



Sertifikat

Nomor : 15.10.07/UN.32.2.5.3/KONASBARA II/2016

Diberikan kepada

Nur Hizbullah, S.Ag., M.Hum

sebagai

Pemakalah

dalam "**Konferensi Nasional Bahasa Arab (KONASBARA) II**"

dengan tema "**Kreativitas dan Inovasi dalam Pembelajaran Bahasa Arab di Indonesia**"

yang diselenggarakan oleh Jurusan Sastra Arab Fakultas Sastra Universitas Negeri Malang

pada hari Sabtu, tanggal 15 Oktober 2016

Malang, 15 Oktober 2016



Prof. Utami Widiati, M.A., Ph.D.
NIP : 19650813 199002 2 001

Ketua Jurusan Sastra Arab

Dr. Hanik Mahliatussikah, M.Hum
NIP : 19740427 199803 2 002

ISSN : 9772540-941006



KONFERENSI NASIONAL BAHASA ARAB

(KONASBARA) II

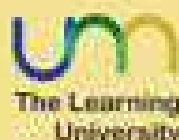
PROSIDING

“Kreativitas dan Inovasi dalam
Pembelajaran bahasa Arab di Indonesia”

Keynote Speaker:

1. Prof. Dr. Tajudin Nur, M.Hum (UNPAD Bandung)
2. Dr. Nasarudin Idris Jauhar, M.Ed (UINSA Surabaya)
3. Dr. Kholisin, M.Hum (UM)

15 Oktober 2016



Jurusan Sastra Arab Fakultas Sastra
Universitas Negeri Malang



KONFERENSI NASIONAL BAHASA ARAB (KONASBARA)



HOME ABOUT LOGIN REGISTER SEARCH CURRENT
ARCHIVES

Home > Archives > No 2 (2016)

NO 2 (2016)

PROSIDING KONFERENSI NASIONAL BAHASA ARAB II

Dalam rangka ikut serta mendorong kreativitas dan inovasi pembelajaran bahasa Arab dan untuk memberikan wadah bagi pertemuan ilmiah dan srawung akademik dalam bidang bahasa Arab dan pengajarannya, Jurusan Sastra Arab Fakultas Sastra Universitas Negeri Malang (UM) menyelenggarakan Seminar Nasional dengan tema: “Kreativitas dan Inovasi dalam Pembelajaran Bahasa Arab di Indonesia”.

TABLE OF CONTENTS

ARTICLES

VERBA BERPREPOSISI DALAM BAHASA ARAB:
ANALISIS STRUKTUR DAN MAKNA
TAJUDIN NUR

(تعليم اللغة العربية بوساطة الإنترنت في إندونيسيا (وقائع وتحديات
نصر الدين إدريس جوهر

MENYIAPKAN GURU BAHASA ARAB PROFESIONAL
DALAM MENGHADAPI DINAMIKA KURIKULUM
KHOLISIN KHOLISIN

تحليل تعليم المحادثة باستخدام الطريقة المباشرة في معهد طالبات بجامعة والي
سونجو الإسلامية الحكومية سمارانج
دوى ماونتي و محمد نحارير فهمي مولانا

STRATEGI PEMBELAJARAN BAHASA ARAB
HASNA QONITA KHANSA

الألعاب الدورية لتعليم مهارة الكلام في مجموعة التعليم المنزلي
(Homeschooling Group)
سيف المصطفى و ميمونة

MODEL PEMBELAJARAN KOOPERATIF
(COOPERATIVE LEARNING) DALAM STRATEGI
BELAJAR MENGAJAR BAHASA ARAB
SITI MUNAWAROH

PROSES PELAKSANAAN PEMBELAJARAN BAHASA

USER

Username

Password

☐ Remember me

Login

NOTIFICATIONS

View
Subscribe

LANGUAGE

Select Language

English ▼

Submit

JOURNAL CONTENT

PDF Search

PDF Search Scope
All ▼

PDF Search

PDF Browse
By Issue
By Author
By Title
Other Journals

PDF FONT SIZE

PDF INFORMATION

PDF For Readers
For Authors
For Librarians

ARAB YANG EFEKTIF DAN MENYENANGKAN
ULFA ZAINUL MUBAROK

METODE PEMBELAJARANKETERAMPILAN
BERBAHASA DI PROGRAM STUDI SARJANA
PENDIDIKAN BAHASA ARAB UNIVERSITAS NEGERI
MALANG (UM)
ABDUL AZIZ KHOIRI

PDF

منهج تزكية النفس كنمط مهنية معلم اللغة العربية
فطرة فطرة

PDF

BELAJAR BAHASA ARAB DENGAN METODE AL-
KASYSYAF: MEMBERANTAS BUTA MAKNA DALAM
MEMBACA
IBNU SAMSUL HUDA

PDF

MENGADOPSI TEORI MEDAN MAKNA MENJADI
METODE PENGAJARAN BAHASA ARAB DI ERA
POSMETODE
Imam Asrori

PDF

MAKNA POST TERJEMAHAN JENGGO TAN (PTJ)
DALAM KONTEKS METODOLOGI PEMBELAJARAN
BAHASA
IRHAMNI ALI MA"SUM

PDF

استخدام لعبة الرموز والصّور في تعليم الإنشاء
ADE DESTRI DEVIANA

PDF

الاطار في تطوير منهج تعليم اللغة العربية
كرمىلا أندرينا

PDF

PEMANFAATAN INTERNET UNTUK BELAJAR
BAHASA ARAB SECARA MANDIRI
IQNAUS SALAM

PDF

STRATEGI PEMBELAJARAN BAHASA ARAB
INTENSIF DI UNIVERSITAS TRUNOJOYO MADURA
LILIS SUAIBAH

PDF

PENDEKATAN DAN PENERAPAN STRATEGI FAUQAL
MA'RIFIAH DALAM PEMBELAJARAN BAHASA
ARAB YANG PAS (PEMBELAJARAN ASIK DAN
SENANG)
MOHAMAD MISHBAHUDDIN

