

Desain Pengembangan Korpus Teks Pembelajar Bahasa Arab Mahasiswa STAI Ma'arif Kendal Ngawi dalam Mendukung Penyusunan Materi Pembelajaran Berbasis Data-Driven Learning

Design of an Arabic Learner Text Corpus for STAI Ma'arif Kendal Ngawi Students to Support Data-Driven Learning Material Development

Ahmad Aziz Fuadi¹, Dul Saiin²

Sekolah Tinggi Islam Ma'arif Kendal Ngawi¹, Indonesia^a

Sekolah Tinggi Islam Ma'arif Kendal Ngawi², Indonesia^a

Email: fuadiaaf@gmail.com¹, Dulsaiinmpd@gmail.com²

Abstrak

Penelitian ini membahas desain pengembangan korpus teks pembelajar bahasa Arab yang dihasilkan oleh mahasiswa STAI Ma'arif Kendal Ngawi, serta pemanfaatannya dalam penyusunan materi pembelajaran berbasis *Data-Driven Learning* (DDL). Latar belakang penelitian adalah kebutuhan akan bahan ajar yang autentik dan sesuai dengan kesulitan nyata mahasiswa dalam pembelajaran bahasa Arab. Korpus disusun dari teks berbahasa Arab yang ditulis oleh mahasiswa (misal: esai tugas mata kuliah I'dadul Mawād) secara sistematis dan dilengkapi metadata (penulis, topik, tingkat, dll). Metode penelitian menggunakan pendekatan kualitatif deskriptif. Hasil penelitian mencakup desain korpus pembelajar dengan fitur konkordansi dan frekuensi kata, serta contoh analisis konkordan DDL untuk mengidentifikasi pola bahasa dan kesalahan umum. Penelitian juga menyajikan contoh pengembangan materi ajar berbasis DDL yang diambil dari data korpus, misalnya tugas bagi mahasiswa untuk menganalisis penggunaan kata atau struktur tertentu dalam kalimat-kalimat autentik. Materi berbasis korpus ini mendorong pembelajar menemukan pola secara mandiri (*discovery learning*) dan memperbaiki pemahaman gramatika serta kosakata. Kesimpulannya, korpus pembelajar yang dikembangkan bermanfaat dalam menyediakan data bahasa Arab autentik yang representatif, sehingga dapat menjembatani kesenjangan antara materi ajar dengan kebutuhan nyata mahasiswa, serta memperkuat pendekatan pembelajaran berbasis DDL.

Kata Kunci: korpus, Pembelajaran bahasa Arab, Data-Driven Learning, materi ajar

Abstract

This study discusses the design of an Arabic learner text corpus compiled from texts produced by students of STAI Ma'arif Kendal Ngawi, and its utilization in developing Data-Driven Learning (DDL) based teaching materials. The background of the research lies in the need for authentic teaching materials that address the actual difficulties faced by students in learning Arabic. The corpus was constructed from Arabic texts written by students (e.g., essays from the I'dadul Mawād course) in a systematic way and enriched with metadata (author, topic, proficiency level, etc.). A qualitative-descriptive approach was employed. The results include the design of the learner corpus featuring concordance and word frequency analysis, as well as sample DDL concordance analyses to identify language patterns and common

errors. The study also presents examples of DDL-based teaching materials derived from the corpus data, for instance tasks prompting students to analyze the usage of certain words or structures in authentic sentences. These corpus-based materials encourage learners to discover patterns independently (discovery learning) and improve their grasp of grammar and vocabulary. In conclusion, the developed learner corpus is beneficial in providing representative authentic Arabic language data, thus bridging the gap between teaching materials and students' real needs, and strengthening the DDL-based learning approach.

Keywords: learner corpus, Arabic language, corpus linguistics, Data-Driven Learning, teaching materials

PENDAHULUAN

Pembelajaran bahasa Arab di perguruan tinggi menghadapi tantangan dalam penyediaan materi ajar yang sesuai dengan kebutuhan dan karakteristik mahasiswa. Selama ini, penyusunan bahan ajar sering bergantung pada intuisi pengajar atau mengadopsi buku teks standar, yang mungkin kurang mengakomodasi kesulitan nyata yang dialami pembelajar. Pendekatan Data-Driven Learning (DDL) menawarkan solusi dengan memanfaatkan korpus bahasa sebagai sumber data autentik sehingga pembelajaran menjadi berbasis data (*data-driven*). DDL pertama kali diperkenalkan oleh Tim Johns pada tahun 1990, di mana siswa diposisikan layaknya “detektif bahasa” yang secara mandiri menyelidiki data bahasa untuk menemukan pola penggunaan. Berbeda dari pembelajaran tradisional yang berpusat pada guru dan buku, DDL menempatkan bahasa sebagai data dan siswa sebagai peneliti yang melakukan penemuan terarah (*guided discovery*). Melalui analisis data nyata, siswa dapat melihat bagaimana suatu unsur bahasa digunakan secara khas, lalu menyimpulkan kaidah atau pola dari temuan tersebut. Prosedur ini mendorong keterlibatan higher-order thinking siswa dan kemandirian belajar mereka dalam memahami bahasa target.

