

مجموعة بحوث

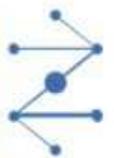
مؤتمرا دوليا واستكتاب
لتجاهات اللغة العربية في العصر الرقمي
تعليميا آديا برمجيا



١٠ أغسطس ٢٠١٧



مركز البحوث
والتواصل المعرفي
Center for Research &
Intercommunication Knowledge



ازحاد مدرسي اللغة العربية باندونيسيا
ARABIC TEACHERS ASSOCIATION OF INDONESIA
KATAN PENGAJAR BAHASA ARAB SE-INDONESIA

مجموعة بحوث

مؤتمر دولي واستكتاب
اتجاهات اللغة العربية في العصر الرقمي
(تعليميا، أدبيا، برمجيا)

١٠ أغسطس ٢٠١٧



مركز البحوث
والتواصل المعرفي
Center for Research &
Intercommunication Knowledge



مجموعة بحوث مؤتمر دولي :

اتجاهات اللغة العربية في العصر الرقمي تعليميا، أدبيا، برمجيا

بجامعة محمدية يوكياكرتا – إندونيسيا أغسطس ٢٠١٧

المحررون :

الأستاذ الدكتور توفيق أحمد دردي

الأستاذ الدكتور سعيدو

الدكتور تولوس مصطفى

تصميم المحتوى:

تلقيس نور دينتو

الطبعة الأولى: ٢٠١٧ م

ردمك (ISBN) : 978-602-7577-99-2

LP3M Press

Kampus Terpadu UMY, Gedung D. Lt 2. Jalan Lingkar Selatan, Tamantirto Kasihan, Bantul,
Yogyakarta 55183 Telp. (0247) 387656

الفهرس

- ١ التطبيقات الحاسوبية التعليمية الموجهة لتعليم اللغة العربية للطفل الأهميَّة والأداء
- ١٧ القيم التعليمية في اللغة العربية
- ٢٧ التعليم التكنولوجي مبدأ أساسي في التعليم
- ٣٢ التطور التعليمي في العصر الرقمي: تحديات وآمال
- ٥٠ تعليم اللغة العربية في جامعة مالانج الحكومية بمساعدة الحاسوب الآلي
- ٦١ الوسائل الحديثة في تعليم اللغة العربية للناطقين بغيرها في المدارس الإسلامية الحكومية بنجرماسين
- ٧٩ الوسائل الرقمية في تعلم اللغة العربية لمحدودي القدرة البصرية
(دراسة تحليلية في جامعة سونن كاليجاكا الإسلامية الحكومية بوكياكرتا إندونيسيا)
- ٨٦ المنهج الأيسري في تعليم وتعلم اللغة العربية لغير الناطقين بها (التجربة الأسمرية)
- ٩٩ أسس بناء منهج تعليم اللغة العربية للطلبة
- ١١١ تعلم العربية على سبيل السكن لطلاب كلية عامة
- ١١٨ تدريس النحو الأول على ضوء التحليل التقابلي
- ١٣٥ واستخدامها في تعلم اللغة العربية (*Animation*) تاريخ تطوير الرسوم المتحركة
- ١٤١ في تعليم النحو *Learning Together* تأثير تطبيق طريقة
- ١٧٠ معوّقات استخدام الوسائط المتعدّدة في تعليم اللغة العربية
- ١٨٠ تعلم مهارتي الاستماع والكلام باستخدام يوتوب لدى المعلمين وطلبة شعبة تعليم اللغة العربية
بجامعة بالنكارايا الإسلامية الحكومية
- ١٩٣ استخدام الأفلام الكرتونية لتنمية مهارة الاستماع في قسم تعليم اللغة العربية بجامعة أنتساري
الإسلامية الحكومية بنجرماسين
- ٢١٤ ترقية إدراك الطلاب في تعليم اللغة العربية عن طريق القرص التفاعلي فلاش (*Flash*)
- ٢٢٠ دور التكنولوجيا في تدريس اللغة العربية
- ٢٣٣ تصميم تعلم اللغة العربية باستخدام التكنولوجيا
- ٢٤٤ تذوق اللهجات العربية عبر التعليم الإلكتروني: تدريس كتاب العربية المعاصرة نموذجاً
- ٢٥٨ مثيرات الفيديو في تدريس الاستماع
- ٢٧٤ التعبير عن الحب مقارنة بين ما ورد في سورة يوسف وما ورد في البردة البوصيرية الشريفة
- ٢٩٣ حي بن يقظان في كتاب الأدغال: دراسة حديثة
- ٣٠١ الكناية وأمكانية ترجمتها إلى الإندونيسية
- ٣٠٦ استنتاج القيم التربوية عن طريق المحاولة على تمركز المصطلحات العربية في ترجمة الكتب الإسلامية
إلى اللغة الإندونيسية.
- ٣١٦ النموذجية لتعليم اللغة العربية في مدرسة محمديّة الكوثر الابتدائية الإسلامية المزدوجة بكارناسورا