PDF

PEMBELAJARAN MEMBACA BAHASA ARAB
DENGAN PERMAINAN KARTU "UNO" DI KELAS X
IPA 2 MAN 3 MALANG
MEIDIAS ABROR w

PDF

PENGUNAAN SOFTWARE "CAMTASIA STUDIO"
SEBAGAI MEDIA PEMBELAJARAN BAHASA ARAB
BERBASIS ICT
MOCH WAHIB DARIYADI

PDF

MEDIA PEMBELAJARAN GAME TARKIB UNTUK
PEMBELAJARAN TARKIB KELAS XI MADRASAH
ALIYAH

PDF

MOCH ARIS ANDIKA

PEMBUATAN ELECTRONIC BOOK (E-BOOK)
BAHASA ARAB DENGAN SOFTWARE FLIP BOOK
MAKER
MOH FAUZAN PDF

PENULISAN BUKU AJAR BAHASA ARAB
Muhaiban Muhaiban PDF

PENGEMBANGAN BAHAN AJAR INTEGRATIF
BAHASA ARAB DAN AL-QUR'ĀN UNTUK SISWA
SD/MI
SALIM SAPUTRA PDF

إعداد المواد التعليمية من الثقافة المحلية لتعليم مهارة الكلام (تجربة في جامعة
(حمزوني نهضة الوطن فنشور الإسلامية)
siti ruhillah PDF

PEMANFAATAN INTERNET UNTUK BELAJAR
BAHASA ARAB SECARA MANDIRI
IQNAUS SALAM PDF

PENYUSUNAN MODUL UNTUK PEMBELAJARAN
INTENSIF LOMBA BAHASA ARAB
UMMUL FAIDA PDF

KESAHIHAN DALAM PENYUSUNAN TES BAHASA
ARAB DI MADRASAH/SEKOLAH
MOH. AININ PDF

ANALISIS HASIL TOAFL MAHASISWA JURUSAN
SASTRA ARAB FAKULTAS SASTRA UNIVERSITAS
NEGERI MALANG
MOHAMMAD AHSANUDDIN PDF

KESALAHAN-KESALAHAN PADA MEMBACA
NYARING (QIRAAH JAHRIYAH) MAHASISWA BARU
DAN UPAYA PERBAIKANNYA DALAM MATA
KULIAH QIRAAH MUKATSTSAFAH DI JURUSAN
SASTRA ARAB UNIVERSITAS NEGERI MALANG
MUHAMMAD ALFAN PDF

PENERAPAN HUKUMAN UNTUK MENINGKATKAN
KETERAMPILAN BERBICARA DALAM
PEMBELAJARAN BAHASA ARAB
ALIBA'UL CHUSNA PDF

مشكلات ترجمة حرف تعدية "ب" من اللغة العربية إلى اللغة الإندونيسية لدى
طلّاب قسم تعليم اللغة العربية بجامعة والي سنجوا الإسلامية الحكومية
الماس فلذة ظفرنا PDF

KAJIAN AWAL REALITAS DAN PRAKTIK
PEMBELAJARAN BAHASA ARAB MADRASAH
MINORITAS MUSLIM PAPUA BARAT
Ismail Suardi Wekke, Agung Muttaqien PDF

TRADISIONALISME DAN MODERNISME
PEMBELAJARAN BAHASA ARAB: PEMBELAJARAN PDF

MADRASAH BERBASIS PESANTREN MINORITAS MUSLIM

Ismail Suardi Wekke, Kisno Umbar, Azhar Arsyad

BELAJAR DAN PEMBELAJARAN BAHASA ARAB
MADRASAH MINORITAS MUSLIM PAPUA BARAT
ISMAIL SUARDI WEKKE, MUAMMAR NUR

PDF

APLIKASI DAN PRINSIP BELAJAR DALAM
PEMBELAJARAN BAHASA ARAB PERGURUAN
TINGGI MINORITAS MUSLIM
ISMAIL SUARDI WEKKE, FAISOL

PDF

LINGUISTIK KORPUS DALAM KAJIAN DAN
PEMBELAJARAN BAHASA ARAB DI INDONESIA
NUR HIZBULLAH, FAZLURRAHMAN
FAZLURRAHMAN, FUZI FAUZIAH

PDF

RANCANGAN PEMBELAJARAN BAHASA ARAB
BERBASIS BLENDED LEARNING
Nurhidayati Nurhidayati

PDF

BERMAIN PERAN SEBAGAI
DUBBER:IMPLEMANTASI INOVATIF
PENDEKATANAURAL ORAL DALAM
PEMBELAJARAN KETERAMPILAN KALAM
NURUL WAHDAH

PDF

المدخل الاتصالي وتطبيقه في تعليم اللغة العربية
ألف نجا إمامنا, إمام أسرار

PDF

TEORI BELAJAR DAN PENERAPANNYA DALAM
PEMBELAJARAN BAHASA ARAB
DEWI LATIFAH

PDF

PENERAPAN TEORI BEHAVIORISTIK DALAM
PEMBELAJARAN BAHASA ARAB (KAJIAN
TERHADAP PEMIKIRAN BF. SKINNER)
MUHAMMAD MAHMUDI

PDF

TAKSONOMI BLOOM – REVISI : RANAH KOGNITIF
SERTA PENERAPANNYA DALAM PEMBELAJARAN
BAHASA ARAB
FARA DIBA FAUZET

PDF

BEBERAPA POKOK PIKIRAN UNTUK PENINGKATAN
EFEKTIFITAS PEMBELAJARAN TARKIIB
MOH. KHASAIRI

PDF

PANDANGAN DAN PENDEKATAN PEMBELAJARAN,
DAN IMPLEMENTASINYA DALAM PEMBELAJARAN
BAHASA ARAB
MUNIROTUN NAIMAH

PDF

BAGAIMANAKAH PENGELOLAAN KELAS UNTUK
MEMBENTUK LINGKUNGAN BAHASA ARAB (BI'AH
ARABIYAH)?
UBAIDILLAH QUDSI

PDF

MANAJEMEN PEMBELAJARAN DI SEKOLAH

PDF

AHMAD KHATIB RIDHANI

PERENCANAAN PEMBELAJARAN BERBASIS
PESERTA DIDIK SEBAGAI PENUNJANG
KEBERHASILAN KEGIATAN BELAJAR MENGAJAR
BAHASA ARAB DI KELAS
KHOLISNAH MINAL BALWA PDF