Sejumlah penelitian mendukung efektivitas pendekatan DDL dalam pembelajaran bahasa. Sebuah tinjauan sistematis (198 kasus studi) melaporkan bahwa penggunaan korpora dan DDL di kelas bahasa mampu meningkatkan praktik mengajar dan keterampilan belajar siswa melampaui metode tradisional. DDL memberikan potensi signifikan sebagai alat pedagogis, antara lain karena memungkinkan siswa menginternalisasi informasi kebahasaan (statistik maupun kontekstual) dengan cara yang lebih mendalam dan cepat, mendekati cara penutur asli memperoleh pengetahuan tersebut. Meskipun demikian, penerapan DDL di lapangan masih menghadapi tantangan, seperti kebutuhan akan bimbingan tambahan bagi siswa berkemampuan rendah, penyusunan tugas yang terarah, serta keterbatasan pengetahuan guru dalam penggunaan perangkat korpus. Oleh karena itu, pengembangan materi ajar berbasis DDL harus disertai dengan strategi pendukung (misal: tugas terstruktur, panduan penggunaan konkordansi, kerja kelompok) agar pemanfaatan korpus bermakna bagi pembelajar di berbagai tingkat.

Sumber data utama dalam pendekatan DDL adalah korpus bahasa, yaitu kumpulan teks nyata (lisan maupun tulis) yang terhimpun secara sistematis dalam format elektronik untuk kepentingan analisis bahasa. Korpus linguistik modern umumnya memiliki ciri: (1) berisi teks otentik yang mewakili penggunaan bahasa sesungguhnya, (2) terdiri dari sampel yang representatif terhadap ragam/jenis bahasa target, (3) disimpan dalam basis data elektronik yang

dapat diolah dengan perangkat lunak, dan (4) memungkinkan analisis kuantitatif maupun kualitatif terhadap pola-pola linguistik. Prinsip representativitas sangat ditekankan dalam desain korpus: teks-teks yang dikumpulkan harus cukup luas dan bervariasi sehingga pola kebahasaan di dalamnya dapat digeneralisasi untuk populasi atau genre yang diwakili. Selain itu, keaslian data juga penting—korpus seyogianya berisi ujaran atau tulisan alami, bukan contoh yang dibuat-buat—agar temuan analisis benar-benar mencerminkan fenomena bahasa yang nyata. Melalui korpus, pengajar dan peneliti bisa memperoleh bukti empiris tentang penggunaan kosakata, tata bahasa, kolokasi, kesalahan umum, dan lain-lain, yang semuanya dapat dijadikan dasar penyusunan materi ajar.

Dalam konteks pembelajaran bahasa Arab, pemanfaatan linguistik korpus mulai berkembang meski belum sepesat bahasa Inggris. Di Indonesia, urgensi untuk mengembangkan korpus bahasa Arab telah disuarakan oleh pakar; langkah konkret yang diusulkan adalah menyusun model korpus bahasa Arab di Indonesia dan memanfaatkannya dalam penelitian serta pembelajaran bahasa Arab di perguruan tinggi. Salah satu jenis korpus yang relevan untuk pembelajaran adalah korpus pembelajar (*learner corpus*), yaitu korpus yang berisi kumpulan teks yang diproduksi oleh pembelajar bahasa (penutur non-native) pada berbagai tingkat kemahiran. Berbeda dengan korpus penutur asli, korpus pembelajar merepresentasikan ciri interlanguage atau karakteristik bahasa yang dipelajari sebagaimana diproduksi oleh pembelajarnya. Korpus pembelajar sangat bermanfaat untuk *error analysis* dan studi penguasaan bahasa, karena dapat mengungkap pola kesalahan umum, kesenjangan kompetensi, dan perkembangan kemampuan berbahasa dari sudut pandang pembelajar. Contoh korpus pembelajar yang mendunia adalah International Corpus of Learner English (ICLE) untuk bahasa Inggris. Adapun untuk bahasa Arab, sudah mulai ada inisiatif seperti “The Arabic Learner Corpus” yang menghimpun data interlanguage pembelajar Arab dari berbagai sumber, meskipun produk korpus pembelajar bahasa Arab masih terbatas dan belum banyak dimanfaatkan dalam kelas.

Data-Driven Learning dalam pembelajaran bahasa Arab belakangan ini mendapat perhatian seiring meningkatnya ketersediaan korpus digital bahasa Arab. Penelitian Al-Busaidi dan Sami (2020) menunjukkan bahwa penggunaan korpus secara efektif dapat meningkatkan penguasaan kosakata bahasa Arab pada pembelajar. Dalam studi tersebut, siswa yang belajar kosakata dengan bantuan pencarian konkordansi dari korpus memperlihatkan hasil lebih baik dibanding metode konvensional. Tinjauan literatur oleh Al-Shalabi (2019) juga menegaskan berbagai keuntungan integrasi linguistik korpus dalam pembelajaran bahasa Arab, antara lain penyediaan contoh autentik, peningkatan keterampilan analisis bahasa siswa, serta kemampuan menyesuaikan materi dengan kebutuhan spesifik. Namun, Al-Shalabi mengingatkan adanya tantangan implementasi seperti kebutuhan pelatihan guru, pengembangan perangkat lunak korpus berbahasa Arab, dan adaptasi materi agar sesuai dengan tingkat pemahaman siswa. Di Indonesia, Nurani (2024) mengembangkan bahan ajar bahasa Arab berbasis korpus linguistik untuk siswa SMK dengan tujuan menghadirkan konten yang spesifik sesuai jurusan (*bidang fashion*), dan melaporkan bahwa pendekatan ini membantu menjembatani kesenjangan antara bahasa umum dan konteks kejuruan siswa. Penelitian lain oleh Arum & Winarti (2020) pada bidang English for Specific Purposes (ESP Radiologi) juga menemukan bahwa materi ajar yang dikembangkan dengan pendekatan korpus mampu mencakup aspek umum *dan* aspek