- ٣٢٦ القيم الاجتماعية و القيم الإنسانية في قصة "ابن الملك وأصحابه" من حكايات كلية و دمنة, قراءة تفكيكية
- ٣٣٢ اشتها ب قيمة المعرفة الموضوعية لتنمية المعهد الإسلامي للإندونيسيا (دراسة في المعهد الإسلامي التقليدي في ولاية بانتين و جاوة غربية و جاكرتا)
- ٣٥٠ تطوير كتاب التدريبات اللغوية على أساس كتاب مدارج الدروس العربية الجزء الثاني
- ٣٦٨ دينميكية برنامج معهد الطلبة لقسم تعليم اللغة العربية بجامعة سوراكرتا الإسلامية الحكومية
- ٣٨١ أنواع الدلالة أو المعنى في علم السيمانتك
- ٣٩٣ مهارات الفهم في قراءة النص و اللفظ العربي للناطقين بغيرها في إن دون سي ا
- ٤٢١ إعادة صياغة النحو العربي من منظور اللسانيات الحديثة و أهميتها في تيسير تعليمه للناطقين بغير العربية
- ٤٣٠ تعليم اللغة العربية المساعد بالأجهزة المعلوماتية في مركز التعلم الذاتي
- ٤٤٢ الفكر البلاغي في المعاني عند السكاكي في كتابه مفتاح العلوم: دراسة بلاغية
- ٤٥٦ التحليل الأدبي لقصيدة الطين لإيليا أبي ماضي
- ٤٦٨ إمكانية اتصال الأفعال العربية بحروف الجر و تأثيرها في المعنى
- ٤٨٠ أهمية علم الدلالة في فهم معاني ألفاظ القرآن الكريم
- ٤٩٢ تعليم اللغة العربية في إندونيسيا: من التكوين إلى التنظيم (تاريخية منهجية)
- ٥٠٤ التغيرات المعنوية للأفعال العربية في ترجمتها إلى اللغة المادورية (دراسة تحليلية دلالية في كتاب بداية الهداية لبغداد المالكي)
- ٥١٨ اتجاهات في ترقية جودة تدريس اللغة العربية
- ٥٣١ قضية الأخطاء اللغوية و تحليلها عند المحادثة العربية
- ٥٥٠ المشترك الصرفي في القرآن الكريم و ترجمته باللغة الإندونيسية
- ٥٦٥ الأخطاء اللغوية الشائعة في استعمال العربية في المجتمع الملايوي
- ٥٧٤ استراتيجيات تعلم مهارة الكتابة الأكاديمية العربية (دراسة وصفية بقسم تعليم اللغة العربية جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج)
- ٥٨٥ مزالق تطبيق (نصر حامد أبو زيد) مناهج النقد الأدبي الحديثة في كتابه ((نقد الخطاب الديني))
- ٥٩٧ مشكلات تعليم اللغة العربية في الجامعات الإندونيسية (دراسة عن تطوير كفاءة المعلم في تعليم المهرات اللغوية)
- ٦٠٩ Al-Qur'an dan Al-Hadits, Sumber Nilai-Nilai Pendidikan yang Tetap Relevan di Era Digital
- ٦١٧ Cara Pengucapan Bunyi-bunyi Konsonan Arab Bagi Penutur Bahasa Jawa
- ٦٢٦ تقنين اختبار المقدرة العربية لطلاب الجامعات العامة بإندونيسيا
- ٦٣٢ Equivalence At Word Level In Creative Writing Process Based On The WSQA (Word, Sentence, Question, And Answer) Method: Case Studies In The Translation Of Indonesian Short Stories Into Arabic Language
- ٦٣٨ *Tarakh Sib Sya'iah* According to Arabic Student of Indonesia University of Education Bandung)
- ٦٤٩ Kekhasan Stilistika Arab: *Al-Iltifaat, Al-Tanaawub, dan Al-Taqdiim Wa Al-Ta'khiir*

| | |
|-----|--|
| ٦٦٣ | Beberapa Model dan Karakteristik Korpus Bahasa Arab sebagai Acuan Penyusunan Korpus Bahasa Arab di Indonesia |
| ٦٧٦ | Ealing With Anxiety From New Students Perspective |
| ٦٨٥ | Origins Or Cognates of Negative Terms In World Languages: A Radical Linguistic Theory Approach |
| ٧٠٧ | Leksikon untuk Unta dalam Bahasa Arab Kajian Etnosemantik |
| ٧٢٥ | Pembelajaran <i>Qawa'id Al-Lughah Al-'Arabiyah</i> (<i>Sharf</i> dan <i>Nahwu</i>) dengan Strategi Pembelajaran Aktif di Jurusan Pendidikan Bahasa Arab Fakultas Ilmu Tarbiyah dan Keguruan IAIN Surakarta Tahun Akademik 2015-2016)". |
| ٧٣٤ | Evaluasi Buku Ajar Bahasa Arab SMA Pesantren Modern Immim Putera Makassar |
| ٧٤٦ | Pembelajaran Ketrampilan Berbahasa Arab Integral |
| ٧٥٧ | Multiple Categorical Independent Variables and Factorial Design Two Way Anova Simulation |
| ٧٧٠ | العلاقة بين العلوم العربية واستنباط الأحكام الشرعية في ضوء علم أصول الفقه |

BEBERAPA MODEL DAN KARAKTERISTIK KORPUS BAHASA ARAB SEBAGAI ACUAN PENYUSUNAN KORPUS BAHASA ARAB DI INDONESIA

Nur Hizbullah & Fazlur Rachman

Universitas Al Azhar Indonesia

nurhz@uai.ac.id

fazlurrachman@uai.ac.id

ABSTRAK

Makalah ini bertujuan mendeskripsikan dan mengklasifikasikan beberapa model korpus bahasa Arab yang ada di dunia berdasarkan karakteristiknya. Aspek yang dideskripsikan adalah keragaman dan kedalaman isi serta kemanfaatan korpus dalam kaitannya dengan keluasan cakupan kajian bahasa Arab yang menjadikannya sebagai sumber data penelitian. Lebih lanjut, korpus-korpus yang ada diklasifikasikan untuk memetakan kategorinya. Hasil kajian ini mengungkapkan bahwa setiap korpus memiliki karakteristik yang beragam sesuai dengan tujuan penyusunannya dan dari klasifikasi yang ada belum semua jenis klasifikasi tersedia korpusnya. Para peneliti bahasa tinggal memilih korpus mana yang relevan sebagai sumber data penelitiannya. Lebih lanjut, gambaran itu memberikan informasi sekaligus acuan bagi kemungkinan penyusunan korpus bahasa Arab dengan jenis yang lebih beragam serta pertimbangan bagi upaya penyusunan korpus bahasa Arab di Indonesia.

Kata kunci : korpus bahasa Arab, karakteristik korpus, klasifikasi korpus

PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi yang pesat ikut mempengaruhi perkembangan berbagai disiplin ilmu pengetahuan, tak terkecuali disiplin ilmu dan kajian bahasa. Salah satu hal yang penting dalam pengaruh perkembangan teknologi informasi dan komunikasi terhadap ilmu dan kajian bahasa adalah digitalisasi. Semua bentuk dan produk bahasa dapat diubah menjadi data digital dan memungkinkan untuk dianalisis dengan berbagai teknik, metode, dan pendekatan mutakhir yang kian beragam. Proses digital itu pun melahirkan berbagai hasil penelitian yang inovatif dan dinamis. Tidak hanya itu, data kebahasaan yang menjadi objek penelitian dapat terdokumentasikan dengan baik dalam bentuk data digital.

Data kebahasaan digital itu pada tahap selanjutnya diolah lagi secara sistematis sehingga menjadi korpus bahasa. Dalam konteks linguistik, keberadaan korpus bahasa digital itu melahirkan suatu “cabang” ataupun “anak cabang” baru yang bernama linguistik korpus. Menurut Adolphs (2006: 137), linguistik korpus mengkaji bahasa dengan mengandalkan data kebahasaan yang natural dan otentik. Data kebahasaan itu dapat berupa data tertulis maupun data lisan yang lebih dulu ditranskripsikan. Data kebahasaan digital itu mencakup keseluruhan tingkatan, mulai dari fonem, morfem, kata, frase, klausa, kalimat, sampai dengan wacana. Meskipun pada awalnya data itu hanya dikaji dalam lingkup linguistik murni dan teoretis, pada perkembangan selanjutnya, kajian yang memanfaatkan korpus bahasa meluas ke lingkup linguistik terapan, seperti pengajaran, dan bahkan bidang kajian yang berkaitan erat dengan bahasa, seperti penerjemahan, susastra, dan sebagainya.