PUNYAKU, PUNYAMU, DAN PUNYA KITA
(STIMULUS PESERTA DIDIK DALAM BELAJAR
BAHASA ARAB SECARA MANDIRI DARI BERBAGAI
SUDUT PANDANG)
SITI NUR MAMBAUS SHAFa PDF

KONSTRUKSI APODOSIS DALAM SUSUNAN
KONDISIONAL BAHASA ARAB (A'T-TARKĪB A'SY-
SYARTHIY)
AFNAN ARUMMI PDF

REPRESENTASI MAKNA KALA OLEH ADJEKTIVA
DALAM LINGUISTIK ARAB
ARIFUDDIN TRI YANTI NURUL HIDAYATI
MUHAMMAD RIDWAN PDF

PEMBELAJARAN APRESIASI SAJAK “AL-MASÂ`”
KARYA SASTRAWAN ARAB DIASPORA ILIYA ABU
MADHI (Sebuah Studi Sastra Berdasarkan Teori Moody)
HANIK MAHLIATUSSIKAH PDF

SASTRA ISLAM DALAM MANUSKRIP (KAJIAN TEKS
BER-AKSARA JAWI SEBAGAI SALAH SATU BAHAN
AJAR MAHASISWA
IIN SURYANINGSIH PDF

KAJIAN SEMIOTIKA DALAM PENAFSIRAN AL-
QUR'AN
DONY BURHAN NOOR HASAN PDF

PENGUNAAN NUN TAUKID TSAQILAH DALAM
ALQURAN: WUJUD DAN FUNGSINYA
LAILY MAZIYAH PDF

MISTERI ESTETIKA BUNYI AL-QUR'AN
YUSUF HANAFI PDF

ISSN: 25409417



PROSIDING

KONFERENSI NASIONAL BAHASA ARAB (KONASBARA)



Jurusan Sastra Arab
FS UM

[HOME](#) [ABOUT](#) [LOGIN](#) [REGISTER](#) [SEARCH](#) [CURRENT](#)
[ARCHIVES](#)

Home > About the Journal > Editorial Team

EDITORIAL TEAM

EDITORS

panel
konasbara konasbara konasbara

ISSN: 25409417

USER

Username

Password

☐ Remember me

NOTIFICATIONS

[View](#)
[Subscribe](#)

LANGUAGE

Select Language

English ▼

JOURNAL CONTENT

Search

Search Scope

All ▼

Browse

[By Issue](#)
[By Author](#)
[By Title](#)
[Other Journals](#)

FONT SIZE

INFORMATION

[For Readers](#)
[For Authors](#)
[For Librarians](#)

LINGUISTIK KORPUS DALAM KAJIAN DAN PEMBELAJARAN BAHASA ARAB DI INDONESIA*

NUR HIZBULLAH, FAZLURRAHMAN, FUZI FAUZIAH

Universitas Al Azhar Indonesia

hizbi77@gmail.com, fazz679@gmail.com, fuzi1366@gmail.com

Abstrak: Makalah ini bertujuan mendeskripsikan beberapa hal tentang linguistik korpus serta perkembangan dan dinamikanya dalam kajian bahasa Arab di dunia internasional. Secara khusus akan digambarkan sejauh mana perkembangan keberadaan korpus bahasa Arab dan aplikasi yang dapat dimanfaatkan untuk pengolahan dan analisis korpus. Hal itu kiranya dapat dijadikan dasar bagi wacana pemikiran dan diskusi akademis mengenai kemungkinan penyusunan suatu model korpus bahasa Arab di Indonesia sekaligus pemanfaatannya dalam kajian dan pembelajaran bahasa Arab di tingkat perguruan tinggi.

PENDAHULUAN

Perkembangan ilmu linguistik modern disemarakkan dengan kemunculan suatu cabang ilmu yang relatif “baru”, yaitu linguistik korpus. Ilmu ini secara khusus meneliti bahasa melalui seperangkat data yang bersifat alamiah, riil sesuai penggunaannya, baik itu data tulisan maupun data lisan yang ditranskripsikan (Adolphs, 2006: 137). Menurut pengertiannya, korpus merupakan sekumpulan data, baik data biasa maupun data digital, dalam bentuk tertulis yang berisi bermacam-macam informasi kebahasaan, mulai dari tataran kata, struktur, makna, dan wacana, yang dapat dimanfaatkan untuk penelitian. Dalam pandangan Cheng (2012: 6), linguistik korpus bekerja dengan menganalisis data melalui proses kompilasi berdasarkan kategori dan satuan analisisnya. Di tahap berikutnya, kategori dan satuan analisis yang paling penting dalam lingkup linguistik korpus adalah daftar kata, kata kunci, dan konkordansi. Daftar kata mengandung sejumlah besar kata yang terdapat dalam data kebahasaan dan jumlah frekuensi kemunculannya dalam teks sumber. Selain daftar kata, ada pula kata kunci yang merupakan kumpulan kata dengan frekuensi kemunculan terbanyak dan diasumsikan sebagai representasi topik inti dalam teks, sedangkan konkordansi adalah daftar kata yang berkolokasi dengan kata lain dan membentuk suatu makna baru berdasarkan konteks serta koteksnya (Adolphs, 2006: 136). Sebelum ditemukan dan berkembangnya teknologi komputer, semua data kebahasaan dalam bentuk tertulis masih berupa manuskrip atau tulisan biasa di atas kertas ataupun objek lain. Sejak ditemukannya teknologi komputer dan dimulainya era digital, kini data kebahasaan yang berbentuk tulisan atau teks serta bahasa lisan yang diubah menjadi tulisan digital dapat disimpan dan diolah secara lebih mudah. Di sinilah linguistik korpus mengambil bagian, yaitu mengolah data kebahasaan digital.

