

## KEGIATAN PEMBELAJARAN 1

### QURANIC ARABIC CORPUS (QAC)

#### A. Tujuan

Setelah mempelajari materi pokok 1 tentang *Quranic Arabic Corpus* (QAC), diharapkan anda dapat :

1. Menjelaskan pengolahan bahasa Arab Al-Qur'an
2. Menjelaskan korpus bahasa Arab dan kajiannya
3. Menjelaskan dalam mengolah teks Arab dengan aplikasi pengolah korpus
4. Menjelaskan proyeksi kajian linguistik korpus bahasa Arab di Indonesia

#### B. Indikator Pencapaian Kompetensi

Setelah mempelajari materi pokok 1 tentang *Quranic Arabic Corpus* (QAC) diharapkan Anda menguasai kompetensi tentang :

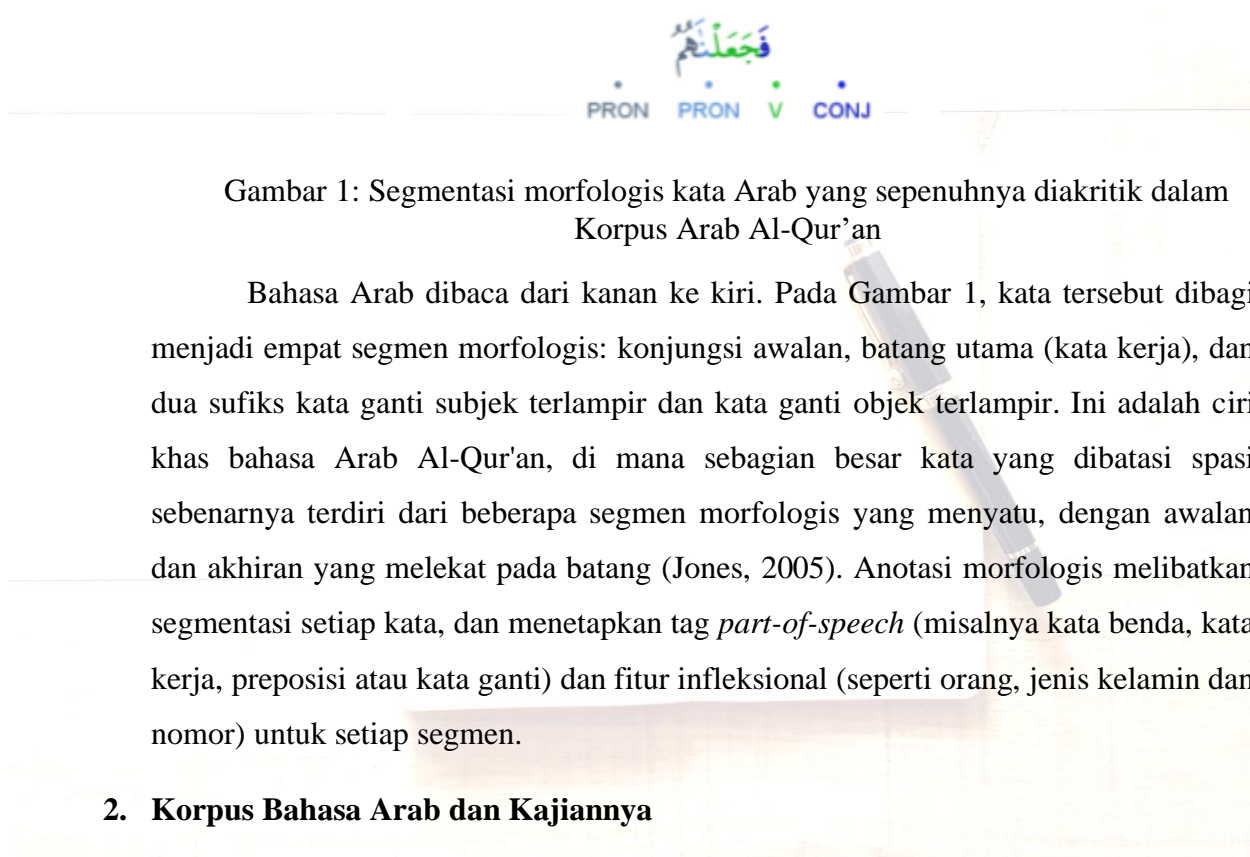
1. Mengolah bahasa Arab Al-Qur'an
2. Korpus bahasa Arab dan kajiannya
3. Pengolah teks Arab dengan aplikasi pengolah korpus
4. Proyeksi kajian linguistik korpus bahasa Arab di Indonesia

