selling point is that the app works completely offline and that nothing is sent to the internet. Google Play's data safety section for the app states that no data is collected and no data is shared with third parties; these are declarations by the developer.

For language work, this matters more than it might first appear. Student essays, interview transcripts, unpublished manuscripts and exam scripts are exactly the kind of text teachers and researchers hesitate to upload to a cloud service. An offline tool also works where connectivity is slow, expensive or unreliable, which describes many classrooms. The trade-off is that there are no user accounts and no cloud synchronisation. One reviewer asked for a history dashboard and sign-in; the developer replied that the app is intentionally offline to protect research data and that a local history feature and more advanced corpus tools are planned.

The toolkit

The developer groups the features into tools for everyday users and tools for corpus researchers.

Everyday text analysis

- **Word count and type-token ratio (TTR).** Counts tokens (running words) and types (distinct words) and expresses lexical variety as a ratio.
- **Word frequency.** Ranks the words in a text or corpus by how often they occur, the starting point of almost every corpus study.
- **Word cloud.** Draws the frequency list as an image in which larger words are more frequent. One reviewer called it a nice bonus.
- **Sentiment analysis.** Classifies the emotional polarity of a text; version 5.0 added an on-device sentiment and emotion analyzer.
- **Summarizer.** Condenses a long text into a shorter one. Version 5.0 lists an academic text summarizer with export to multiple formats.

Corpus linguistics tools

- **KWIC concordancing.** Displays every occurrence of a search word with its surrounding context, aligned so that patterns on the left and right become visible.
- **Dispersion plots.** Show where in a document a word occurs, which helps distinguish a word that is spread evenly from one clustered in a single section.
- **Collocational clusters and N-grams.** Find words and multi-word sequences that regularly occur together, such as lexical bundles like “in order to”.
- **Association metric collocate statistics.** Rank the collocates of a node word by how strongly they are attracted to it rather than by raw frequency alone.
- **Comparative keyword keyness.** Compares two texts to identify words that are unusually frequent in one of them, a way of revealing what a text is about.
- **Full document file views.** Lets users view the complete text of the documents they have loaded.

Version 5.0 additions

The latest update added a redesigned dashboard called the Executive Hub, an offline corpus merger, cleaner and document converter, and a top-bar quick search. The corpus preparation tools deserve attention: cleaning and combining files is unglamorous work, but it is often where a corpus project succeeds or fails, and being able to do it on the phone completes the workflow from raw files to results.

Five classroom activities

The tools become more useful when they are tied to a teaching purpose. The following activities draw on the data-driven learning tradition and need nothing but students' phones and a few text files.

- 1. Discover the pattern (KWIC).** Give students a short corpus and ask them to search for a verb such as *provide*. Which words follow it most often, and does the pattern change by text type? Students state the rule in their own words before the teacher confirms it. The aim is inductive learning: evidence first, rule second.
- 2. Collocation detectives.** Ask whether learners should *make* or *do* a decision, *strong* or *powerful* tea. Students predict, then check the collocates and their association scores. Errors that grammar rules cannot explain become something learners can investigate.
- 3. A personal learner corpus (word count and TTR).** Each student saves their writing across a semester as text files and tracks vocabulary range over time. One caution applies: TTR falls as texts get longer, so compare texts of similar length or treat the figures as rough indicators only.
- 4. What is this text about? (keyness).** Compare a set of subject articles with general writing to extract the vocabulary that defines the field. This suits English for Specific Purposes and academic writing classes, where students must learn the terminology and phrasing of their discipline.
- 5. Verify the chatbot.** Ask an AI chatbot to write a story or essay at a given level, import the result and examine it with frequency lists, TTR, KWIC lines and collocates. Does the text really use the target vocabulary? Does it overuse certain transitions? Students learn to treat AI output as a draft to be checked against evidence rather than as an authority.

Using it responsibly

A new tool deserves enthusiasm and scepticism in equal measure.

Treat research results with care. Established programs such as AntConc and #LancsBox have accumulated years of use, documentation and published validation. A new app has not yet had the same scrutiny. The Google Play listing does not specify details that corpus researchers normally need to report, such as how words are tokenised, whether stop-word lists are applied, or which association measures are calculated. For research intended for publication, record the settings used and cross-check key results in an established tool until the app's methods are documented.

Know the limits of the format. A phone is excellent for agile classroom analysis and for individual texts. Compiling and querying very large corpora remains a job for desktop software and scripting languages, as does advanced statistical modelling.

Remember what automation cannot do. Sentiment scores generally struggle with irony, sarcasm and context, and summaries can omit or distort important points. Both are useful for starting discussion; neither should end it.

Keep the evidence in perspective. A word cloud shows frequency, not meaning. Follow up interesting words with concordance lines to see how they are actually used.

Looking ahead

The developer says that a local history dashboard and additional corpus tools are planned. Several additions would make the app more valuable for research and teaching: published documentation of its methods, export of results in standard formats, and built-in support for reference corpora. There is also an opportunity for regional varieties. Teachers and researchers in Pakistan and India could build small local corpora and examine features of South Asian English that global AI systems and general reference corpora may overlook.

Conclusion

LinguaX Arc does not replace the established desktop tools, and it does not claim to. Its contribution is access: it lets a student or teacher with only a phone run frequency lists, concordances, collocations and keyness analyses anywhere, privately and without cost. Used with a clear pedagogical purpose and with care over its limits, it can help turn corpus linguistics from a topic in a syllabus into something learners actually do.

Availability

LinguaX Arc is available free on Google Play for Android devices; search for *LinguaX Arc* or use the listing at <https://play.google.com/store/apps/details?id=com.aistudio.linguaxarc.lhvqkp>.