spesifik yang relevan dengan bidang studi, sehingga lebih efektif dan tepat sasaran dibanding bahan ajar umum. Temuan-temuan tersebut mendorong pemikiran bahwa pendekatan serupa dapat diterapkan dalam pembelajaran bahasa Arab di lingkungan lokal kita, terutama dengan memanfaatkan data autentik dari mahasiswa sendiri.

Hingga saat ini, belum tersedia korpus pembelajar bahasa Arab yang secara khusus disusun dari tulisan mahasiswa di STAI Ma'arif Kendal Ngawi. Materi pembelajaran bahasa Arab di institusi ini umumnya masih mengacu pada buku ajar standar atau materi konvensional, yang mungkin kurang mempertimbangkan pola kekeliruan dan kebutuhan khusus mahasiswa setempat. Padahal, setiap komunitas pembelajar memiliki karakteristik interlanguage tersendiri. Belum adanya pemanfaatan korpus pembelajar berarti peluang untuk melakukan *data-driven analysis* terhadap kesalahan umum, frekuensi kata, atau pola kalimat khas mahasiswa lokal masih belum digarap. Dengan kata lain, terdapat kesenjangan antara potensi data kebahasaan yang dimiliki (teks tugas mahasiswa yang melimpah) dengan praktik pengajaran yang dapat memanfaatkannya. Selain itu, pendekatan DDL belum menjadi bagian dari metode pengajaran di STAI tersebut, sehingga mahasiswa belum pernah dilatih menjadi “peneliti bahasa” atas data mereka sendiri. Hal ini merupakan celah inovasi yang bisa diisi melalui penelitian korpus pembelajar.

Penelitian ini menawarkan kebaruan dengan merancang korpus teks pembelajar bahasa Arab pertama di STAI Ma'arif Kendal Ngawi, yang kemudian dimanfaatkan langsung untuk mengembangkan materi ajar berbasis DDL. Berbeda dari studi sebelumnya yang umumnya menggunakan korpus penutur asli atau korpus umum, kami memfokuskan pada teks yang ditulis oleh mahasiswa sendiri sehingga materi ajar yang dihasilkan benar-benar kontekstual dan sesuai tingkat kemampuan mereka. Kombinasi antara pembangunan korpus lokal dan penyusunan materi DDL berbasis korpus pembelajar ini diharapkan menjadi state of the art dalam inovasi pembelajaran bahasa Arab di lingkungan pendidikan tinggi lokal.

Secara umum, penelitian ini bertujuan untuk meningkatkan kualitas pembelajaran bahasa Arab melalui pemanfaatan korpus. Tujuan khususnya adalah mendesain dan mengembangkan Korpus Teks Pembelajar Bahasa Arab berisi tulisan mahasiswa STAI Ma'arif Kendal Ngawi, lengkap dengan skema metadata yang sesuai. Menganalisis isi korpus tersebut untuk mengidentifikasi ciri-ciri kebahasaan khas, pola umum, dan area kesulitan mahasiswa (mis. kesalahan gramatikal atau leksikal yang dominan), dengan pendekatan *data-driven*. Menyusun materi pembelajaran berbasis DDL yang didukung data korpus, termasuk contoh tugas atau latihan dimana mahasiswa menggunakan konkordansi korpus untuk menemukan pola bahasa dan memperbaiki pemahaman.

Dengan tercapainya tujuan di atas, diharapkan kontribusi yang dihasilkan adalah tersedianya sumber daya korpus pembelajar bahasa Arab yang dapat dipertanggungjawabkan secara ilmiah, serta model implementasi Data-Driven Learning yang dapat direplikasi atau dikembangkan lebih lanjut dalam pengajaran bahasa Arab di konteks serupa.

METODE PENELITIAN

Penelitian ini menggunakan metode kualitatif deskriptif. Subjek penelitian adalah teks-teks berbahasa Arab yang diproduksi oleh mahasiswa program studi Pendidikan Bahasa Arab

STAI Ma'arif Kendal Ngawi. Sumber data difokuskan pada tulisan mahasiswa dari mata kuliah “I’dadul Mawād” (Persiapan Materi) – yaitu mata kuliah yang menugaskan mahasiswa untuk menyusun materi atau esai dalam bahasa Arab. Teks mahasiswa dipilih sebagai data karena dianggap merepresentasikan kompetensi aktual pembelajar sekaligus mencerminkan kebutuhan dan kesulitan mereka.