Keluasan lingkup kajian tersebut mendorong kepada penyusunan korpus bahasa dalam berbagai konteks yang beragam, mulai dari klasifikasi, keragaman, dan kedalaman isinya. Ada korpus yang memang dirancang khusus untuk bidang kajian kebahasaan tertentu saja, dan ada juga korpus lain yang memang mencakup seluas mungkin aspek bahasa yang menjadi sumber korpus yang bersangkutan.

Banyak bahasa di dunia menggunakan prinsip itu dalam penyusunan korpusnya, tak terkecuali bahasa Arab. Lebih lanjut, penelitian ini akan mendeskripsikan beberapa model korpus bahasa Arab yang ada di dunia dan membandingkannya untuk memperoleh gambaran tentang persamaan dan perbedaan masing-masing korpus. Lebih lanjut, gambaran tentang beragam korpus itu diharapkan menjadi acuan bagi penguatan pemikiran dan pertimbangan ilmiah strategis bagi penyusunan korpus bahasa Arab di Indonesia. Langkah ini menjadi penting mengingat sejarah, dinamika, potensi, dan cakupan keberadaan bahasa Arab di Indonesia serta pembelajarannya sangat menarik untuk digali. Bahasa Arab memiliki sejarah dan akar kuat dalam masyarakat Indonesia, terutama dalam kaitannya sebagai bahasa agama Islam. Keberadaan dan perkembangannya di tengah masyarakat Indonesia sangat dinamis. Dari situ, potensi pemanfaatannya dalam berbagai bidang kehidupan masyarakat amat terbuka luas, dan, karenanya, pembelajarannya menjadi penting. Korpus bahasa Arab di Indonesia nantinya akan menjadi “bukti rekaman” bagi perjalanan dan perkembangan bahasa Arab dan menjadi objek “baru” yang otentik dalam kajian bahasa Arab di Indonesia.

KARAKTERISTIK DAN KLASIFIKASI KORPUS

Ihwal suatu korpus dapat dilihat berdasarkan karakteristiknya. Al-Sulaiti (2010) mendata sejumlah korpus dan mencantumkan beberapa butir karakteristik sebagai informasi dasar bagi kumpulan korpus yang diidentifikasikannya, yaitu :

1. sumber atau pemilik korpus yang bersangkutan,
2. ragam bahasanya,
3. ukuran atau besaran isi korpus yang umumnya berjumlah jutaan kata,
4. tujuan penyusunan korpus tersebut,
5. materialnya, dan
6. metode akses terhadap korpus, ada yang berbayar, cuma-cuma, dan percobaan (*trial*).

Seiring dengan perkembangan korpus dan ragamnya, ikut beragam pula model klasifikasi yang dilakukan oleh beberapa ahli. Makalah ini akan menggunakan dua model, yaitu klasifikasi yang dibuat oleh Nesselhauf (2011: 3-4) dan klasifikasi versi Sketch Engine,¹ salah satu mesin penyedia korpus multibahasa yang terkemuka.

Menurut Nesselhauf (2011: 3-4), dari sekian banyak ragam korpus yang ada di dunia, dapat diklasifikasikan sebagai berikut.

1. Korpus general/referensi dan—“kebalikannya”—korpus khusus. Klasifikasi ini berkenaan dengan suatu jenis korpus yang dibuat untuk menggambarkan representasi suatu bahasa secara general ataupun spesifik berdasarkan topik-topik tertentu. Korpus ini berisi data bahasa ragam tulis dan ragam lisan yang ditranskripsikan serta diambil dari dari berbagai jenis teks yang beragam maupun khusus pada bidang tertentu. Salah satu contoh dari ragam korpus umum yang sangat populer dalam hal ini adalah British National Corpus (BNC).
2. Korpus historis dan—kebalikannya—korpus bahasa modern. Korpus jenis ini menggambarkan fenomena kebahasaan yang berlaku pada suatu kurun waktu dalam masyarakat penuturnya. Ada korpus yang secara khusus menggambarkan periode klasik dari suatu bahasa dan ada pula korpus yang menggambarkan fakta nyata kebahasaan di masa modern. Contoh korpus jenis ini adalah Helsinki Corpus.
3. Korpus regional, yang disusun untuk mewakili variasi bahasa dari suatu kawasan tertentu. Wellington Corpus of Written New Zealand English (WCNZE) adalah salah satu jenis korpus yang masuk klasifikasi ini.
4. Korpus pembelajar dan—kebalikannya—korpus penutur asli. Korpus pembelajar berisi fakta bahasa yang umumnya diproduksi sebagai bahasa kedua oleh pembelajar bukan penutur asli bahasa tersebut. Sementara itu, korpus penutur asli berisi tulisan ataupun tuturan yang diproduksi oleh penutur asli suatu bahasa. Contoh korpus pembelajar dalam hal ini adalah International Corpus of Learner English (ICLE).

¹ <https://www.sketchengine.co.uk/user-guide/user-manual/corpora/corpus-types/>, diakses tanggal 17 Juni 2017

5. Korpus multilingual dan—kebalikannya—korpus ekabahasa. Korpus ini memiliki keunikan karena merepresentasikan sedikitnya dua bahasa dan bersumber dari suatu ragam teks yang sama. Korpus ini umumnya digunakan untuk keperluan kajian kontrastif antarbahasa.
6. Korpus lisan dan—kebalikannya—korpus tulisan dan/atau campuran lisan-tulisan. Korpus jenis ini secara khusus merepresentasikan ragam bahasa lisan dari suatu masyarakat penutur bahasa dan/atau ragam bahasa tulis mereka ataupun kedua-duanya.
7. Korpus ortografis dan—kebalikannya—korpus beranotasi. Korpus ini dibedakan berdasarkan bukan pada sumbernya, tapi berdasarkan pengolahannya dalam analisis bahasa. Korpus beranotasi, dalam hal ini, sudah tidak lagi berisi fakta asal suatu bahasa, tetapi sudah memuat anotasi yang ditambahkan terhadap fakta kebahasaan aslinya.