* Makalah yang disajikan pada Konferensi Nasional Bahasa Arab (KONASBARA) II yang diselenggarakan oleh Jurusan Sastra Arab, Fakultas Sastra, Universitas Negeri Malang, 15 Oktober 2016

Korpus yang dihasilkan melalui suatu proses digitalisasi kemudian dapat dimanfaatkan dalam berbagai bidang linguistik, seperti morfologi, sintaksis, semantik, leksikologi, maupun bidang kebahasaan lain seperti penerjemahan, kesusastraan, dan sebagainya. Di banyak bahasa dunia, sudah tersedia berbagai bentuk dan ragam data kebahasaan yang dikumpulkan. Pada praktiknya di dalam bahasa lain, pemanfaatan linguistik korpus sudah sangat meluas dan menguat. Tak terkecuali bahasa Arab, meski perkembangannya mungkin tidak sedinamis penyusunan korpus bahasa lainnya. Korpus bahasa yang tersedia tidak hanya bersumber dari bahasa era modern, bahkan bahasa era klasik pun ikut “dikorpuskan” dengan memanfaatkan dokumen-dokumen tertulis yang masih ada, bisa ditemukan, dan dipelajari. Namun, di negara ini, kiranya masih perlu sosialisasi lebih luas kepada kalangan bahasawan dan linguist akan manfaat linguistik korpus, sebelum mendorong pemanfaatan bidang ini lebih maju lagi oleh mereka yang berminat dan yang berkepentingan.

Makalah ini ingin mendeskripsikan beberapa aspek dasar dari linguistik korpus serta perkembangan dan dinamikanya dalam kajian di dunia internasional secara umum dan khususnya di dunia bahasa Arab. Akan digambarkan pula dinamika penyusunan dan pengembangan korpus bahasa Arab dan aplikasi yang dapat dimanfaatkan secara praktis untuk pengolahan dan analisis korpus berbahasa Arab. Besar harapan agar kiranya hal ini dapat dijadikan dasar pemikiran bagi wacana dan diskusi mengenai perlunya akademisi bahasa Arab di Indonesia untuk merespons dan mengikuti perkembangan tersebut. Langkah konkret yang perlu dilakukan kemudian adalah penyusunan suatu model korpus bahasa Arab di Indonesia sekaligus pemanfaatannya dalam bidang kajian dan bidang pembelajaran bahasa Arab di tingkat perguruan tinggi.

SEKILAS TENTANG LINGUISTIK KORPUS

Mudahnya, linguistik korpus adalah sebuah metode empiris dalam analisis dan deskripsi linguistik untuk meneliti bahasa yang dipakai oleh para penutur secara alamiah (Cheng, 2012: 6, 29-30). Metode ini semakin luas digunakan dan bisa untuk meneliti berbagai masalah kebahasaan dalam lingkup yang luas dan bahkan mampu menyingkap hal-hal yang menarik, fundamental, bahkan “mengejutkan” dari data-data kebahasaan yang ada.

Data kebahasaan seperti apa yang bisa dianalisis secara linguistik dengan metode ini? Nesselhauf (2011: 2) menyebutkan, secara umum, data itu dibagi ke dalam dua kelompok besar, yaitu data yang diperoleh dari intuisi, baik intuisi si peneliti maupun intuisi informan, dan bahasa yang digunakan sehari-hari secara natural, bisa dari percakapan ringan dan bisa juga sekumpulan teks yang dikoleksi dan dikelola secara sistematis. Data jenis itulah yang disebut korpus (jamak: “korpora”). Lebih lanjut, masih menurut Nesselhauf (2011: 3), korpus adalah teks dari penggunaan bahasa lisan dan tulisan yang praktis serta alamiah dan dikumpulkan secara sistematis. Maksud dari sistematis adalah penyusunan struktur dan isi teks itu didasarkan atas prinsip luar bahasa, seperti konteks waktu, tempat, dan lainnya, serta memungkinkan untuk diteliti lebih jauh mengenai aspek-aspek yang ada di dalamnya.

Dalam hal penyusunan korpus bahasa, kini sudah tersedia banyak sekali model korpus bahasa. Lebih khusus, dilihat dari tipenya, ada beberapa jenis korpus yang bisa digunakan bergantung pada jenis analisisnya. Berikut macam-macamnya:

1. Korpus general/referensi, contoh: British National Corpus (BNC) – “vs” korpus khusus, yang dibuat untuk merepresentasikan suatu bahasa secara umum dan menyeluruh, terdiri atas data bahasa tertulis dan lisan yang kemudian ditranskripsikan serta bersumber dari berbagai jenis teks.
2. Korpus historis, contoh Helsinki Corpus – vs korpus bahasa modern, yang disusun untuk merepresentasikan periode-periode tertentu dari perkembangan suatu bahasa.
3. Korpus regional, seperti Wellington Corpus of Written New Zealand English (WCNZE), disusun untuk merepresentasikan variasi bahasa dari suatu kawasan.
4. Korpus pembelajar, seperti International Corpus of Learner English (ICLE) – vs korpus penutur asli, yang dibuat untuk merepresentasikan bahasa yang diproduksi oleh para pembelajar asing bagi bahasa tersebut dan bukan merupakan penutur asli.
5. Korpus multilingual – vs korpus ekabahasa, yang merepresentasikan sedikitnya suatu fenomena dari dua bahasa yang bersumber dari ragam teks yang sama, biasanya digunakan untuk keperluan analisis kontrastif.
6. Korpus lisan – vs korpus tulisan vs campuran lisan-tulisan, untuk merepresentasikan bahasa lisan dari suatu masyarakat penutur bahasa.

Ada satu jenis korpus lain yang penyusunannya didasarkan bukan pada sumbernya, tapi berdasarkan suatu tahapan dalam proses awal pengolahannya, yaitu korpus beranotasi – vs korpus ortografis. Dalam korpus beranotasi, beberapa jenis analisis linguistik biasanya sudah dilakukan terhadap teks sehingga sudah memuat anotasi yang sejatinya bukan bagian asli dari teks tersebut (Nesselhauf, 2011: 3-4).

