Dr. Nur Hizbullah, S.Ag., M.Hum.
https://nur-hizbullah.my.id
Prodi Bahasa dan Kebudayaan Arab, Universitas Al Azhar Indonesia

Korpus untuk Riset dan Pembelajaran Bahasa Arab

***Workshop* Aplikasi Korpus**

"Perencanaan Riset dan Pembelajaran Bahasa Arab Berbasis Korpus bagi Dosen"

Program Studi Pendidikan Bahasa Arab, Fakultas Tarbiyah dan Ilmu Keguruan

UIN Raden Intan Lampung

Selasa, 19 Juli 2022

Bagian 1

Pengenalan dan Pengantar tentang Korpus

Definisi “Korpus”

- Korpus → *corpus* (sing., Latin ‘*body*’), *corpora* (pl.)
- Korpus:
  - wujud penggunaan bahasa yang riil dan otentik;
  - berupa bahasa lisan dan tulisan;
  - dipilih dengan kriteria tertentu;
  - representasi gambaran penggunaan bahasa oleh seseorang maupun sekelompok orang;
  - terdokumentasikan secara sistematis;
  - dapat diolah dengan pendekatan mekanistik-komputasi;
  - dapat dikaji dengan berbagai pendekatan keilmuan yang relevan.

Definisi “Linguistik Korpus”

- ‘Study of language based on examples of “real life” language use’ and ‘a methodology rather than an aspect of language requiring explanation or description’ (McEnery and Wilson, 1996:1)

Jenis-jenis Korpus

No	Berdasarkan	Jenis 1	Jenis 2	Jenis 3
1	Konten	Umum	Khusus	-
2	Periode Bahasa	Historis	Modern	-
3	Penutur	Penutur jati	Pembelajar asing	Regional
4	Konten bahasa	Monolingual	Paralel	-
5	Ragam bahasa	Lisan	Tulisan	Campuran
6	Jenis material	Teks	Teks + suara	Teks + audiovisual
7	Format tulisan	Ortografis	Beranotasi	-

https://sites.google.com/a/aucegypt.edu/infoguistics/directory/Corpus-Linguistics/arabic-corpora

Potensi Korpus Berbahasa Arab di Indonesia …

1. Korpus pendidikan: buku ajar, karya pengajar/pembelajar, media pembelajaran, dll
2. Korpus kebahasaan: artikel, paper, laporan penelitian, tugas akhir, dsb.
3. Korpus studi Islam: berbagai karya intelektual di bidang studi Islam
4. Korpus karya sastra: puisi, prosa, novel, dsb.
5. Korpus karya populer: berita, reportase, iklan, tulisan populer, dsb.