Lebih lanjut, Sketch Engine membagi sekian banyak jenis korpus ke dalam kategori berikut

1. Korpus monolingual. Korpus jenis ini adalah yang paling banyak ditemukan karena umumnya korpus terfokus pada satu bahasa. Korpus ini biasanya digunakan untuk analisis satuan kata dan variasi kolokasinya dalam konteks perkembangan bahasa lintas masa.
2. Korpus paralel. Korpus ini terdiri atas dua korpus monolingual. Satu teks berbahasa sumber dan satu teks lainnya merupakan terjemahan teks sebelumnya dalam bahasa sasaran. Secara teknis, korpus jenis ini menyandingkan dua bahasa dalam segmen-segmen tertentu, biasanya dalam cakupan kalimat ataupun paragraf. Dalam pemanfaatan korpus ini, peneliti bisa mengamati suatu kata ataupun ungkapan dalam bahasa sumber lalu melihat terjemahan ataupun padanannya dalam bahasa sasaran. Korpus ini amat membantu dalam penelitian penerjemahan utamanya.
3. Korpus multilingual. Korpus jenis ini hampir sama dengan korpus paralel. Hanya saja, ragam bahasa yang dimuat bisa lebih dari dua. Pendekatan teknis penyusunannya pun sama, begitu pula segmen pembagian unsur-unsur di dalam teks. Peneliti bisa lebih leluasa memilih bahasa mana yang akan diteliti dari antara sekian bahasa yang tersedia dalam korpus.
4. Korpus pembelajar. Maksud dari korpus pembelajar adalah korpus yang berisi teks yang diproduksi oleh pembelajar suatu bahasa. Korpus jenis ini umumnya dimanfaatkan dalam konteks pengajaran bahasa dan penelitian di bidang itu, antara lain untuk analisis problem pembelajaran bahasa dan analisis kesalahan berbahasa, terutama dalam pembelajaran bahasa asing ataupun bahasa kedua.
5. Korpus diakronis. Korpus ini memuat teks dari berbagai zaman dan biasanya digunakan secara khusus untuk kajian perkembangan maupun perubahan bahasa.
6. Korpus multimedia. Korpus ini memuat teks yang dilengkapi dengan materi suara, video, ataupun media lain.

Kesemua klasifikasi yang disusun oleh Sketch Engine itu lebih lanjut dapat diklasifikasikan lebih khusus lagi. Itu dilakukan dengan cara membuat subkorpus tersendiri berdasarkan kategori-kategori tertentu, seperti area, ranah, topik, dan sebagainya.

Apabila dicermati lebih lanjut, dua klasifikasi tersebut memiliki kemiripan dalam beberapa jenis korpus. Oleh karena itu, untuk menjadi dasar dan kerangka bagi klasifikasi korpus dalam makalah ini, berikut sintesis klasifikasi dari keduanya berdasarkan kriteria yang mendasari klasifikasinya.

1. Berdasarkan konten korpus :
 - a. korpus umum
 - b. korpus khusus
2. Berdasarkan periode bahasa disebut korpus diakronis, yang terdiri atas :
 - a. korpus historis
 - b. korpus modern
3. Berdasarkan penutur bahasa :
 - a. korpus pembelajar dari kalangan penutur asing

- b. korpus penutur asli suatu bahasa ibu
- c. korpus regional tentang suatu bahasa di suatu wilayah
- 4. Berdasarkan variasi bahasa konten korpus :
 - a. korpus monolingual atau ekabahasa
 - b. korpus paralel yang terdiri atas dua bahasa atau lebih
- 5. Berdasarkan ragam bahasa :
 - a. korpus bahasa lisan
 - b. korpus bahasa tulisan
 - c. korpus campuran bahasa lisan dan tulisan
- 6. Berdasarkan jenis materialnya :
 - a. korpus teks
 - b. korpus teks dengan tambahan materi audio
 - c. korpus teks dengan tambahan materi audiovisual
- 7. Berdasarkan tulisannya :
 - a. korpus ortografis
 - b. korpus beranotasi

BERBAGAI KORPUS BAHASA ARAB YANG SUDAH ADA

Pada bagian ini akan ditampilkan tabel data tentang korpus bahasa Arab yang sejauh ini sudah ada dan berhasil diidentifikasi. Data ini utamanya bersumber dari Al-Sulaiti (2010) dan Sketch Engine (2012) serta beberapa sumber lain yang menyediakan data korpus bahasa Arab secara umum maupun korpus Al-Quran secara khusus.

| No | Nama dan Tahun Korpus | Karakteristik | | | | | | | Metode Akses |
|----|------------------------------------|--------------------------------|--------------|-----------------------|---|--|----------------|--|--------------|
| | | Sumber/Pemilik | Ragam Bahasa | Ukuran/Besaran | Tujuan Penyusunan | Sumber/Material Data | Metode Akses | | |
| 1 | Al-Hayat Corpus (2002) | ELRA | Tulisan | 18,6 juta kata | <i>Language Engineering and Information Retrieval</i> | Koran Al-Hayat (Lebanon) | Berbayar | | |
| 2 | Al-Nahar Corpus (2001) | ELRA | Tulisan | 140 juta kata | Penelitian Umum | Koran An-Nahar (Lebanon) | Berbayar | | |
| 3 | Arabic Gigaword (2003) | University of Pennsylvania LDC | Tulisan | ± 400 juta kata | <i>Natural language processing, information retrieval, language modelling</i> | Agence France Presse, Al-Hayat news agency, An-Nahar news agency, Xinhua news agency | Berbayar | | |
| 4 | Arabic Newswire Corpus (1994) | University of Pennsylvania LDC | Tulisan | 80 juta kata | Pendidikan dan pengembangan teknologi | Agence France Presse, Xinhua News Agency, and Umma Press | Berbayar | | |
| 5 | Arabic Web 2012 (arTenTen12) | TenTen corpus family | Tulisan | 11.2 juta kata | Penelitian umum | Berbagai sumber dari Internet | Berbayar | | |
| 6 | arabiCorpus Egyptian Colloquial | arabiCorpus | Tulisan | 164,457 kata | Penelitian Umum | Ebsite penyedia layanan chat, wawancara dengan Husni Mubarak, dan serial TV | Tidak berbayar | | |
| 7 | arabiCorpus Modern Literature | arabiCorpus | Tulisan | 1,1 juta kata | Penelitian umum | Novel-novel dari beberapa negara Timur Tengah | Tidak berbayar | | |
| 8 | arabiCorpus newspapers | arabiCorpus | Tulisan | 135,3 juta kata | Penelitian umum | Sepuluh koran berbahasa Arab dari Timur Tengah | Tidak berbayar | | |
| 9 | arabiCorpus Nonfiction | arabiCorpus | Tulisan | 82,6 juta kata | Penelitian umum | Buku-buku nonfiksi berbahasa arab | Tidak berbayar | | |
| 10 | arabiCorpus Premodern | arabiCorpus | Tulisan | 9,1 juta kata | Penelitian umum | Literatur Arab pramodern (Al-Quran, hadist, fiksi, gramatika arab, dan filsafat) | Tidak berbayar | | |
| 11 | Broadcast News Speech (2000) | University of Pennsylvania LDC | Lisan | > 110 siaran | Speech recognition | Siaran berita Radio Voice of America. | Berbayar | | |
| 12 | Buckwalter Arabic Corpus 1986-2003 | Tim Buckwalter | Tulisan | 2.5 - 3 milyar kata | Leksikografi | Konten umum di internet | Berbayar | | |
| 13 | CALLFRIEND Corpus (1995) | University of Pennsylvania LDC | Lisan | 60 Percakapan telepon | Pengembangan " <i>language identification technology</i> " | Percakapan via telepon oleh penutur asli bahasa Mesir | Berbayar | | |

