

Penggunaan Formula Kueri dalam Pengembangan Purwarupa Kamus Tingkat Tutur Bahasa Bali

Ida Wayan Eka Werdi Putra
Universitas Udayana
ekawerdi7@gmail.com

Abstrak

Makalah ini memaparkan pemahaman dasar formula kueri pada perangkat lunak pengolah data Google Sheet dalam membangun model purwarupa kamus digital tingkat tutur bahasa Bali. Berdasarkan data yang diperoleh dari korpus Majalah Suara Saking Bali, formula KUERI Google Sheet dan pendukungnya digunakan untuk membuat berbagai konfigurasi perintah terhadap basis data leksikon yang kemudian disusun dalam suatu model kamus tingkat tutur bahasa Bali yang interaktif dan efisien. Adapun beberapa klausul KUERI dan fungsi pendukung lainnya yang dapat digunakan dalam penyusunan kamus tingkat tutur bahasa Bali antara lain SELECT, WHERE, FILTER, dan IFS. Kelebihan penggunaan fungsi KUERI pada program aplikasi web Google Sheet memiliki kelebihan antara lain: kustomisasi perintah, mudah diakses, penyimpanan berbasis komputasi awan, kolaboratif, hingga mampu mengolah data bahasa yang besar. Pemahaman dasar terhadap konsep KUERI dalam pengembangan purwarupa kamus digital tingkat tutur bahasa Bali ini diharapkan mampu menjadi langkah awal dalam pengembangan berbagai model linguistik terapan lainnya dalam media digital

Kata Kunci: Kueri, Korpus, Leksikografi, Bahasa Bali, Tingkat Tutur, Linguistik Terapan

Abstract

This paper aims to describes the basic understanding of query formulas on Google Sheet data processing software in building a prototype model of a Balinese spoken-language digital dictionary. Based on data obtained from the Suara Saking Bali Magazine corpus, the Google Sheet QUERY formula and its supporting elements are used to create various command configurations for the lexicon database which are then arranged into an interactive and efficient Balinese spoken-language dictionary model. Some QUERY clauses and other supporting functions that can be used in compiling a Balinese spoken-language dictionary include SELECT, WHERE, FILTER, and IFS. The advantages of using the QUERY function in the Google Sheet web application program include: command customization, easy to access, cloud-based storage, collaborative, and the ability to process large language data. A basic understanding of the QUERY concept in developing a prototype of a Balinese spoken-language digital dictionary is expected to be the first step in developing various other applied linguistic models in digital media

Keywords: Query, Corpus, Lexicography, Balinese, Speech Level, Applied Linguistics

1. Pendahuluan

Perpaduan antara ilmu linguistik dengan teknologi komputer telah melahirkan suatu interdisiplin ilmu yang disebut dengan linguistik komputasional. Computational

Linguistics atau Linguistik komputasional didefinisikan sebagai bidang interdisipliner yang menggabungkan aspek linguistik dan ilmu komputer untuk mempelajari bahasa menggunakan metode komputasi (Mitkov, 2004). Fokusnya adalah mengembangkan algoritma dan model yang dapat memproses dan menganalisis bahasa alami. Adapun beberapa produk berbasis aplikasi linguistik komputasional yang sering digunakan antara lain Google Terjemahan, Alexa, Siri, hingga dalam kecerdasan buatan seperti ChatGPT dan Gemini. Seiring dengan perkembangan dunia leksikografi, ada beberapa perangkat lunak komputer yang dapat membantu pekamus dalam mengolah data leksikon, mulai dari aplikasi penyusun kamus seperti Lexique Pro, hingga bahasa pemrograman seperti R, Python, SQL yang memberi keleluasaan fungsi dalam membangun sebuah model komputasi bahasa. Khusus dalam hal pengembangan model atau purwarupa berbasis teori linguistik, kendala yang dihadapi seorang linguist maupun pekamus dalam menggunakan program komputasi R, Python, maupun SQL umumnya adalah kompleksitas bahasa pemrograman yang memerlukan kompetensi dalam merangkai kode. Solusi dan alternatif yang dapat ditempuh adalah menggunakan program populer dengan padanan fitur memadai seperti Google Sheet. Google Sheet menyediakan fitur-fitur pengolah data yang lazim digunakan berbagai kalangan seperti pengolah angka, rangkaian karakter (string), hingga perintah-perintah KUEI yang dapat dikustomisasi dalam rangka mengakomodasi kebutuhan penggunaannya, termasuk dalam bidang leksikografi. Makalah ini akan mengulas pemahaman dasar formula KUEI pada perangkat lunak pengolah data Google Sheet dalam membangun model purwarupa Kamus (Digital) Tingkat Tutar Bahasa Bali (KTTBB). Purwarupa KTTBB dipilih karena bahasa Bali memiliki keunikan variasi tingkat tutur yang mencerminkan variable spesial yang tidak dapat diaplikasikan pada program pengolah kamus lain seperti Lexique Pro dan AntConc. Tujuan makalah ini adalah memaparkan fitur kueri pada Google Sheet beserta turunannya, dalam membantu merumuskan model purwarupa (dalam hal ini berupa kamus digital KTTBB) berdasarkan teori dan asas-asas linguistik.