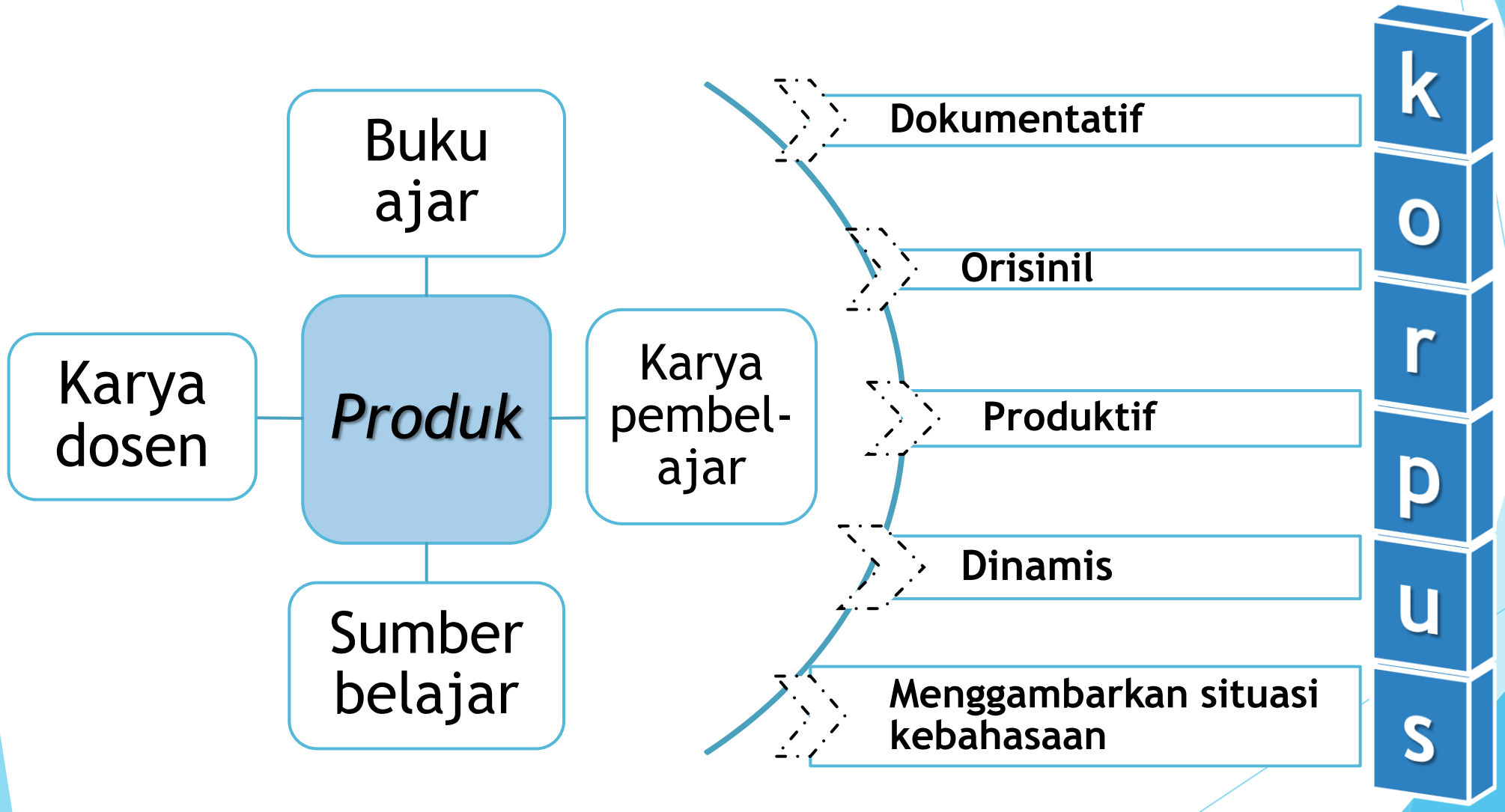
Potensi Korpus Berbahasa Arab di Indonesia (Hizbullah & Muchlis, 2017)

6. Korpus Al-Quran dan terjemahannya: Al-Quran terjemahan bahasa Indonesia dan bahasa-bahasa nusantara
7. Korpus Hadis dan terjemahannya: Hadis terjemahan bahasa Indonesia dan bahasa-bahasa nusantara
8. Korpus dokumen: berbagai dokumen terjemahan Arab-Indonesia

++ daftar kata serapan Arab-Indonesia

https://www.atlantis-press.com/proceedings/icclas-17/25891004

Produk + Karakteristik = KORPUS



Bagian 2

Ihwal dan Karakteristik Riset dan Pembelajaran Bahasa Arab Berbasis Korpus

Karakteristik Penelitian Berbasis Korpus

- Bersifat empiris, fokus kepada penggunaan bahasa yang riil dalam teks yang alami
- Lebih memadai jika volume korpus data cukup besar
- Memanfaatkan komputer dan aplikasi dengan mekanisme tertentu
- Memadukan teknik analisis kuantitatif dan kualitatif (Yuliawati, 2018:7)

Korpus untuk Riset & Pembelajaran BA

Objek	Pemanfaatan
Karya pembelajar	• Penjaringan bahan dan pendokumentasian korpus secara berkelanjutan • Telaah karakteristik & kualifikasi bahasa pembelajar
Dinamika bahasa pembelajar	• Telaah perkembangan kemampuan bahasa pembelajar • Evaluasi pengajaran dengan bahan/metode tertentu
Sumber belajar potensial	• Telaah bahan untuk dibuat bahan ajar • Penyusunan kamus kosakata