Tabel 1. Karakteristik korpus Arab

| No | Nama dan Tahun Korpus | Karakteristik | | | | | | | Metode Akses |
|----|--|--|-------------------|-----------------------------------|--------------------------------------|---|--------------|----------------|--------------|
| | | Sumber/Pemilik | Ragam Bahasa | Ukuran/Besaran | Tujuan Penyusunan | Sumber/Material Data | Metode Akses | | |
| 15 | CLARA (1997) | Charles University, Prague | Tulisan | 50 juta kata | Untuk Leksikografi | Majalah, buku-buku, bahan-bahan dari internet dari tahun 1975-sekarang | Metode Akses | Berbayar | |
| 16 | Classical Arabic Corpus (CAC) (2004) | UMIST, UK | Tulisan | 5 juta kata | Lexical analysis research | www.muhammad.org and www.alwaraq.com | Metode Akses | Berbayar | |
| 17 | Corpus of Contemporary Arabic (CCA) 2004 | University of Leeds, UK | Lisan dan Tulisan | ± 1 juta kata | TAFI | Website majalah online | Metode Akses | Tidak berbayar | |
| 18 | DINAR Corpus (2000) | Nijmegen Univ., SOTETEL-IT, coordination of Lyon2 Univ | Tulisan | 10 juta kata | Lexicography, general research, NLP | Tidak ada keterangan | Metode Akses | Berbayar | |
| 19 | E-A Parallel Corpus | University of Kuwait | Tulisan | 3 juta kata | Pengajaran terjemah dan leksikografi | Publikasi dari Kuwait National Council | Metode Akses | Berbayar | |
| 20 | Egypt (1999) | John Hopkins University | Tulisan | - | Mesin Penerjemah | Korpus paralel Al-Qur'an dalam Bahasa Inggris dan Arab | Metode Akses | Berbayar | |
| 21 | Essex Arabic Summaries Corpus (EASC 1.0) | Mahmoud El-Haj, University of Essex, UK | Tulisan | - | - | 153 artikel berbahasa Arab dan 765 ringkasan artikel | Metode Akses | Tidak berbayar | |
| 22 | General Scientific Arabic Corpus (2004) | UMIST, UK | Tulisan | 1,6 juta kata | Penelitian susunan kata bahasa arab | http://www.kisr.edu.kw/science/ | Metode Akses | Berbayar | |
| 23 | International Corpus of Arabic (ICA) 2008-2009 | Bibliotheca Alexandrina (BA) | Tulisan | 100 juta kata | Penelitian linguistik umum | Konten Internet dari beberapa negara Arab | Metode Akses | Berbayar | |
| 24 | King Saud University Arabic Speech Database | Mansour Alsulaiman et.al | Tulisan | - | Speaker recognition research | 590 jam rekaman audio dari 269 orang laki-laki dan perempuan dari 28 negara arab dan non arab | Metode Akses | Berbayar | |
| 25 | Leuven Corpus (1990-2004) | Catholic University Leuven, Belgium | Lisan dan Tulisan | 2,5 - 3 juta kata | Penyusunan kamus Arab-Belanda v.v | Data-data dari Internet, radio & TV, Buku bahan ajar sekolah dasar | Metode Akses | Berbayar | |
| 26 | Multilingual Corpus 2004 | UMIST, UK | Tulisan | 10,7 juta kata (Arab 1 juta kata) | Bahan terjemahan | Website khusus konten IT | Metode Akses | Berbayar | |
| 27 | NijmegenCorpus (1996) | Nijmegen University | Tulisan | Lebih dari 2 juta kata | Penyusunan kamus Arab-Belanda v.v | Majalah dan Cerita Fiksi | Metode Akses | Berbayar | |
| 28 | Opus2 Arabic | OPUS 2 parallel corpora | Tulisan | - | Penelitian Umum | data korpus multilingual dengan 40 bahasa | Metode Akses | Tidak berbayar | |
| 29 | SOTETEL Corpus | SOTETEL-IT, Tunisia | Tulisan | 8 juta kata | Leksikografi | Literature, academic and journalistic material | Metode Akses | Berbayar | |

Tabel 2. Karakteristik korpus Arab (lanjutan)

Berikut ulasan tentang 29 korpus yang berhasil dihimpun datanya seperti tercantum dalam tabel tersebut berdasarkan karakteristik yang dirumuskan oleh Al-Sulaiti (2010).

1. Berdasarkan sumber atau pemilik korpus yang bersangkutan, terlihat bahwa pada umumnya satu lembaga memiliki satu korpus tersendiri. Tercatat ada 12 lembaga yang memiliki satu korpus. Yang menarik adalah bahwa ada sejumlah lembaga yang memiliki lebih dari satu korpus. ArabiCorpus, misalnya, bahkan memiliki lima jenis korpus. Ada pula lembaga yang memiliki satu korpus tersendiri tetapi juga memiliki satu korpus yang dimiliki bersama lembaga lain, yaitu Nijmegen University yang bermitra dengan lembaga SOTETEL-IT dan Lyon 2 University. Menurut hemat kami, kerja sama kolaboratif semacam ini dapat dijadikan acuan model bagi pengembangan korpus karena setiap lembaga dapat berkontribusi banyak sumber yang dimilikinya dan nilai lebih lainnya adalah semakin tingginya tingkat variasi isi korpus yang disumbang oleh lembaga-lembaga yang bermitra.
2. Apabila dilihat dari ragam bahasanya, pada umumnya korpus bahasa Arab berisi bahasa ragam tulis. Hanya satu korpus yang berisi data bahasa lisan dan ada dua korpus yang berisi kombinasi antara dua bahasa lisan dan tulisan. Bentuk tulisan sendiri merupakan standar baku dalam penyusunan korpus karena lazimnya korpus berbentuk data tertulis digital yang diproses melalui sejumlah mekanisme tertentu, seperti konversi *file*, sebelum diolah dengan mesin atau aplikasi pengolah korpus. Namun, seiring waktu dan perkembangan teknologi, data lisan yang didigitalkan ada yang tidak perlu dikonversi menjadi bentuk tertulis karena sudah tersedianya mesin atau alat pengolah korpus suara digital. Hanya saja, aplikasi pengolah korpus lebih banyak diarahkan untuk mengolah korpus tulis sebagai yang paling populer digunakan daripada aplikasi pengolah korpus suara.
3. Mengenai ukuran atau besaran isi korpus bahasa Arab, umumnya sudah berjumlah dalam satuan juta kata, bahkan ada yang diklaim mencapai satuan miliar kata. Korpus semacam ini ternyata mengambil data dari rentang waktu yang sangat panjang, yaitu 17 tahun, mulai 1986 s.d 2003, seperti yang disusun oleh Buckwalter (2013). Hanya satu korpus yang jumlahnya hanya mencapai hitungan ratusan ribu kata. Adapun korpus bahasa lisan yang disebutkan bersumber dari percakapan lisan via telepon jumlahnya tidak (atau mungkin belum) dihitung dari jumlah kata, tapi dihitung dari jumlah percakapan telepon yang direkam dan dijadikan korpus.
4. Dalam hal tujuan penyusunan, semua korpus tersebut disusun untuk kepentingan penelitian ilmiah kebahasaan. Hanya saja, orientasi penyusunannya berbeda-beda bergantung pada sumber ataupun material korpus itu sendiri. Dari penelusuran terhadap sumber penyedia korpus, diketahui bahwa orientasi penelitian linguistik umum amat dominan. Dalam hal ini, pemanfaatan korpus sepenuhnya menjadi kepentingan peneliti yang ingin menggunakan data yang dalam korpus. Sebagian yang lain bertujuan untuk kepentingan penelitian leksikografi, penerjemahan, pengajaran bahasa, dan pengembangan model analisis digital terhadap bahasa lisan yang dikombinasikan dengan pendekatan teknologi informasi dan komunikasi. Yang menarik dalam bagian ini adalah bahwa ternyata penyusunan korpus tidak hanya dilakukan oleh ahli ataupun peneliti bahasa, tetapi juga melibatkan ahli teknologi informasi dan komunikasi. Hal ini memberi keuntungan tersendiri karena sistem dan infrastruktur korpus menjadi lebih baik dan mapan.
5. Terkait dengan sumber dan materialnya, korpus-korpus bahasa Arab yang sudah ada ternyata amat beragam sekali. Dalam pembagian yang lebih mudah, korpus yang bersumber dari data kebahasaan klasik umumnya mengambil data dari sumber cetak yang kemudian ditulis ulang dan digitalkan, sementara korpus yang mutakhir lebih banyak memanfaatkan data tulis yang umumnya sudah dalam bentuk digital berbasis internet. Kemajuan dan perkembangan dunia internet amat krusial perannya dalam hal ini. Hanya saja, korpus bahasa lisan bersandar pada rekaman suara konvensional yang pada prosesnya kemudian mengalami digitalisasi.
6. Soal metode akses terhadap korpus, mayoritas korpus bisa diakses dengan berbayar. Sebagian kecil di antaranya menyediakan fasilitas *trial* atau coba cuma-cuma. Hanya sembilan korpus yang dapat diakses tanpa berbayar, namun demikian sebagiannya meminta prosedur pendaftaran keanggotaan terlebih dulu bagi pengunjung lamannya.