#### C. Uraian Materi

##### 1. Mengolah Bahasa Arab Al-Qur'an

Memproses Bahasa Arab Al-Quran merupakan tantangan unik dari sudut pandang komputasi, karena kosakata dan ejaan berbeda dari (MSA). Namun, Al-Qur'an Arab datang dengan keuntungan yang sepenuhnya diakritik, tidak seperti kebanyakan teks Arab lainnya. Setiap kata Al-Qur'an berisi diakritik (tanda) rinci di atas semua huruf yang menggambarkan vokalisasi yang tepat (lihat Gambar 1). Menggunakan informasi ini memberikan keuntungan pada anotasi otomatis jika dibandingkan dengan bentuk bahasa Arab lainnya.

Gambar 1 di bawah ini menunjukkan contoh kata dalam *Quranic Arabic Corpus*, seperti yang ditampilkan kepada pengguna situs web yang melihat anotasi morfologis. Tiga angka di atas gambar memberikan nomor bab, nomor ayat dan nomor kata. Qur'an dibagi menjadi 114 bab. Setiap bab berisi urutan ayat bernomor. Dalam contoh ini, anotasi sesuai dengan kata keempat dari pasal 21, ayat 70. Baris berikutnya pada Gambar 1 adalah terjemahan interlinear segmen, diikuti dengan terjemahan penuh dan transkripsi fonetik. Pengucapan yang ditampilkan diturunkan secara otomatis dari anotasi morfologis dan diakritik yang sudah ada dalam teks.



Gambar 1: Segmentasi morfologis kata Arab yang sepenuhnya diakritik dalam Korpus Arab Al-Qur'an

Bahasa Arab dibaca dari kanan ke kiri. Pada Gambar 1, kata tersebut dibagi menjadi empat segmen morfologis: konjungsi awalan, batang utama (kata kerja), dan dua sufiks kata ganti subjek terlampir dan kata ganti objek terlampir. Ini adalah ciri khas bahasa Arab Al-Qur'an, di mana sebagian besar kata yang dibatasi spasi sebenarnya terdiri dari beberapa segmen morfologis yang menyatu, dengan awalan dan akhiran yang melekat pada batang (Jones, 2005). Anotasi morfologis melibatkan segmentasi setiap kata, dan menetapkan tag *part-of-speech* (misalnya kata benda, kata kerja, preposisi atau kata ganti) dan fitur infleksional (seperti orang, jenis kelamin dan nomor) untuk setiap segmen.

## 2. Korpus Bahasa Arab dan Kajiannya

Secara khusus, sudah tersedia beberapa corpus bahasa Arab yang dibuat dengan segala kekhususan, kelebihan, dan tak ketinggalan kekurangannya masing-masing. Dalam penyusunan corpus bahasa Arab, melalui suatu aplikasi di laman Sketch Engine, misalnya ada corpus yang isinya lebih kurang 5,8 juta kata (versi 2012) dan diperbarui menjadi 8,3 juta kata (versi 2015) yang diambil dari sejumlah sumber. Selain itu Alsubaiti menginventarisasi daftar yang lebih lengkap lagi, sekitar 18 jenis corpus, yang berasal dari berbagai sumber dan digunakan untuk berbagai

bidang kajian spesifik yang ada di sub laman University of Leeds, diantaranya Corpus of Contemporary Arabic, Arabic Gigaword, dan yang sedang dalam pengembangan adalah International Corpus of Arabic oleh University of Alexandria, Mesir. Sekian banyak corpus itu terbagi ke dalam dua kelompok besar, yaitu yang dapat diunduh dengan membayar dan tak membayar.