Contoh Riset tentang Korpus

Judul	Tautan
Corpus Linguistics and Corpus-Based Research and Its Implication in Applied Linguistics: A Systematic Review	https://ejournal.undip.ac.id/index.php/parole/article/view/31104
Corpora and language teaching	https://lexically.net/wordsmith/corpus_linguistics_links/Roemer%20208%20HSK%20CL%20chapter%20final%20print%20version.pdf
Corpora and Language Learning: Ways to Use Corpora in Classroom	https://journal.uny.ac.id/index.php/lingua-pedagogia/article/view/38169
Corpus linguistics and languageteaching research: bridging the gap	https://www.academia.edu/7292966/Corpus_linguistics_and_language_teaching_research_bridging_the_gap
Corpus Linguistics in LanguageTeaching	https://www.academia.edu/7292961/Corpus_Linguistics_in_Language_Teaching
Corpora: From theoretical linguistics to language teaching	http://seminar.uad.ac.id/index.php/utic/article/view/5731
Corpus Linguistics, Language Corpora, and Language Teaching	http://ejournal.uika-bogor.ac.id/index.php/ENGLISH/article/view/3845
Corpus for Classrooms: Ideas for Material Design	https://dbe.metu.edu.tr/convention/proceedingsweb/Corpus.pdf
Corpus-Based Approaches in Language Teaching: Outcomes, Observations, and Teacher Perspectives	https://dergipark.org.tr/en/download/article-file/1177080
Arabic-English Parallel Corpus: A New Resource for Translation Training and Language Teaching	https://awej.org/arabic-english-parallel-corpus-a-new-resource-for-translation-training-and-language-teaching/
Corpus-Based Pedagogical Materials Development: Guidelines for Teachers and Sample MICASE-Based Materials	https://www.press.umich.edu/pdf/0472030728-part7.pdf

Contoh Riset Berbasis Korpus

Dynamic of Arabic Academic writing in Postgrad UIN Malang.pdf

Sumber : http://ejournal.uin-malang.ac.id/index.php/ijazarabi/article/view/13251/9188

Korpus Arab pesantren.pdf

Sumber : https://ejournal.unida.gontor.ac.id/index.php/tadib/article/viewFile/7067/pdf_82

Kamus kedokteran Arab berbasis korpus.pdf

Sumber : https://tsaqofiya.iainponorogo.ac.id/index.php/tsaqofiya/article/view/37/16

Compiling vocab list for arabic tourism teaching.pdf

Sumber : https://journal.uinjkt.ac.id/index.php/arabiyat/article/view/20818

UAS Kamus_Linguistik Arab Terapan_Mujahidah Fharieza R.pdf

Beberapa Bacaan Tambahan

Corpora and Language Teaching (Ed. Karin Aijmer), 2009: John Benjamis

How to Use Corpora in Language Teaching (Ed. John McH. Sinclair, 2004: John Benjamins

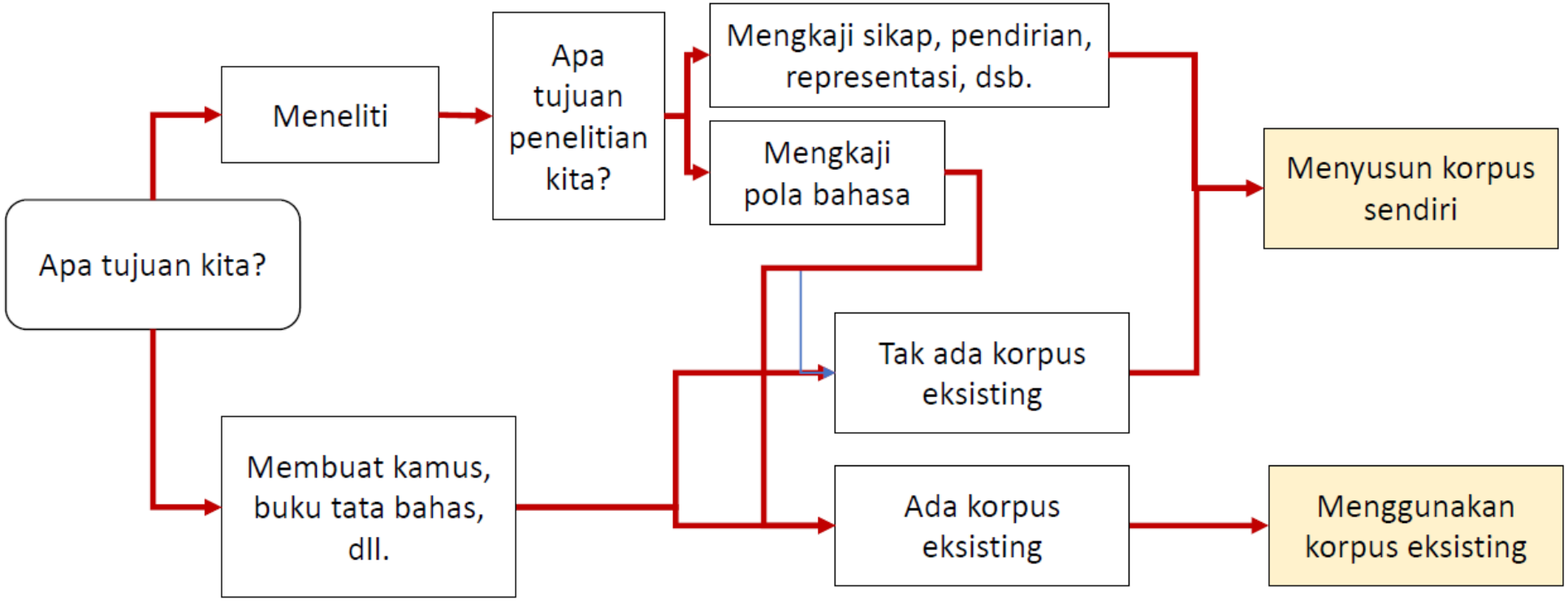
Textual Patterns: Key Words and Corpus Analysis in Language Education (Mike Scott & Christopher Tribble), 2006: John Benjamins