Berbagai bahasa di dunia kini sudah “terjamah” oleh linguistik korpus. Sekian banyak korpus bahasa dengan segala tipenya bermunculan dan beberapa di antaranya bahkan terbuka dan bisa diakses oleh publik. Wikipedia, misalnya, mendata ada sejumlah korpus bahasa yang dibagi ke dalam beberapa kelompok bahasa. Uniknya, bahasa Inggris berada dalam kelompok tersendiri karena ketersediaan banyak sekali ragam dan sumber korpusnya, seperti yang paling terkemuka dan termutakhirkan adalah British National Corpus (BNC), American National Corpus (ANC), Corpus of Contemporary American English (COCA), Bank of English, Brown Corpus, Oxford English Corpus, dan lain sebagainya. Selain kategori bahasa Inggris, dalam rumpun bahasa-bahasa Eropa, misalnya, tercantum nama-nama seperti Bulgarian National Corpus, Croatian Language Corpus, Czech National Corpus, Thesaurus Linguae Graecae, German Reference Corpus, dan lain-lain. Untuk bahasa-bahasa Asia, situs ini mencantumkan kategori Asia Timur yang berisi Kotonoha Japanese Language Corpus dan LIVAC Synchronous Corpus untuk bahasa China. Selain itu, ada pula kategori bahasa-bahasa Timur Tengah yang berisi Hamshahri Corpus untuk bahasa Persia, Neo-Assyrian Text Corpus Project, Turkish National Corpus, dan Quranic Arabic Corpus. Yang menarik, ada pula kategori korpus paralel multibahasa, semacam Open Source Parallel Corpus, NTU-Multilingual Corpus, dan kategori *Comparable corpora*, seperti WaCky – The Web-As-

Corpus Kool Yinitiative Web as Corpus yang memuat kombinasi bahasa Inggris, Prancis, Jerman, dan Italian, Disambiguating Similar Language Corpora Collection (DSLCC) yang berisi banyak sekali bahasa antara lain Bosnia, Kroasia, Serbia, Indonesia, Malaysia, Cek, dan lain-lain, serta Wikipedia Comparable Corpora.

KORPUS BAHASA ARAB DAN KAJIANNYA

Secara khusus, sudah tersedia beberapa korpus bahasa Arab yang dibuat dengan segala kekhususan, kelebihan, dan tak ketinggalan kekurangannya masing-masing. Dalam penyusunan korpus bahasa Arab, melalui suatu aplikasi di laman Sketch Engine, misalnya, ada korpus yang isinya lebih kurang 5,8 juta kata (versi 2012) dan diperbarui menjadi 8,3 juta kata (versi 2015)¹ yang diambil dari sejumlah sumber. Selain itu, Alsubaiti menginventarisasi daftar yang lebih lengkap lagi, sekitar 18 jenis korpus, yang berasal dari berbagai sumber dan digunakan untuk berbagai bidang kajian spesifik yang ada di sublaman University of Leeds, di antaranya Corpus of Contemporary Arabic, Arabic Gigaword, dan yang sedang dalam pengembangan adalah International Corpus of Arabic oleh University of Alexandria, Mesir.² Sekian banyak korpus itu terbagi ke dalam dua kelompok besar, yaitu yang dapat diunduh dengan berbayar dan tak berbayar.

Di samping korpus bahasa Arab, ada pula korpus khusus mengenai Al-Qur'an. Sejauh ini, menurut pengamatan kami, yang terlihat paling representatif adalah karya University of Leeds, yaitu *The Quranic Arabic Corpus*.³ Tidak hanya korpus Al-Qur'an, laman itu memuat model-model analisis kebahasaan Al-Qur'an, seperti terjemah kata-per-kata dalam bahasa Inggris, kamus kosa kata Al-Qur'an, "pohon sintaksis" struktur konsep dan kata, dan gramatika Al-Qur'an. Laman ini menyediakan pula file digital Al-Qur'an dalam format *.txt yang bisa diunduh untuk bahan analisis. Namun, file tersebut hanya berisi teks Al-Qur'an dalam transliterasi latin, bukan Arab, meskipun dilengkapi dengan status gramatikal setiap kata.

Kajian mengenai linguistik korpus Arab, dan khususnya Al-Quran, antara lain dilakukan oleh Sharaf (2009). Ahli yang satu ini memperkenalkan serangkaian risetnya berikut hasil pemetaan konseptual terhadap isi Al-Qur'an dengan aplikasi tertentu. Namun, di laman webnya, tidak ada bahan digital olahannya yang dapat diunduh dan dimanfaatkan untuk penelitian lebih lanjut.⁴

Salah satu langkah penyusunan korpus umum bahasa Arab yang mencantumkan Al-Qur'an sebagai salah satu sumbernya adalah yang dilakukan oleh Al-Rabia (et.al., 2014). Korpusnya dinamakan King Saud University Corpus of Classical Arabic (KSUCCA). Sesuai namanya, mereka hanya fokus mengambil korpus dari bahasa Arab periode klasik dan Al-Qur'an diasumsikan termasuk ke dalam periode itu. Mereka beralasan, pola bahasa Arab klasik adalah merupakan dasar teori linguistik Arab dan harus dipandang sebagai sumber ilmiah-akademis. Korpus ini berisi 50 juta kata dari khazanah pustaka bahasa Arab masa klasik dan digunakan antara lain untuk keperluan studi distribusi semantik leksikal Al-Qur'an dan bahasa Arab klasik dengan pendekatan linguistik komputasional. Namun,

¹ <https://www.sketchengine.co.uk/artenten-corpus/> diakses 30 Agustus 2016

² http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm, diakses 30 Agustus 2016

³ <http://corpus.quran.com/> diakses 30 Agustus 2016

⁴ <http://www.textminingthequran.com/wiki/QurConcord>, diakses 30 Agustus 2016

selain menggambarkan struktur korpus, penelitian mereka juga mendeskripsikan beberapa gejala perubahan bahasa dari masa klasik ke masa modern.