Setelah penjabaran tabel dan uraian tentang karakteristik tentang korpus yang ada, berikut akan dijabarkan tabel dan menyusul uraian tentang klasifikasi korpus-korpus bahasa Arab tersebut.

| No | Nama Korpus | KLASIFIKASI KORPUS | | | | | | | | | | | | | | | | |
|----|--|--------------------|--------|----------------|--------|------------------|--------------|----------|----------------|---------|--------------|---------|----------|----------------|--------------|--------------------|----------------|------------|
| | | Konten | | Periode Bahasa | | Penutur | | | Variasi Konten | | Ragam Bahasa | | | Jenis Material | | | Format Tulisan | |
| | | Umum | Khusus | Historis | Modern | Pembelajar asing | Penutur asli | Regional | Mono-lingual | Paralel | Lisan | Tulisan | Campuran | Teks | Teks + Audio | Teks + Audiovisual | Ortografis | Beranotasi |
| 1 | Al-Hayat Corpus (2002) | √ | | | √ | | √ | | √ | | | √ | | √ | | | | |
| 2 | An-Nahar Corpus | √ | | | √ | | √ | √ | | | | √ | | √ | | | | |
| 3 | Arabic Gigaword (2003) | √ | | | √ | | √ | | √ | | | √ | | √ | | | | |
| 4 | Arabic Newswire Corpus(1994) | | √ | | √ | | √ | | √ | | | √ | | √ | | | | |
| 5 | Arabic Web 2012 (arTenTen12) | √ | | | √ | | √ | | √ | | | √ | | √ | | | | |
| 6 | arabiCorpus Egyptian Colloquial | | √ | | √ | | √ | | √ | | | √ | √ | √ | | | | |
| 7 | arabiCorpus Modern Literature | | √ | | √ | | √ | | √ | | | √ | | √ | | | | |
| 8 | arabiCorpus newspapers | | √ | | √ | | √ | | √ | | | √ | | √ | | | | |
| 9 | arabiCorpus Nonfiction | | √ | √ | | | √ | | √ | | | √ | | √ | | | | |
| 10 | arabiCorpus Premodern | | √ | √ | | | √ | | √ | | | √ | | √ | | | | |
| 11 | Broadcast News Speech (2000) | | √ | | √ | | √ | | √ | | √ | | | | | | | |
| 12 | Buckwalter Arabic Corpus 1986-2003 | √ | | | √ | | √ | | √ | | | √ | | √ | | | | |
| 13 | CALLFRIEND Corpus (1995) | √ | | | | | √ | √ | √ | | √ | | | | √ | | | |
| 14 | CALLHOME Corpus (1997) | √ | | | | | √ | √ | √ | | √ | | | | √ | | | |
| 15 | CLARA (1997) | √ | | √ | | | √ | | √ | | | √ | | √ | | | | |
| 16 | Classical Arabic Corpus (CAC) (2004) | | √ | | √ | | √ | | √ | | | √ | | √ | | | | |
| 17 | Corpus of Contemporary Arabic (CCA) 2004 | | √ | | √ | | √ | | | | | √ | | √ | | | | |

| No | Nama Korpus | KLASIFIKASI KORPUS | | | | | | | | | | | | | | | | |
|----|--|--------------------|--------|----------------|--------|------------------|--------------|----------|----------------|---------|--------------|---------|----------|----------------|--------------|--------------------|----------------|------------|
| | | Konten | | Periode Bahasa | | Penutur | | | Variasi Konten | | Ragam Bahasa | | | Jenis Material | | | Format Tulisan | |
| | | Umum | Khusus | Historis | Modern | Pembelajar asing | Penutur asli | Regional | Mono-lingual | Paralel | Lisan | Tulisan | Campuran | Teks | Teks + Audio | Teks + Audiovisual | Ortografis | Beranotasi |
| 18 | DINAR Corpus (2000) | √ | | | √ | | √ | | √ | | | | √ | | | | | |
| 19 | E-A Parallel Corpus | | √ | | | | √ | | √ | | | | √ | | | | | |
| 20 | Egypt (1999) | √ | | | √ | | √ | | | √ | | | √ | | | | | |
| 21 | Essex Arabic Summaries Corpus (EASC 1.0) | √ | | | √ | | √ | | √ | | | | √ | | | | | |
| 22 | General Scientific Arabic Corpus (2004) | | √ | | √ | | √ | | √ | | | | √ | | | | | |
| 23 | International Corpus of Arabic (ICA) 2008-2009 | √ | | | √ | | √ | | √ | | | | √ | | | | | |
| 24 | King Saud University Arabic Speech Database | | √ | | √ | √ | √ | √ | √ | √ | | | | √ | | | | |
| 25 | Leuven Corpus (1990-2004) | √ | | | √ | | √ | | | √ | | √ | √ | √ | √ | | | |
| 26 | Multilingual Corpus 2004 | √ | | | √ | | √ | | | | √ | | √ | | | | | |
| 27 | NijmegenCorpus | √ | | | | | √ | | | √ | | | √ | | | | | |
| 28 | Opus2 Arabic | √ | | | | | √ | | | √ | | | √ | | | | | |
| 29 | SOTETEL Corpus | √ | | | | | √ | | √ | | | | √ | | | | | |