Bagian 3

Pemanfaatan, Pembuatan, dan Pengolahan Korpus

Opsi Penggunaan Korpus

Korpus Eksisting >< Korpus Buatan Sendiri



Memanfaatkan Korpus yang Ada

- https://www.arabiclearnercorpus.com/
- https://www.arabiclearnercorpus.com/about-the-corpus-en
- https://arabicorpus.byu.edu/ / https://auth.sketchengine.eu/#login

Membuat Korpus Sendiri

Contoh aplikasi korpus sederhana (& gratis) untuk analisis korpus buatan sendiri

- https://laurenceanthony.net/software/antconc/
- https://sourceforge.net/projects/linguistic-analyzer/files/latest/download

Langkah Membuat Korpus Sendiri …

1. Memilih teks (*.pdf, *.bok, dll.) → *.doc/x

- Bahan korpus sebisa mungkin berformat dokumen (*.doc/x)
- Jika belum, perlu dikonversi dulu menjadi format *.doc/x

Langkah Membuat Korpus Sendiri …

2. Mengolah dan “membersihkan” teks (*.doc/x)

- Bahan korpus “dibersihkan” dari gambar, tabel, nomor halaman, *header* dan *footer*, bingkai halaman, catatan kaki, karakter “*superscript*” dan “*subscript*”, *hyperlink*, kode HTML, dll. yang dapat “mengganggu” pembacaan teks oleh aplikasi

Langkah Membuat Korpus Sendiri …

3. Berikan metadata pada data di tiap awal teks:

   <judul>...</judul>
   <tahun>...</tahun>
   <jenis>...</jenis>
   dll

Langkah Membuat Korpus Sendiri

4. Mengonversi ke format *plain text* (*.txt):

- save as → plain text
- gunakan "*other encoding*"
- pilih "UTF-8"
- tutup *file* teks tsb
- buka *file* teks melalui aplikasi notepad
- *file* siap diolah dengan aplikasi korpus

Fungsi Pengolahan pada Mesin Korpus

- **Daftar/frekuensi kata:**
  - Daftar kata yang paling banyak/sering muncul dalam sebuah korpus yang dikumpulkan
- Tipe: variasi kata tunggal (unik/spesifik) yang muncul dalam sebuah korpus
- Token: jumlah total kemunculan kata dalam sebuah korpus

Fungsi Pengolahan pada Mesin Korpus

- **Konkordansi kata/*Keyword in Context* (KWIC):**
  - penyajian suatu kosakata dalam konteks, bisa untuk membandingkan pola-pola penggunaan kata atau struktur secara cepat

Fungsi Pengolahan pada Mesin Korpus

- **N-Gram:**
  - urutan dua kata atau lebih yang muncul berulang dalam sebuah korpus dan memiliki signifikansi secara statistik

Fungsi Pengolahan pada Mesin Korpus

- Kolokasi:
  - kombinasi dua kata atau lebih yang biasanya digunakan oleh penutur bahasa; terdiri atas kata poros (*node*) dan kata pasangan (*collocate*)-nya

Fungsi Pengolahan pada Mesin Korpus

- Kata kunci:
  - daftar kata yang unik/spesifik yang lebih banyak muncul pada suatu korpus

إن أريد إلا الإصلاح ما استطعت
وما توفيقي إلا بالله

Terima kasih & Takzim

© Dr. Nur Hizbullah, S.Ag., M.Hum. << https://nur-hizbullah.my.id >>
Prodi Bahasa dan Kebudayaan Arab, Universitas Al Azhar Indonesia