Di samping corpus bahasa Arab, ada pula corpus khusus mengenai Al-Qur'an. Sejauh ini, menurut pengamatan kami, yang terlihat paling representatif adalah karya University of Leeds, yaitu The Quranic Arabic Corpus. Tidak hanya corpus 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 corpus Arab, dan khususnya Al-Qur'an 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 (Nur Hizbullah, Fazlurrahman, dan Fuzi Fauziah, 2016).

Salah satu langkah penyusunan corpus umum bahasa Arab yang mencantumkan Al-Qur'an sebagai salah satu sumbernya adalah yang dilakukan oleh Al-Rabia (2014). Corpusnya dinamakan *King Saud University Corpus of Classical Arabic (KSUCCA)*. Sesuai namanya, mereka hanya fokus mengambil corpus 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. Corpus 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, selain

menggambarkan struktur corpus, 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 Qur'an, corpus Al-Qur'an beranotasi khusus pada kasus anaphora pronomina. Penelitian itu diarahkan kepada analisis pronominal dalam Al-Qur'an dan pola kemunculan anaphora yang terkait dengan pronominal dan kaitan antesedennya dalam ayat-ayat Al-Qur'an. Model analisis penelitian ini bermanfaat untuk diterapkan pada analisis pronominal, anaphora, dan antesedennya dalam wacana bahasa Arab modern dan perbandingan kedua masa itu.

Salah satu contoh lain penelitian corpus Al-Qur'an yang lebih spesifik adalah karya Imad dan Abdelhak (2016). Mereka melakukan penelitian untuk menyusun corpus 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 isi peneliti sendiri.

Sebagai simpulan terhadap peneliti terdahulu, dapat dikatakan bahwa penggunaan pendekatan linguistik corpus 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 corpus.

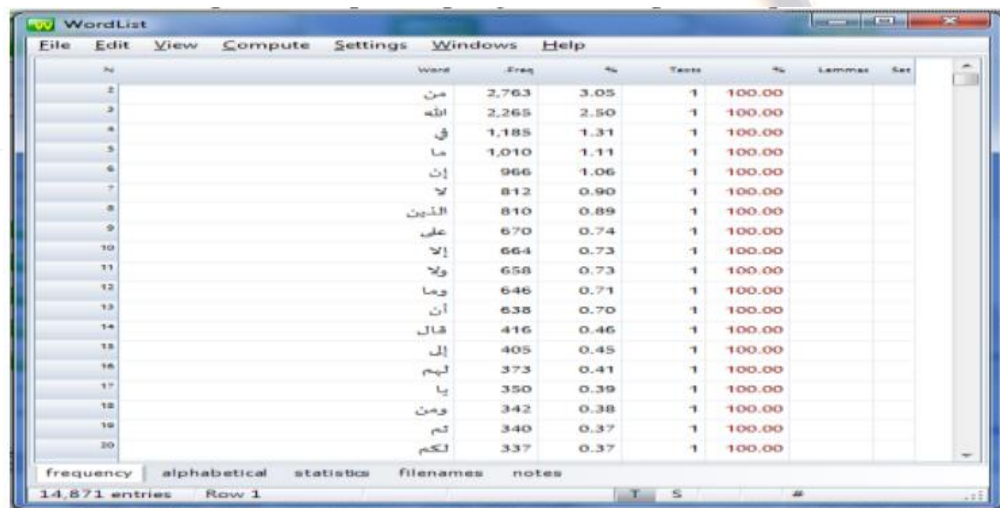
### 3. Pengolah Teks Arab dengan Aplikasi Pengolah Korpus

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 \*.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 corpus. Secara teknis, hanya file dengan ekstensi \*.txt dengan mode konversi UTF-8 saja yang dapat diolah dengan aplikasi corpus pada umumnya.