Tabel 3. Klasifikasi korpus Arab

Tabel 4. Klasifikasi korpus Arab (lanjutan)

Berikut uraian singkat mengenai klasifikasi korpus Arab berdasarkan aspek-aspek yang ada di dalamnya mengacu kepada kombinasi klasifikasi dari Nesselhauf (2011) dan Sketc Engine (2012).

1. Berdasarkan konten korpus : korpus umum dan korpus khusus.
Jumlah korpus umum dan korpus khusus hampir berimbang. Korpus umum berisi materi dan sumber yang relatif lebih beragam dan luas daripada korpus khusus, karena korpus khusus hanya berisi materi tertentu dan dari sumber-sumber terbatas.
2. Berdasarkan periode bahasa disebut korpus diakronis : korpus historis dan korpus modern.
Korpus modern lebih dominan jumlahnya bila dibandingkan dengan korpus historis. Itu bisa terjadi karena mudahnya proses digitalisasi terhadap data-data kebahasaan mutakhir yang umumnya tersebar luas di media internet dan arsip-arsip yang sudah diproses menggunakan komputer. Alhasil, pemerolehan terhadap *soft copy* dari dokumen ataupun data yang diperlukan sangatlah mudah. Berbeda dari korpus modern, korpus historis masih minim jumlahnya meski data kebahasaan klasik di masa lalu tak kalah kaya dan beragam. Bisa jadi, proses pengolahan data cetak ke data digital membutuhkan waktu lama dan teknik yang beragam menjadi alasan bagi masih minimnya korpus historis.
3. Berdasarkan penutur bahasa : korpus pembelajar dari kalangan penutur asing, korpus penutur asli suatu bahasa ibu, dan korpus regional tentang suatu bahasa di suatu wilayah.
Korpus penutur asli tercatat lebih banyak daripada korpus pembelajar-penutur asing dan korpus regional. Namun demikian, di masa yang akan datang, korpus pembelajar dari kalangan penutur dan juga korpus regional, terlebih regional luar asal suatu bahasa—contoh bahasa Arab di Indonesia, harus bertambah jumlahnya. Itu penting antara lain dalam rangka peningkatan kualitas pengajaran bahasa asing di regional ataupun masyarakat bukan penutur bahasa asing itu. Korpus pembelajar memberikan gambaran tentang praktik dan produk pengajaran bahasa asing, sementara korpus regional menggambarkan batasan atau cakupan di mana bahasa itu dipraktikkan dan dikorpuskan. Dalam hal ini pun, korpus penutur asli bisa berguna dalam konteks pengajaran kepada penutur asing sebagai acuan maupun perbandingan dalam proses pembelajaran bahasa.
4. Berdasarkan variasi bahasa konten korpus : korpus monolingual atau ekabahasa dan korpus paralel yang terdiri atas dua bahasa atau lebih.
Sejauh penelusuran terhadap korpus bahasa Arab, dari sekian banyak yang ditemukan, hanya lima korpus yang berisi lebih dari satu bahasa. Selebihnya atau lebih banyaknya merupakan korpus monolingual atau ekabahasa. Situasi ini mestinya membuka cakrawala berpikir dan menumbuhkan minat para pemerhati dan peneliti penerjemahan agar berkontribusi dalam penyusunan korpus multibahasa yang akan bermanfaat luas bagi bidang penelitian lainnya.
5. Berdasarkan ragam bahasa : korpus bahasa lisan, korpus bahasa tulisan, dan korpus campuran bahasa lisan dan tulisan.
Dalam hal ini, korpus bahasa Arab umumnya adalah korpus bahasa tulisan. Hanya sedikit korpus yang berisi bahasa lisan dan campuran antara keduanya. Ini menandakan bahwa penelitian kebahasaan masih dominan mengarah kepada data tulis ketimbang data lisan. Ini pun mestinya membuka mata para pemerhati korpus Arab untuk memanfaatkan peluang bagi penyusunan korpus bahasa lisan yang lebih beragam dan berkualitas.
6. Berdasarkan jenis materialnya : korpus teks, korpus teks dengan tambahan materi audio, dan korpus teks dengan tambahan materi audiovisual.
Hampir senada dengan klasifikasi pada butir 5, korpus teks lebih banyak karena hampir semua korpus memanfaatkan data tertulis dalam bentuk teks. Adapun korpus dengan tambahan audio berkenaan dengan adanya kombinasi isi korpus antara data tulis dan data

lisan. Namun, agak berbeda dari butir sebelumnya, ada pula korpus yang dilengkapi dengan kombinasi data audiovisual, meski jumlahnya amat terbatas.

7. Berdasarkan tulisannya : korpus ortografis dan korpus beranotasi. Korpus jenis ini belum ada dalam khazanah korpus Arab. Sejauh penelusuran kami, belum ada informasi tentang hal ini dalam keterangan yang dimuat dalam laman penyedia korpus bahasa Arab.

PERLUNYA KEBERADAAN KORPUS BAHASA ARAB DI INDONESIA

Keberadaan sejumlah korpus tersebut terus mendorong perkembangan dan pengembangan bidang keilmuan linguistik korpus Arab secara umum. Hal itu berkonsekuensi mendorong tumbuhnya minat para peneliti bahasa Arab untuk memanfaatkan korpus sebagai data penelitian. Lebih dari itu, para peneliti juga mestinya “tertantang” untuk berkontribusi dalam bidang korpus, baik itu menambahkan koleksi data ataupun menyusun korpus tersendiri sesuai orientasi dan kepentingan ilmiah yang diinginkan. Hasil yang diharapkan adalah terjadinya dinamika keilmuan yang semakin semarak di dunia bahasa Arab.

Situasi tersebut seyogyanya juga menjadi perhatian para linguis ataupun pemerhati dan peminat linguistik Arab di Indonesia. Terlebih, keberadaan korpus bahasa Arab di Indonesia mestinya menjadi “angin segar” bagi harapan akan tumbuh dan berkembangnya kajian linguistik Arab modern di Indonesia. Sebagian peneliti bahasa Arab yang relatif sudah akrab dengan pemanfaatan teknologi informasi dan komunikasi, diharapkan mau melirik bidang kajian ini sebagai salah satu pijakan bagi pengembangan kajiannya di masa yang akan datang.