Ada karya lain yang lebih spesifik lagi, misalnya, Sharaf dan Atwell (2012). Mereka membuat "QurAna", korpus Al-Qur'an beranotasi khusus pada kasus anafora pronominal. Penelitian itu diarahkan kepada analisis pronomina dalam Al-Qur'an dan pola kemunculan anafora yang terkait dengan pronomina dan kaitan antesedennya dalam ayat-ayat Al-Qur'an. Model analisis penelitian ini bermanfaat untuk diterapkan pada analisis pronomina, anafora, dan antesedennya dalam wacana bahasa Arab modern dan perbandingan kedua masa itu.

Salah satu contoh lain penelitian korpus Al-Qur'an yang lebih spesifik adalah karya Imad dan Abdelhak (2016). Mereka melakukan penelitian untuk menyusun korpus Al-Qur'an khusus untuk penelitian di bidang morfologi Arab. Mereka menggunakan teks Al-Qur'an dan juga memanfaatkan tabel morfologi Al-Khalil. Mereka menggunakan data dari Quranic Arabic Corpus milik University of Leeds dan Quranic Corpus of Haifa milik University of Haifa sebagai pedoman. Langkah selanjutnya, mereka membuat analisis dan kritik terhadap konsep akar kata, serta diakhiri dengan perbandingan antara keduanya dan pemaparan pola rumusan si peneliti sendiri.

Sebagai simpulan terhadap penelitian terdahulu, dapat dikatakan bahwa penggunaan pendekatan linguistik korpus di kalangan peneliti bahasa Arab sudah sedemikian marak. Semestinya para ahli dan peminat kajian bahasa Arab di Indonesia masuk ke percaturan ini dan ikut menyemarakkan kajian bahasa Arab berbasis korpus.

BEBERAPA APLIKASI PENGOLAH KORPUS

Aplikasi untuk mengolah korpus banyak tersedia dan bisa dipilih tergantung ketersediaan menu dalam aplikasinya. Ada yang bisa diunduh secara cuma-cuma, seperti Nooj, TextStat, MonoconcEsy, Aconcord, dan AntCont. Ada juga aplikasi yang berbayar, seperti WordSmith. Namun, perlu dicermati bahwa pemilihan perangkat analisis korpus harus mempertimbangkan karakteristik bahasa Arab dan ortografinya yang relatif berbeda dari bahasa-bahasa pada umumnya. Perbedaan yang khas ada pada bahasa Arab misalnya arah tulisan dari kanan ke kiri, keberadaan tanda diakritik dalam tulisan kata, dan ragam serta bentuk huruf yang khas. Situasi itu membuat pilihan aplikasi menjadi terbatas. Beberapa aplikasi yang ada, misalnya, hanya mampu membaca teks Arab, tetapi tidak bisa mengadaptasikan pola tulisan Arab (dari kanan ke kiri). Masalah itulah yang paling lazim muncul dalam beberapa aplikasi yang kami amati. Pada akhirnya, dari sekian banyak pilihan dan mempertimbangkan kemampuan teknisnya, pilihan paling rasional kiranya bisa diarahkan kepada aplikasi WordSmith.

Dibanding aplikasi lainnya, WordSmith bisa dianggap yang paling lengkap dan paling representatif. Utamanya, kemampuan aplikasi itu untuk mengolah teks dalam bahasa apa pun sudah memadai karena dilengkapi dengan menu penyesuaian sistem yang diperlukan untuk analisis teks bahasa yang memiliki keunikan ortografi dan kekhasan teknis lainnya. Kelebihan lainnya, aplikasi ini menyediakan fasilitas untuk mengonversi teks yang memungkinkan kustomisasi

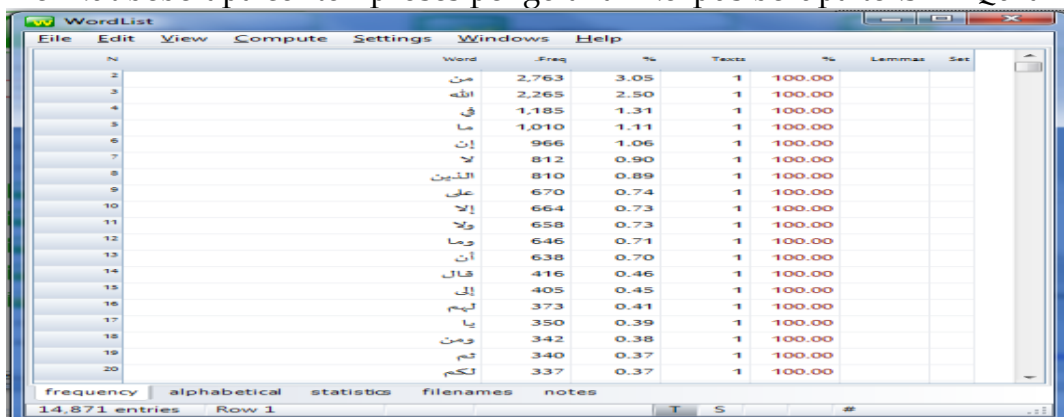
yang optimal bagi teks yang akan dianalisis jika format asalnya belum sesuai dengan format yang bisa diolah oleh aplikasi yang bersangkutan.

PENGOLAHAN TEKS ARAB DENGAN APLIKASI PENGOLAH KORPUS

Pada bagian ini, akan digambarkan sekilas pengalaman dalam menggunakan aplikasi WordSmith untuk mengolah teks berbahasa Arab, dalam hal ini Al-Qur'an. Sebelum teks diolah, perlu dipastikan bahwa *file* yang akan diolah harus dikonversi terlebih dulu. Biasanya, *file* yang ditulis dengan aplikasi Microsoft Word berekstensi *.doc atau *.docx. Format itu harus dikonversi dengan metode *encoding* konversinya menggunakan UTF-8 dan disimpan dengan ekstensi *.txt. Setelah diuji coba dengan beberapa aplikasi lain, dapat dikonfirmasi bahwa mode konversi selain *encoding* UTF-8 tak berterima bagi hampir semua aplikasi pengolah korpus. Secara teknis, hanya *file* dengan ekstensi *.txt dengan mode konversi UTF-8 saja yang dapat diolah dengan aplikasi korpus pada umumnya.