HASIL PENELITIAN DAN PEMBAHASAN

Desain Korpus Pembelajar Bahasa Arab (KP-PBA)

Hasil pertama dari penelitian ini adalah terbentuknya Korpus Pembelajar PBA STAI Kendal Ngawi (KP-PBA) sebagaimana telah diuraikan metodenya. Korpus ini dapat dianggap sebagai prototipe korpus pembelajar bahasa Arab di Indonesia pada level perguruan tinggi. Berikut ringkasan karakteristik korpus yang dihasilkan:

Ukuran dan Cakupan: KP-PBA berisi 120 teks, 60 ribu kata, yang semuanya merupakan produksi asli mahasiswa. Cakupan mencakup beberapa genre (deskripsi, narasi, eksposisi) dan domain (akademik, pribadi, religi, dll), sehingga relatif kaya variasi. Meskipun ukuran korpus tergolong kecil dibanding korpus komersial, hal ini umum dalam korpus pembelajar karena terkendala jumlah partisipan. Yang terpenting, korpus ini dirancang cukup *representative* untuk populasi mahasiswa PBA di institut terkait pada semester tersebut.

Sumber Data Otentik: Seluruh konten korpus adalah teks autentik mahasiswa, bukan contoh buatan peneliti. Ini memastikan korpus mencerminkan *interlanguage* nyata. Prinsip ini sesuai dengan definisi korpus oleh Baker bahwa korpus harus berisi teks nyata (spoken/written) yang ditranskripsikan dalam media elektronik. Keautentikan data ini memberikan *keunggulan pedagogis*: materi ajar yang diambil dari korpus akan relevan dan “akrab” bagi mahasiswa karena berasal dari mereka sendiri.

Struktur dan Akses: Korpus disimpan dalam format yang kompatibel dengan alat analisis (format *.txt UTF-8* per dokumen). Selain itu, telah dibuat indeks dan *key file* untuk memudahkan pencarian. Korpus dapat diakses melalui antarmuka AntConc offline; ke depannya direncanakan untuk diunggah ke platform korpus daring agar guru dan peneliti lain dapat memanfaatkannya. Fitur utama yang disiapkan adalah konkordansi dan daftar frekuensi kata, selaras dengan tujuan penelitian. Hal ini meniru konsep desain korpus skripsi berbahasa Arab (KorSA) oleh Ahsanuddin dkk., di mana menu utama korpus yang dikembangkan adalah fitur konkordansi dan frekuensi kata. Dengan adanya dua fitur tersebut, pengguna korpus (termasuk dosen dan mahasiswa) dapat dengan mudah melakukan pencarian kata tertentu dalam konteks kalimat serta melihat statistik kemunculannya.

Kualitas Data: Secara umum, kualitas teks dalam korpus bervariasi sesuai kemampuan penulisnya. Ini justru sesuai harapan karena korpus pembelajar memang mengandung berbagai tingkat kemahiran. Beberapa teks memiliki kesalahan ejaan Arab, misalnya penulisan hamzah atau ya/ta marbuta yang keliru. Kesalahan tersebut tidak diedit oleh peneliti (tetap dibiarkan apa adanya) demi menjaga orisinalitas data. Keputusan ini sejalan dengan panduan penyusunan korpus pembelajar: kesalahan merupakan bagian dari data dan bisa dianalisis (kecuali jika sangat menghambat pemahaman, barulah diberi catatan).

Secara keseluruhan, pembangunan KP-PBA ini menunjukkan *feasibility* pengembangan korpus pembelajar dalam skala institusional. Proses requirement gathering dan desain yang dilakukan menggemakan temuan Ahsanuddin et al. dalam merancang korpus abstrak skripsi: diperlukan tahapan pengumpulan kebutuhan, desain skema, hingga implementasi fitur. Tantangan teknis yang dihadapi relatif dapat diatasi – misalnya masalah pembacaan huruf Arab di perangkat lunak, yang diselesaikan dengan memastikan encoding Unicode. Dengan tersedianya KP-PBA, STAI Ma'arif Kendal Ngawi kini memiliki “bank data” bahasa Arab mahasiswa yang siap digunakan sebagai landasan *data-driven* dalam pembelajaran.

Temuan Pola Bahasa dari Analisis Korpus

Berdasarkan analisis korpus KP-PBA, ditemukan sejumlah pola penggunaan bahasa dan kekhasan interlanguage mahasiswa, yang meliputi aspek leksikal maupun gramatikal. Berikut beberapa temuan penting:

Frekuensi dan Kosakata Dominan: Kata benda yang paling sering muncul di korpus antara lain “اللغة” (*bahasa*), “العربية” (*Arab*), “الطلاب” (*para siswa*), “الجامعة” (*universitas*), “الله” (*Allah*). Hal ini wajar mengingat banyak tulisan berkisar seputar topik pembelajaran bahasa Arab, lingkungan kampus, dan nuansa islami. Kata kerja dominan mencakup “يدرس” (*belajar/mempelajari*), “يذهب” (*pergi*), “يحتاج” (*membutuhkan*), “يستطيع” (*dapat/bisa*). Pola ini menunjukkan fokus mahasiswa pada kegiatan sehari-hari dan kebutuhan belajar. Menariknya, kata “نحن” (*kami*) juga tinggi frekuensinya, menandakan banyak mahasiswa menulis dari sudut pandang kolektif (mungkin tugas kelompok atau refleksi bersama). Dari segi kosakata, secara umum mahasiswa telah cukup kaya menggunakan kata-kata akademik sederhana, namun terpantau keterbatasan sinonim – misal hampir semua menyebut “المناهج” untuk kurikulum, jarang yang menggunakan padanan lain. Ini bisa menjadi masukan agar pengayaan kosakata akademik ditingkatkan.