Banyak pertimbangan mengapa perlu (bahkan harus) ada korpus bahasa Arab di Indonesia, antara lain akar sejarah pengajaran dan penggunaan bahasa Arab di Nusantara, luasnya lingkup pembelajaran bahasa Arab di sejumlah jenjang lembaga pendidikan, banyaknya produk-produk kebahasaan baik lisan maupun tulisan yang terdokumentasikan, dan yang lebih utama adalah alasan perlunya pemutakhiran paradigma, metodologi, dan pendekatan dalam penelitian linguistik Arab di Indonesia. Kami tidak sependapat dengan pernyataan bahwa “linguistik Arab sudah selesai”. Pernyataan semacam itu dikhawatirkan dapat mematikan daya nalar dan daya tarik akan fakta bahwa ilmu linguistik sudah sedemikian jauh berkembang dan beradaptasi sangat cepat bahkan masif dengan perkembangan disiplin keilmuan lain, dalam kasus ini adalah teknologi informasi dan komunikasi. Oleh karena itu, para peminat linguistik Arab hendaknya terbuka dan objektif terhadap fakta semacam ini. Penulis ingin mengajak semua pemerhati bahasa Arab akan luas dan masifnya ketersediaan data kebahasaan Arab di Indonesia yang layak dijadikan korpus. Berikut ini adalah beberapa butir ragam materi kebahasaan Arab yang setidaknya mudah untuk diidentifikasi, dijangkau, dan dikumpulkan untuk bahan korpus.

1. Makalah ilmiah yang ditulis dalam bahasa Arab oleh penulis Indonesia yang dipresentasikan dalam seminar IMLA dan seminar lainnya di Indonesia.
2. Penelitian atau karya ilmiah yang ditulis dalam bahasa Arab oleh penulis Indonesia dan dipresentasikan ataupun dipublikasikan di forum internasional dan dimuat dalam prosidingnya.
3. Berkala ilmiah, populer, maupun informatif dalam bahasa Arab yang terbit rutin dan diedarkan di Indonesia.
4. Kajian ilmiah dalam bahasa Arab karya penulis Indonesia yang dimuat di dalam jurnal ilmiah yang dipublikasikan di Timur Tengah.
5. Karya ilmiah maupun karya sastra lepas yang ditulis dalam bahasa Arab oleh penulis Indonesia, baik yang diterbitkan maupun yang tidak diterbitkan.
6. Segala bentuk karya yang ditulis oleh penulis Indonesia dengan huruf Arab meskipun tidak menggunakan bahasa Arab.
7. Segala bentuk teks sastra berbahasa Arab karya penulis Indonesia.
8. Buku pelajaran bahasa Arab karya penulis Indonesia yang digunakan di berbagai jenjang sekolah dan perguruan tinggi di Indonesia.

9. Segala bentuk media pembelajaran bahasa Arab tertulis karya pengajar Indonesia yang dimanfaatkan dalam pengajaran.
10. Segala bentuk rekaman kegiatan pembelajaran, perkuliahan, ceramah, pidato, maupun seminar dalam bahasa Arab berbentuk audio maupun audiovisual. Bentuk seperti ini dapat menjadi dua versi, yaitu data versi audio/audiovisual dan data yang perlu ditranskripsikan lebih dulu sebelum diolah menjadi data digital.
11. Rekaman kegiatan berbahasa Arab oleh mahasiswa ataupun pelajar dalam acara perlombaan, festival, dan sebagainya berbentuk audio maupun audiovisual yang ditranskripsikan dan diolah menjadi data digital.
12. Data koleksi dan kompilasi kosa kata bahasa Indonesia secara terstruktur yang berasal dari bahasa Arab berikut makna ataupun padanannya dalam bahasa Indonesia.
13. Segala bentuk data kebahasaan riil yang dianggap perlu didokumentasikan, diolah, dan diproses menjadi korpus bahasa Arab di Indonesia.

Perlu ditegaskan bahwa keberadaan korpus bahasa Arab semestinya menjadi titik tonggak sejarah baru bagi keberadaan dan perkembangan bahasa Arab di Indonesia di abad modern ini. Dunia internasional, atau setidaknya dunia Timur Tengah, harus tahu betapa kayanya khazanah bahasa Arab di Indonesia, betapa panjangnya sejarah keberadaan dan pengajaran bahasa Arab di negeri ini, dan betapa prospektifnya pengajaran dan perkembangan bahasa Arab ke depan. Itu akan terjadi bila para pengajar, pembelajar, dan peneliti bahasa Arab di Indonesia memiliki idealisme tentang pentingnya melakukan pendokumentasian dalam skala luas terhadap fakta kebahasaan Arab di negeri ini. Lebih lanjut, korpus bahasa Arab sebagai data empirik yang ada di Indonesia nantinya dapat dimanfaatkan seluas-luasnya untuk kepentingan penelitian kebahasaan, seperti di bidang pengajaran, penerjemahan, linguistik itu sendiri, kesusastraan, dan sebagainya.

TERIMA KASIH DAN PENGHARGAAN

- Terhatur ucapan terima kasih dan penghargaan kepada **Universitas Al Azhar Indonesia** dan **Lembaga Penelitian dan Pengabdian Masyarakat (LP2M) UAI** atas dukungan dan bantuan bagi partisipasi kami mewakili Prodi Sastra Arab Fakultas Sastra UAI di Seminar Internasional Bahasa Arab dan Lokakarya Linguistik Korpus Arab di Universitas Muhammadiyah Yogyakarta, 10—11 Agustus 2017.

PUSTAKA ACUAN

- Adolphs S, *Introducing Electronic Text Analysis - A Practical Guide for Language and Literary Studies*, Routledge, New York, 2006
- Al-Sulaiti L, *Arabic Corpora*, http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm, 2010, [17 Juni 2017]
- Cheng W, *Exploring Corpus Linguistics: Language in Action*, Routledge, Oxon, 2012
- Dukes K, *The Quranic Arabic Corpus*, <http://corpus.quran.com/>, 2011 [17 Juni 2017]
- Nesselhauf, Nadja, *Corpus Linguistics: A Practical Introduction*, <http://www.as.uni-heidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf>, 2011 (15 Juni 2017)
- Parkinson, D., arabiCorpus.byu.edu (18 Juni 2017)
- Sketch Engine, *Arabic TenTen Corpus*, <https://www.sketchengine.co.uk/artenten-corpus/>, 2015, [17 Juni 2017]

Wikipedia, *List of Text Corpora*, https://en.wikipedia.org/wiki/List_of_text_corpora, 2016 , [17 Juni 2017]

صالح ، محمود إسماعيل ، *لسانيات المدونات اللغوية : مقدمة للقارئ العربي*

http://dr-mahmoud-ismail-saleh.blogspot.co.id/2014/04/blog-post_5.html [18 Juni 2017)