Khusus menyangkut teks Al-Quran, sebelum diolah dengan aplikasi, bahan korpus yang sudah dikonversi harus diverifikasi dulu dengan cara pembacaan manual untuk memastikan kelengkapan redaksi ayat Al-Qur'an secara keseluruhan dan juga untuk menghindari jangan sampai ada ayat yang hilang akibat proses konversi format. Adapun untuk teks selain Al-Quran, langkah itu bisa perlu bisa pula tidak. Bagaimanapun, secara teknis, proses konversi teks dari suatu format ke format yang lain selalu berisiko terhadap kondisi teks tersebut, apalagi jika teks itu ditulis dalam bahasa seperti Arab yang memiliki sistem ortografi yang tidak sama dengan sistem ortografi bahasa latin pada umumnya.

Berikut beberapa contoh proses pengolahan korpus berupa teks Al-Quran.



N	Word	.Freq	%	Texts	%	Lemmas	Set
2	من	2,763	3.05	1	100.00		
3	الله	2,265	2.50	1	100.00		
4	في	1,185	1.31	1	100.00		
5	ما	1,010	1.11	1	100.00		
6	إن	966	1.06	1	100.00		
7	لا	812	0.90	1	100.00		
8	الذين	810	0.89	1	100.00		
9	على	670	0.74	1	100.00		
10	إلا	664	0.73	1	100.00		
11	ولا	658	0.73	1	100.00		
12	وما	646	0.71	1	100.00		
13	أن	638	0.70	1	100.00		
14	قال	416	0.46	1	100.00		
15	إلى	405	0.45	1	100.00		
16	لهم	373	0.41	1	100.00		
17	يا	350	0.39	1	100.00		
18	ومن	342	0.38	1	100.00		
19	ثم	340	0.37	1	100.00		
20	لكم	337	0.37	1	100.00		

Gambar 1. Contoh proses penyusunan daftar kata dan kemunculan hasil dalam peringkat 20 besar teratas.

Concordance									
File Edit View Compute Settings Windows Help									
N	Concordance	Set	Tag	# Word	Sent % Pos	Sent % Neg	Para % Pos	Para % Neg	File
									Dr.
1	من ريك ليعولن إنا كنا معكم أويين الله بأعلم بما في صدور العالين	29[11]			51,513	0.36%	0.36%	0.36%	n-simple-clean.txt log/14-00-00/201
2	العالم [29]ومن جاهد بينهم وأجملت لنفسه إن الله لعلي من العالين	29[11]			51,442	0.36%	0.36%	0.36%	n-simple-clean.txt log/14-00-00/201
3	الأعين في البعثة المبكرة من الحجرة أن ما موسى إنا على العالين	29[11]			50,445	0.34%	0.34%	0.34%	n-simple-clean.txt log/14-00-00/201
4	إلا لهموا إنكم لذاتون الفاحشة ما يسعكم ما من أحد من العالين	29[29]			51,784	0.36%	0.36%	0.36%	n-simple-clean.txt log/14-00-00/201
5	[86]فلما أتت يوم النور [37]87فمن العالين	37[88]			58,264	0.74%	0.74%	0.74%	n-simple-clean.txt log/14-00-00/201
6	الباب [37]78وكانا علي من العالين	37[79]			58,229	0.74%	0.74%	0.74%	n-simple-clean.txt log/14-00-00/201
7	الرسم المزمع إليه [32]إن الكتاب لا يخرج فيه من العالين	32[32]			53,754	0.99%	0.99%	0.99%	n-simple-clean.txt log/14-00-00/201
8	قاور تآكل برني فقلت نفسي وأسلمت عن سليمان علي العالين	27[45]			49,352	0.33%	0.33%	0.33%	n-simple-clean.txt log/14-00-00/201
9	العالم [26]166علي من أجبر إن أجبي لا على العالين	26[165]			48,423	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
10	العالم [26]147علي من أجبر إن أجبي لا على العالين	26[146]			48,314	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
11	العالم [26]127علي من أجبر إن أجبي لا على العالين	26[128]			48,222	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
12	أجرى إن أجبي لا على العالين	26[166]			48,227	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
13	جاءهم بعد أن يروك من النار ومن حوله ويسميت إليه من العالين	27[91]			48,858	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
14	العالم [26]191علي من أجبر إن أجبي لا على العالين	26[192]			48,592	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
15	العالم [26]180علي من أجبر إن أجبي لا على العالين	26[181]			48,514	0.32%	0.32%	0.32%	n-simple-clean.txt log/14-00-00/201
16	مكنون [26]79علي لا لمظنون [36]19علي من العالين	36[19]			69,742	0.99%	0.99%	0.99%	n-simple-clean.txt log/14-00-00/201
17	وسعدتون [36]45علي بالصدقة والبر والعتبات وري الأبرار من العالين	36[13]			65,183	0.33%	0.33%	0.33%	n-simple-clean.txt log/14-00-00/201
18	والكتاب والنسوة وزلفاهن وفلسطاهن والعالي	45[17]			64,884	0.33%	0.33%	0.33%	n-simple-clean.txt log/14-00-00/201
19	الإسلام كفر كما قال إلي أفاد الله في العالي	59[17]			71,207	0.91%	0.91%	0.91%	n-simple-clean.txt log/14-00-00/201
20	أهم مبعوثون [35]13عليوم يعرف الناس إلي العالين	35[37]			76,503	0.90%	0.90%	0.90%	n-simple-clean.txt log/14-00-00/201
concordance collocates plot patterns clusters timeline filenames source text notes									
61 entries	Row 1	T	S	يس الله بأعلم بما في صدور العالين 29[11]إبراهيم الله النبي					

Gambar 2. Contoh proses penyusunan konkordansi dan hasilnya.