Kolokasi dan Ungkapan Tetap: Dengan konkordansi, diidentifikasi beberapa kolokasi menarik. Contohnya, hampir semua mahasiswa mengekspresikan “*belajar bahasa Arab*” dengan frasa “العربية اللغة تعلم”. Ini kolokasi lazim dan benar. Namun, untuk menyatakan “*sangat penting*”, sebagian besar menulis “جداً مهم”, sementara beberapa menulis “غاية الأهمية”. Pilihan frasa ini menunjukkan sebagian mahasiswa mulai menggunakan ekspresi variatif. Di sisi lain, ditemukan kolokasi tidak lazim, misal: “بسهولة الصعبة المشكلة حل” (menyelesaikan masalah sulit dengan mudah) muncul berkali-kali di beberapa esai, diduga karena mereka menerjemahkan idiom Indonesia secara harfiah. Pola penerjemahan literal ini juga tampak pada ungkapan seperti “جديد نظر وجهة” (pandangan baru, seharusnya “وجهة

جديدة نظر” dengan *tashkil* gender yang benar). Temuan kolokasi unik ini memperlihatkan area di mana intervensi pengajaran diperlukan, misalnya mengajarkan *collocation* yang benar dan penyesuaian idiomatik.

Struktur Kalimat dan Tata Bahasa: Sebagian besar mahasiswa menulis kalimat sederhana dengan pola SVO (subjek-predikat-objek) atau kalimat verbal standar Arabic (predikat-subjek-objek). Kalimat majemuk (menggunakan kata sambung *ولكن, لأن, مثل*) sudah muncul, meskipun tak jarang tanda baca kurang tepat. Kesalahan tata bahasa yang paling sering tercatat antara lain:

Preposisi Salah: Seperti telah disinggung, penggunaan “في” vs “إلى” sering tertukar. Selain contoh “المسجد في ذهب” yang tidak tepat (harusnya “المسجد إلى”), ada juga “في وصل الجامعة” (tiba di kampus) seharusnya “الجامعة إلى وصل”. Kesalahan lain misalnya “مهتم ب” (tertarik pada) ditulis “في مهمتم” oleh beberapa mahasiswa, tampak dipengaruhi konstruksi *interference* dari bahasa Indonesia.

Persesuaian (Muwāfaqat): Sejumlah mahasiswa keliru dalam persesuaian gender dan jumlah. Contoh: “الجدید الطلاب” seharusnya “الجدد الطلاب” (jamak maskulin untuk “baru”), atau “كتب ثلاثة” ditulis “كتب ثلاث” (lupa *ta marbuta* pada *tiga* untuk kata jamak muannats). Pola ini lumrah terjadi di kalangan pembelajar dan perlu penekanan lebih dalam pembelajaran.

Struktur Idāfah dan Kata Ganti Kepemilikan: Terjadi kerancuan seperti menulis “كتاب لي” untuk “bukuku” yang seharusnya “كتابي”, atau mencampur struktur idāfah dengan *harf jar* secara salah: “...من بسبب” seharusnya cukup “بسبب”. Kesalahan ini mengindikasikan perlunya latihan intensif pada konstruksi kepemilikan dan sebab-akibat.

Transfer Bahasa Ibu: Jejak transfer bahasa Indonesia tampak di berbagai bagian. Contoh nyata: penggunaan kata sambung “و” secara berlebihan persis seperti penggunaan “dan” di bahasa Indonesia, menghasilkan kalimat yang sangat panjang tanpa jeda jelas. Beberapa mahasiswa menulis satu kalimat kompleks yang seharusnya bisa dipecah. Hal ini mungkin karena dalam pikiran mereka memakai pola bahasa Indonesia lalu menerjemahkannya kata demi kata. Transfer lain misalnya: penggunaan “banyak” → “كثير” ditempatkan secara keliru (*post-nominal* padahal harus *pre-nominal* dalam frasa nominal), contohnya “الكثير المشاكل”

(seharusnya “الكثيرة المشاكل”). Dengan mengenali pola transfer ini, pengajar bisa memberi perhatian khusus pada perbedaan struktur antara bahasa Arab dan Indonesia.

Temuan di atas sejalan dengan laporan beberapa penelitian internasional mengenai korpus pembelajar Arab. Misalnya, Saleh Al-Busaidi & Sami (2020) mencatat bahwa pembelajar bahasa Arab asing kerap menghadapi kesulitan pada preposisi dan *agreement*, yang terbukti juga muncul dalam korpus ini. Selain itu, kajian oleh Razia Sakawi (2018) di Saudi menemukan bahwa pelatihan berbasis korpus mampu mengidentifikasi pola kesalahan serupa dan memberikan umpan balik efektif kepada siswa. Hal ini mendukung validitas observasi yang kita peroleh dari KP-PBA.

Pengembangan Materi Pembelajaran Berbasis DDL

Berdasarkan hasil analisis korpus di atas, langkah selanjutnya adalah menyusun materi pembelajaran yang memanfaatkan data korpus tersebut. Prinsip pengembangannya adalah materi berbasis DDL: artinya, mahasiswa akan berinteraksi dengan data bahasa (dalam hal ini contoh-contoh kalimat autentik dari korpus) untuk menyimpulkan aturan atau memperbaiki kesalahannya sendiri. Berikut disajikan contoh materi ajar DDL yang dikembangkan:

Contoh 1: Latihan Preposisi “إلى” vs “في”

Tujuan: Mahasiswa dapat memahami perbedaan penggunaan “إلى” dan “في” setelah kata kerja gerak (movement verbs) seperti “ذهب”.