2. Data dan Metode

Purwarupa Kamus Tingkat Tutar Bahasa Bali (PKTTBB) menggunakan sumber data utama yang berasal dari korpus Majalah Suara Saking Bali Edisi 1-64. Data korpus tersebut kemudian mengalami dua kali proses ekstraksi yang menghasilkan DATA PRIMER dan DATA SEKUNDER. DATA PRIMER yang akan digunakan sebagai

acuan lema bahasa sumber pada purwarupa kamus diperoleh dengan memproses data korpus menggunakan teknik *Word Frequency List* pada perangkat lunak AntCont Seri 4.3.1 (Anthony, 2005). Informasi tambahan terkait kelas kata, bahasa sasaran, tingkat tutur, dan lainnya disesuaikan berdasarkan Kamus Bahasa Bali-Indonesia konvensional dan juga intuisi peneliti. DATA SEKUNDER yang akan digunakan sebagai contoh penggunaan lema pada kamus diperoleh dengan melakukan proses ekstraksi kalimat pada sumber data korpus utama menggunakan fungsi pemisah rangkaian karakter (*string*) bahasa pemrograman R. Fungsi pemisah kalimat tersebut dapat diakses melalui tautan <https://tukangkorp.us.shinyapps.io/shiny> (Putra, 2024). Ekstraksi data korpus menjadi daftar kalimat bertujuan untuk menyediakan contoh penggunaan kata dalam setiap lema kamus yang dicari oleh pengguna. Ragam tingkat tutur bahasa Bali memungkinkan adanya beberapa lema dengan kandungan makna leksikal yang sama, namun memiliki varian tingkat tutur yang berbeda. Kamus Bahasa Bali-Indonesia terbitan Balai Bahasa Bali yang dijadikan rujukan tingkat tutur pada penelitian ini memuat beberapa tingkat tutur yang dilekatkan pada beberapa lema, antara lain: Asi (*Alus Singgih*), Aso (*Alus Sor*), serta Ami (*Alus Mider*). Selain itu, tambahan ragam tingkat tutur juga diperoleh dari Kamus Daring BASAbali Wiki dengan tambahan tingkat tutur *Alus Madia*, *Andap*, *Kasar*, dan *Mider*.

3. Hasil

3.1 Komputerisasi Leksikon Bahasa Bali

Berbagai definisi dan klasifikasi tingkat tutur dari penelitian terdahulu mengidentifikasi adanya sebuah kaidah kesantunan yang mengatur level berbahasa. Kaidah tersebut direalisasikan berdasarkan nilai rasa bahasanya dalam beberapa jenis tingkat tutur (anggha-ungguhing) bahasa Bali, antara lain (1) basa kasar, (2) basa andap, (3) basa madia, dan (4) basa alus (Narayana, 1983). Pada tataran leksikal, dokumentasi terhadap leksikon bahasa Bali berdasarkan tingkat tuturnya telah termuat di beberapa sumber dan penelitian, dua di antaranya adalah Kamus Daring Bahasa Bali-Indonesia yang diterbitkan oleh Balai Bahasa Provinsi Bali serta Kamus Daring BASAbali Wiki. Kedua kamus tersebut memuat entri leksikon bahasa Bali beserta terjemahannya yang dilengkapi label singkatan tingkat tutur untuk kata-kata tertentu yang diklasifikasikan sebagai; Alus Singgih (Asi), Alus Sor (Aso), Alus Mider (Ami), dan Kasar (Ks). Entri-entri tersebut berasal dari berbagai kelas kata seperti nomina (n), adjektiva (a), verba (v), pronomina (pro), numeralia (n), dan adverbialia (adv). Selain menyediakan bahasa