PROYEKSI KAJIAN LINGUISTIK KORPUS BAHASA ARAB DI INDONESIA

Melihat pesatnya perkembangan produk ilmiah kajian bahasa Arab dengan menggunakan linguistik korpus di kalangan para ahli di mancanegara, itu membuktikan bahwa kedudukan linguistik semakin penting bagi pengembangan bahasa Arab lintas masa. Uniknya lagi, kajian terhadap bahasa Arab di luar negara Arab juga semarak sebagai bukti intensifnya perhatian para ahli terhadap bahasa Arab. Berkaca kepada situasi itu, penting kiranya bagi para ahli dan peminat kajian bahasa Arab di Indonesia untuk terbuka terhadap fakta itu, lalu lebih jauh dan lebih luas lagi memanfaatkan linguistik korpus dalam penelitian mereka.

Secara khusus, makalah ini melihat akan pentingnya keberadaan sebuah korpus bahasa Arab di Indonesia. Secara khusus, pilihan yang bisa diambil untuk tahap awal adalah penyusunan korpus pembelajar bahasa Arab di Indonesia. Menurut hemat kami, langkah itu mungkin diambil dan amat penting dalam rangka membuat suatu peta sudah sejauh mana dan sudah seluas apa pembelajaran bahasa Arab dan kajiannya di berbagai jenjang pendidikan di negeri ini. Akan lebih baik jika nanti disusun korpus bahasa Arab yang bersumber dari teks-teks berbahasa Arab karya penutur Indonesia asli. Setelah korpus itu tersedia, dilakukanlah penelitian berbasis linguistik korpus secara bertahap, berkelanjutan, dan komprehensif terhadap bahasa Arab. Di sisi lain, kajian itu bisa dihubungkan dengan bidang ilmu linguistik yang lain, seperti morfologi, sintaksis, semantik, pragmatik, wacana, leksikografi, kesusastraan, penerjemahan, dan lain-lain.

Di sisi yang lebih umum, kami menganggap perlu untuk lebih digalakkan lagi pendekatan linguistik korpus sebagai titik tolak dalam penelitian linguistik dalam rangka memaksimalkan pemanfaatan data-data kebahasaan Arab terkini yang sudah lebih banyak berbentuk digital. Harapannya adalah, kajian bahasa Arab di Indonesia tidak hanya “bertaklid” kepada negara lain terhadap pemanfaatan linguistik korpus dalam penelitian kebahasaan. Namun, sudah sepantasnya, para ahli bahasa Arab di Indonesia ikut menyumbangkan kepada dunia kajian bahasa Arab global suatu model korpus bahasa Arab dengan mengangkat khazanah sumber pustaka pembelajaran bahasa Arab dan sumber lain produk ilmiah asli putra-putri negeri ini.

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 Konferensi Nasional Bahasa Arab (KONASBARA) II yang diselenggarakan di Universitas Negeri Malang, Jawa Timur

PUSTAKA ACUAN

- Adolphs S, *Introducing Electronic Text Analysis - A Practical Guide for Language and Literary Studies*, Routledge, New York, 2006
- Adams A, *QuranCode*, <http://heliwave.com>, 2015 [30 Agustus 2016]
- Alhaj A, *New Theory of the Holy Qur'an Translation. A Textbook for Advanced University Students of Linguistics and Translation*, Hamburg, Anchor Academic Publishing, 2015
- Al-Sulaiti L, *Arabic Corpora*,
http://www.comp.leeds.ac.uk/eric/latifa/arabic_corpora.htm, 2010,
[30 Agustus 2016]
- Cheng W, *Exploring Corpus Linguistics: Language in Action*, Routledge, Oxon, 2012
- Dukes K, *The Quranic Arabic Corpus*, <http://corpus.quran.com/>, 2011 [30 Agustus 2016]
- Imad Z, Abdelhak L, *New Quranic Corpus rich in Morphological Information*, International Journal of Speech Technology, 2016 [30 Agustus 2016]
- Ismail D, Asyraf HA, *Mushaf Imla'i dan Implikasinya dalam Pembacaan Al-Quran (Imla'i Mushaf and It's Implication in Reading Al-Quran)*, 'Ulum Islamiyyah Journal,
<http://ddms.usim.edu.my/bitstream/123456789/12106/1/ulum%20islamiyyah%2017%20-%203.pdf>, 2016 [30 Agustus 2016]
- Khan MK, Alginahi YM, *The Holy Quran Digitization: Challenges and Concerns*, Life Science Journal, p.156-164, <http://www.lifesciencesite.com>, 2013 [30 Agustus 2016]
- Nesselhauf, Nadja, *Corpus Linguistics: A Practical Introduction*,
<http://www.as.uni-heidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf>, 2011 (30 Agustus 2016)
- Sharaf A, Atwell ES, *QurAna: Corpus of the Quran annotated with Pronominal Anaphora*, University of Leeds,
<http://www.textminingthequran.com/wiki/QurConcord>, 2012, [30 Agustus 2016]
- Sketch Engine, *Arabic TenTen Corpus*, <https://www.sketchengine.co.uk/artenten-corpus/>, 2015, [30 Agustus 2016]
- Tanzil, *Tanzil Quran Text Features*, <http://tanzil.net/docs>, 2007 [30 Agustus 2016]
- Wikipedia, *List of Text Corpora*, https://en.wikipedia.org/wiki/List_of_text_corpora, 2016, [30 Agustus 2016]
- <http://www.laurenceanthony.net/software/antconc/> [30 Agustus 2016]
- <http://www.lexically.net/wordsmith/> [30 Agustus 2016]
- <http://www.monoconc.com/> [30 Agustus 2016]

<http://neon.niederlandistik.fu-berlin.de/en/textstat/> [30 Agustus 2016]

<http://www.nooj4nlp.net/pages/nooj.html> [30 Agustus 2016]

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

http://dr-mahmoud-ismail-saleh.blogspot.co.id/2014/04/blog-post_5.html [30 Agustus 2016).