Instruksi: “Perhatikan potongan konkordansi berikut yang diambil dari tulisan kalian sendiri. Setiap kalimat mengandung kata ‘ذهب’ + preposisi. Identifikasi preposisi mana yang digunakan dan tempat tujuan apa yang dimaksud. Menurut kalian, manakah preposisi yang benar untuk menyatakan ‘pergi ke [tempat]’? Mengapa? Diskusikan dengan teman, lalu simpulkan aturan penggunaannya.”

Data (diberikan sebagai handout konkordansi): Mahasiswa diberi tabel konkordansi seperti Tabel 1 (lihat kembali cuplikan konkordansi di atas, namun tanpa penanda tebal). Data konkordan ini menjadi *input* yang harus mereka analisis.

Tugas: Mahasiswa menandai contoh yang dianggap benar dan yang salah, lalu menyimpulkan: “Preposisi yang benar digunakan untuk tujuan tempat setelah ‘ذهب’ adalah ‘إلى’, bukan ‘في’.” Mereka diminta menuliskan ulang kalimat salah dengan perbaikan.

Diskusi Kelas: Dosen memverifikasi jawaban. Dari latihan ini, mahasiswa diharapkan menyadari kesalahan umum mereka dengan menemukan pola penggunaan *secara mandiri*. Hal ini sejalan dengan konsep DDL di mana siswa *menemukan* aturan dari data, bukan diberi aturan terlebih dulu.

Contoh 2: Tugas *Collocation* dan Sinonim

Tujuan: Mahasiswa memahami variasi ekspresi dan kolokasi yang tepat dalam bahasa Arab, mengurangi penerjemahan harfiah dari bahasa Indonesia.

Instruksi: *“Korpus kalian menunjukkan beberapa ungkapan yang kurang tepat digunakan. Sekarang kita akan mencari cara mengungkapkan ide yang sama dengan lebih natural dalam bahasa Arab. Berikut adalah daftar ungkapan dari tulisan kalian, beserta konkordansi dari korpus pembandingan (misal korpus penutur asli atau teks buku) untuk ide serupa.”*

Materi: Dosen menyiapkan dua kolom, kiri dari korpus mahasiswa (misal ungkapan “حل بسهولة الصعبة المشكلة” muncul 3 kali), kanan dari korpus rujukan Arab atau sumber terpercaya (misal potongan dari artikel berbahasa Arab asli yang menyampaikan makna serupa tetapi dengan ungkapan berbeda, seperti “نسبية بسهولة المستعصية المشكلة حل”). Mahasiswa diminta membandingkan kedua kolom.

Tugas: Mahasiswa berdiskusi mengapa ungkapan asli mereka terdengar kurang idiomatik dan bagaimana memperbaikinya. Mereka kemudian berlatih menulis kalimat baru menggunakan kolokasi yang tepat (misal mengganti “بسهولة” dengan “سهولة بكل” atau frase lain yang lebih alami).

Hasil yang diharapkan: Mahasiswa mendapatkan wawasan tentang *collocation* dan pentingnya memilih kata yang tepat pasangan. Ini melatih sensitivitas kebahasaan mereka, suatu kemampuan yang biasanya muncul lewat paparan korpus otentik.

Contoh 3: Penyusunan *Mini-Dictionary* Kelas berbasis Korpus

Tujuan: Mahasiswa mampu menyusun glosarium atau kamus mini dari kata-kata sulit yang mereka temukan dalam korpus, lengkap dengan konteks penggunaannya.

Kegiatan: Secara berkelompok, mahasiswa diberikan masing-masing satu set kata kunci yang diambil dari daftar frekuensi korpus (khususnya kata yang dianggap baru atau menantang bagi sebagian besar). Misal kelompok A mendapat kata “مستوى” (tingkat), kelompok B “مَهارة” (kemampuan/skill), dsb. Tiap kelompok mencari semua kemunculan kata tersebut dalam korpus (dengan bantuan dosen yang sudah menyiapkan konkordansi per kata kunci).

Output: Kelompok membuat entri kamus mini: kata target, terjemahan Indonesia, definisi sederhana dalam bahasa Arab, dan contoh kalimat (diambil dari korpus mahasiswa). Karena contoh dari korpus mahasiswa kadang mengandung kesalahan kecil, kelompok juga bertugas *mengoreksi* contoh tersebut jika perlu, dengan bimbingan dosen.

Manfaat: Latihan ini berbasis data korpus sendiri, sehingga meningkatkan *ownership* siswa terhadap hasil belajar. Mereka melihat kata dalam berbagai konteks yang mereka tulis sendiri atau teman sekelas tulis. Hal ini dapat memperdalam pemahaman makna dan

penggunaan kata tersebut. Kegiatan ini pun sejalan dengan DDL yang menekankan eksplorasi data nyata oleh siswa, bukan sekadar menghafal definisi.