Khusus menyangkut teks Al-Qur'an, sebelum diolah dengan aplikasi bahan corpus 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-Qur'an, 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.



No	Word	Freq	%	Tests	%	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	ان	966	1.06	1	100.00		
7	لا	812	0.90	1	100.00		
8	الذين	810	0.89	1	100.00		
9	علي	670	0.74	1	100.00		
10	لا	664	0.73	1	100.00		
11	ولا	658	0.73	1	100.00		
12	وما	646	0.71	1	100.00		
13	ان	638	0.70	1	100.00		
14	قال	416	0.46	1	100.00		
15	الى	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 2: Contoh proses penyusunan daftar kata dan kemunculan hasil dalam peringkat 20 besar teratas

Concordance											
File Edit View Compute Settings Windows Help											
Concordance											
In		Out		Delta		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag		Delta		Delta		Delta	
File		Text		Tag</							

Gambar 3: Contoh proses penyusunan konkordansi dan hasilnya

#### 4. Proyeksi Kajian Linguistik Korpus Bahasa Arab Di Indonesia

Melihat pesatnya perkembangan produk ilmiah kajian bahasa Arab dengan menggunakan linguistik corpus 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 dan peminat kajian bahasa Arab di Indonesia untuk terbuka terhadap fakta itu, lalu lebih jauh dan lebih luas lagi memanfaatkan linguistik corpus dalam penelitian mereka.

Secara khusus, penulis melihat akan pentingnya keberadaan sebuah corpus bahasa Arab di Indonesia. Pilihan awal yang bisa diambil untuk tahap awal adalah penyusunan corpus pembelajar bahasa Arab di Indonesia. Menurut penulis, 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 corpus bahasa Arab yang bersumber dari teks-teks berbahasa Arab karya penutur Indonesia asli. Setelah corpus itu tersedia, dilakukanlah penelitian berbasis linguistik corpus 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, penulis menganggap perlu untuk lebih digalakkan lagi pendekatan linguistik corpus 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 corpus dalam penelitian kebahasaan. Namun, sudah sepantasnya para ahli bahasa Arab di Indonesia ikut menyumbangkan kepada dunia kajian bahasa Arab global suatu model corpus bahasa Arab dengan mengangkat khazanah sumber pustaka pembelajaran bahasa Arab dan sumber produk ilmiah asli putra-putri negeri ini.

#### D. Aktivitas Pembelajaran

Untuk lebih meningkatkan pemahaman anda tentang materi kegiatan 1 ini, disarankan untuk melakukan aktivitas-aktivitas sebagai berikut:

1. Rumuskan dalam mengolah bahasa Arab Al-Qur'an!
2. Bagaimana pandangan anda tentang korpus bahasa Arab dan kajiannya?
3. Berdasarkan uraian di atas tentang *Quranic Arabic Corpus* (QAC), manakah di antara uraian-uraian tersebut yang menurut anda lebih mudah untuk diterapkan? Berikan alasannya!
4. Melalui diskusi dengan rekan sejawat anda, bagaimana cara anda menerapkan *Quranic Arabic Corpus* (QAC) pada mata kuliah Terjemah? Apabila menemukan hal-hal yang sama dan berbeda dengan rekan sejawat anda, dimanakah persamaan dan perbedaannya? Hasil kerja anda dan rekan sejawat anda tentang poin-poin yang dikerjakan dapat dituliskan pada kolom analisa anda dan analisa rekan sejawat anda.
5. Dalam melakukan aktivitas poin d, anda dapat menggunakan format lembar kerja di bawah:

**Lembar Kerja 1: *Quranic Arabic Corpus* (QAC)**

No.	Konsep yang diperdalam	Analisa anda	Analisa rekan sejawat
1.	Prinsip individual		
2.	Prinsip pengalaman dalam pengaplikasian <i>Quranic Arabic Corpus</i> (QAC)		
3.	Prinsip totalitas		
4.	Prinsip aktivitas mandiri		

**E. Rangkuman**

Memproses Bahasa Arab Al-Quran merupakan tantangan unik dari sudut pandang komputasi, karena kosakata dan ejaan berbeda dari (MSA). Namun, Al-Qur'an Arab datang dengan keuntungan yang sepenuhnya diakritik, tidak seperti kebanyakan teks Arab lainnya. Secara khusus, sudah tersedia beberapa corpus bahasa Arab yang dibuat dengan segala kekhususan, kelebihan, dan tak ketinggalan kekurangannya masing-masing. Khusus menyangkut teks Al-Qur'an, sebelum diolah dengan aplikasi bahan corpus 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. Di sisi yang lebih umum, penulis menganggap perlu untuk lebih digalakkan lagi pendekatan linguistik corpus sebagai titik tolak dalam penelitian linguistik dalam rangka memaksimalkan pemanfaatan data-data kebahasaan Arab terkini yang sudah lebih banyak berbentuk digital.



## F. Latihan/Tugas

### 1. Latihan

1.  $\text{قُلْ هُوَ اللَّهُ أَحَدٌ}$  Dhomir pada kata  $\text{هُوَ}$  dalam surat al-ikhlas ayat 1 adalah...

- a. She
- b. He
- c. Is
- d. Don't
- e. Has not

2.  $\text{قُلْ هُوَ اللَّهُ أَحَدٌ}$  Bentuk fi'il dari kata  $\text{قُلْ}$  adalah ....

- a. فعل أمر
- b. فعل ماضى
- c. فعل مضارع
- d. فعل ناهي
- e. مصدر

3.  $\text{يَا أَيُّهَا الَّذِينَ آمَنُوا كُتِبَ عَلَيْكُمُ الْقِصَاصُ فِي الْقَتْلَى}$  Kata  $\text{كُتِبَ}$  dari surat Al-Baqarah ayat

178 merupakan ....

- a. فعل ماضى مبني للمنصوب
- b. فعل ماضى مبني للمرفوع
- c. فعل ماضى مبني للمجهول
- d. فعل ماضى
- e. مصدر

4. Bentuk *ruba'i* ( رَبَاعِ ) dari kata  $\text{دَخَلَ}$  yang benar dari kata pilihan dibawah ini

adalah ....

- a.  $\text{يَدْخُلُونَ}$
- b.  $\text{يَدْخُلُونَ}$
- c.  $\text{أَدْخَلُوا}$

- d. أَذْخُلُوا  
e. يَدْخُلُ
5. قُلْ أَعُوذُ بِرَبِّ النَّاسِ Kata النَّاسِ dari surat An-Nas ayat 1 merupakan ....
- a. اسم مكان  
b. اسم زمان  
c. مصدر  
d. فعل ماضى  
e. اسم مجرور

## 2. Kunci Jawaban

1. B
2. A
3. C
4. A
5. E

## G. Umpan Balik dan Tindak Lanjut

Anda sebaiknya mempelajari kembali semua jawaban dari soal latihan yang telah dikerjakan. Jawaban anda tersebut dicocokkan dengan rambu-rambu jawaban yang telah dibuat dengan uraian materi, ada baiknya anda sudah dipandang sesuai dengan materi yang ada dalam modul, anda dapat meneruskan mempelajari ke materi selanjutnya. Namun apabila jawaban anda masih belum dengan rambu-rambu jawaban sebagaimana tertuang dalam uraian materi, anda disarankan untuk mempelajari kembali bagian materi yang dipandang belum lengkap.