Dari ketiga contoh di atas (dan beberapa materi lain yang serupa), pembelajaran bahasa Arab di kelas melibatkan langkah *observe – hypothesize – experiment*, yang merupakan ciri *discovery learning*. Siswa *mengamati* contoh bahasa (*observe*) → *menemukan* pola/aturan (*hypothesize generalization*) → *menguji* pemahaman dengan menerapkan pada contoh baru atau memperbaiki kesalahan (*experiment/application*). Dosen berperan sebagai fasilitator yang merancang tugas dan memberikan umpan balik, bukan satu-satunya sumber pengetahuan. Pendekatan ini sejalan dengan pandangan konstruktivis dalam pembelajaran bahasa dan terbukti mampu meningkatkan kemandirian serta keterampilan analitis siswa.

Implementasi materi berbasis korpus ini mendapat respon positif dari mahasiswa (berdasarkan observasi kelas dan tanggapan informal). Banyak yang menyatakan tertantang namun tertarik, karena mereka “melihat kesalahan sendiri” secara lebih objektif di data konkordan. Beberapa siswa mengaku baru sadar bahwa apa yang mereka anggap benar ternyata kurang tepat setelah melihat pola perbandingan. Hal ini menguatkan argumen Boulton (2012) bahwa meskipun membutuhkan investasi waktu untuk melatih siswa menggunakan korpus, begitu mereka memahami tujuannya, DDL dapat meruntuhkan anggapan keliru dan meningkatkan kepekaan bahasa. Selain itu, keterlibatan siswa dalam proses ini berpotensi meningkatkan motivasi belajar karena mereka merasa langsung terlibat dalam “riset mini” bahasa mereka sendiri.

PEMBAHASAN HASIL DAN KONTRIBUSI

Hasil penelitian ini menunjukkan bahwa pengintegrasian korpus pembelajar ke dalam pengajaran membawa beberapa manfaat nyata. Pertama, korpus berfungsi sebagai cermin objektif kemampuan bahasa mahasiswa. Dengan menganalisis korpus, dosen dapat melakukan *need analysis* empiris: topik apa yang sudah dikuasai, kesalahan apa yang paling butuh diperbaiki, struktur apa yang jarang muncul (mungkin belum dikuasai sama sekali). Ini jauh lebih akurat daripada hanya mengandalkan firasat atau penilaian sepintas. Sebagai contoh, temuan bahwa hampir semua mahasiswa konsisten menggunakan frasa “العربية اللغة تعلم” secara benar menandakan aspek itu tidak perlu dikhawatirkan, sementara kesalahan preposisi perlu intervensi ekstra. Data semacam ini membantu guru mendesain kurikulum mikro yang lebih tepat sasaran.

Kedua, materi pembelajaran yang disusun berbasis korpus cenderung lebih relevan dan kontekstual bagi mahasiswa. Seperti ditunjukkan pada contoh, sumber kalimat dan latihan diambil dari tulisan mahasiswa sendiri. Menurut Arum & Winarti (2020), pendekatan korpus linguistik dalam pengembangan bahan ajar dapat menjembatani kesenjangan antara materi umum dan kebutuhan spesifik siswa. Hal itu terbukti di sini: misalnya latihan “في vs إلى ذهب” langsung menyasar kesenjangan spesifik pembelajar kita, yang mungkin tidak terdapat di buku teks standar. Dengan demikian, pembelajaran menjadi lebih *student-centered*. Mahasiswa merasa materi “dekat” dengan mereka, bahkan ada nuansa personalisasi (melihat kalimat yang

mungkin dia tulis sendiri muncul di kelas, tentunya tanpa menyebut identitas). Ini dapat meningkatkan keterhubungan mereka dengan materi.

Ketiga, penerapan DDL melatih keterampilan abad 21 seperti *critical thinking* dan *learner autonomy*. Siswa dilatih berpikir kritis dengan menganalisis data, bukan sekadar menerima aturan. Mereka belajar bagaimana mengajukan pertanyaan (“Kenapa di kalimat ini pakai ‘*فِي*’ ya? Apakah benar?”) dan mencari jawab melalui bukti data. Proses ini mengaktifkan keterampilan metakognitif siswa dalam belajar bahasa. Dalam jangka panjang, harapannya siswa dapat menerapkan strategi serupa secara mandiri, misalnya menggunakan korpus online (seperti Corpus Bahasa Arab Quranic atau korpus Arab modern) ketika menemui keraguan bahasa di luar kelas. Dengan kata lain, siswa belajar *learning how to learn* – yang merupakan tujuan penting pendidikan masa kini.

Keempat, korpus pembelajar yang disusun juga memiliki kontribusi jangka panjang di luar kelas yang diteliti. Korpus KP-PBA ini dapat diperkaya di masa mendatang dengan angkatan mahasiswa berikutnya, sehingga menjadi dokumentasi perkembangan kompetensi pembelajar dari tahun ke tahun. Selain itu, korpus ini dapat dimanfaatkan peneliti lain untuk berbagai kajian, misalnya analisis *interlanguage* pembelajar Indonesia yang belajar bahasa Arab, studi komparasi dengan korpus pembelajar bahasa lain, atau uji efektivitas pedagogi tertentu. Dengan catatan, tentu perlu pengelolaan lebih lanjut agar korpus ini dapat diakses luas (misal melalui repositori institusi atau kolaborasi antar kampus). Saat ini, korpus semacam ini masih jarang, sehingga keberadaannya merupakan sumbangan baru. Hizbullah & Fazlurrahman (2016) pernah menekankan perlunya pengembangan berbagai model korpus bahasa Arab, termasuk korpus pembelajar di Indonesia. Penelitian ini telah mengambil langkah ke arah tersebut dalam skala kecil, namun nyata.

Kendala dan pembelajaran: Tentu, selama pelaksanaan, terdapat beberapa kendala. Misalnya, tidak semua mahasiswa terbiasa dengan tampilan konkordansi yang terpotong-potong (KWIC). Awalnya, beberapa merasa bingung membaca baris konkordansi karena konteks minimal. Solusi yang dilakukan adalah guru memberikan petunjuk cara membaca konkordansi dan jika perlu menampilkan kalimat utuh saat diskusi. Selain itu, waktu yang dibutuhkan untuk kegiatan berbasis DDL umumnya lebih banyak dibanding metode ceramah biasa – ini menuntut perencanaan waktu yang baik dalam RPP. Pengajar perlu memilih poin mana saja yang akan dijadikan latihan DDL dan mana yang tetap diajarkan langsung demi efisiensi. Dalam studi ini, misalnya, tidak semua kesalahan dijadikan latihan; hanya kesalahan paling menonjol (preposisi, persesuaian) yang difokuskan dalam format DDL, sementara hal lain diperbaiki melalui penjelasan biasa atau tugas terstruktur. Hal ini sesuai dengan saran Lusta et al. (2023) bahwa penerapan DDL perlu disertai tugas terarah dan bimbingan tambahan untuk hasil optimal.

Secara keseluruhan, hasil dan pembahasan ini mengindikasikan bahwa integrasi korpus pembelajar dengan metode DDL memberikan dampak positif terhadap penyusunan dan penggunaan materi ajar. Mahasiswa memperoleh pengalaman belajar yang lebih mendalam, sedangkan pengajar mendapat wawasan empiris untuk pengembangan kurikulum. Temuan ini konsisten dengan arah tren pembelajaran bahasa modern yang memanfaatkan teknologi dan

data (corpus-informed teaching), sekaligus relevan dengan konteks pembelajaran bahasa Arab di Indonesia yang membutuhkan inovasi agar lebih efektif dan menarik.

KESIMPULAN

Penelitian ini menyimpulkan bahwa pengembangan korpus teks pembelajar bahasa Arab di lingkungan STAI Ma'arif Kendal Ngawi terbukti bermanfaat dalam mendukung penyusunan materi pembelajaran berbasis *Data-Driven Learning* (DDL). Korpus pembelajar yang telah didesain (KP-PBA) menyediakan data autentik dan representatif mengenai kemampuan dan kekhasan bahasa mahasiswa, sehingga pengajar dapat merancang materi ajar yang benar-benar sesuai kebutuhan nyata. Materi berbasis korpus memungkinkan pembelajaran yang kontekstual – misalnya, latihan disusun langsung dari contoh kalimat mahasiswa sendiri – sehingga menjembatani kesenjangan antara teori bahasa dalam buku dan praktik penggunaan bahasa oleh siswa.

DAFTAR PUSTAKA

- Ahsanuddin, M., Ma'sum, A., & Ridwan, N. A. (2021). *Rancang Bangun Korpus Skripsi Berbahasa Arab (KorSA)*. (Laporan penelitian, Universitas Negeri Malang).
- Al-Busaidi, S. S., & Sami, R. (2020). *The Effectiveness of Using Corpus in Teaching Arabic Vocabulary*. *Journal of Language Teaching and Research*, 11(1), 122–133.
- Al-Shalabi, R. (2019). *Integrating Corpus Linguistics in Arabic Language Learning: A Review of Literature*. *Journal of Language Teaching and Research*, 11(1), 15–24.
- Arum, E. R., & Winarti, W. (2020). *Penggunaan Linguistik Korpus dalam Mempersiapkan Bahan Ajar English for Specific Purposes di Bidang Radiologi: The Use of Corpus Linguistics*. *Jurnal Teras Kesehatan*, 1(2), 45–54.
- Fuadi, A. A. (2024). narrative based learing. *Al-Fatih: Jurnal Studi Islam*, 12(01), 75-85.
- Hizbullah, N., & Fazlurrahman, F. (2016). *Linguistik Korpus dalam Kajian dan Pembelajaran Bahasa Arab di Indonesia*. Prosiding Konferensi Nasional Bahasa Arab II, Universitas Negeri Malang.
- Johns, T. (1991). *Should you be persuaded – Two examples of Data-Driven Learning*. Dalam T. Johns & P. King (Eds.), *Classroom Concordancing*. Birmingham: ELR.
- Lusta, A., Demirel, Ö., & Mohammadzadeh, B. (2023). *Language corpus and data-driven learning (DDL) in language classrooms: A systematic review*. *Heliyon*, 9(12), e22731.
- KP-PBA (Korpus Pembelajar PBA STAI Kendal Ngawi) – Korpus hasil penelitian (2025).
- Sakawi, R. (2018). العربية اللغة وتعليمية اللغوية الازدواجية (Diglossia and Arabic Language Teaching). Dar al-Manzur.
- Setiawan, E. (2018). *Definisi dan Jenis Korpus*. (Dalam Ahsanuddin et al., 2021).