sumber, bahasa sasaran, dan informasi tingkat tutur pada setiap lema, hal menarik dari kedua kamus tersebut adalah tersedianya contoh penggunaan lema dalam bentuk frasa, klausa, maupun kalimat. Kamus Daring Bahasa Bali-Indonesia terbitan Balai Bahasa Bali menggunakan contoh penggunaan lema yang berasal dari kamus konvensional. Artinya contoh kalimat tersebut bersifat definit dan tidak dapat disunting maupun ditambah oleh pengguna. Mekanisme yang lebih fleksibel digunakan oleh Kamus Daring BASAbali Wiki dalam penyediaan contoh penggunaan lemanya. Ada dua cara yang digunakan, yaitu penambahan contoh kalimat oleh pengguna terdaftar, serta penarikan contoh kalimat dari data korpus Ruang Komunitas yang menjadi satu kesatuan dalam situs BASAbali Wiki. Sistem inilah yang akan diuji coba menggunakan fitur kueri pada perangkat lunak Google Sheet.

3.2 Kueri dan Google Sheet

Sama seperti fungsi pada perangkat lainnya (misalnya SQL), fungsi KUEI di Google Sheets adalah alat yang sangat serbaguna dan memungkinkan pengguna untuk memanipulasi serta menganalisis data menggunakan sintaks (Cunha et al., 2013). Fungsi ini memungkinkan berbagai operasi data, seperti menyaring, mengurutkan, mengelompokkan, dan mengagregasi data di beberapa kolom. Dengan KUEI, pengguna dapat memilih baris tertentu yang memenuhi kondisi tertentu, menerapkan operator logika, dan menggunakan klausa seperti WHERE, ORDER BY, LIMIT, dan GROUP BY untuk mengoperasikan data yang diinginkan. Selain itu, fungsi ini juga dapat melakukan perhitungan seperti penjumlahan, rata-rata, dan menghitung jumlah data. Berbeda dengan alat penyaringan biasa di spreadsheet, KUEI memberi lebih banyak kontrol atas bagaimana data diambil, diformat, dan disajikan. Fungsi ini sangat berguna untuk mengelola kumpulan data yang besar, mengotomatiskan laporan, atau membuat dasbor yang dinamis. Dengan mengombinasikan KUEI dengan fitur lain di Google Sheets, pengguna dapat menyederhanakan alur kerja, memastikan data yang selalu akurat dan disesuaikan dengan kebutuhan, tanpa perlu perangkat lunak basis data eksternal atau pengetahuan pemrograman yang mendalam. Secara umum, berikut adalah rangkaian argumen atau sintaks dalam menjalankan KUEI di Google Sheet:

SINTAKSIS DASAR KUEI

=KUEI (data; kueri; [header])

data

- Rentang sel tempat menjalankan kueri

- Setiap kolom data hanya dapat mengandung nilai boolean, numerik (termasuk tanggal/waktu) atau string (rangkaian karakter).
- Untuk jenis data campuran dalam satu kolom, jenis data mayoritas menentukan jenis data kolom untuk tujuan kueri.

kueri

- Sintaksis Bahasa Kueri Google Visualization API dirancang agar serupa dengan sintaksis SQL. Bahasa kueri terdiri dari klausa-klausa yang dimulai dengan satu atau dua kata kunci. Semua klausa bersifat opsional.

Tabel 1. Klausul Kueri Goole Sheet

ausul	nggunaan
lect	emilih kolom yang akan ditampilkan, dan urutannya. Jika dihilangkan, semua kolom tabel akan ditampilkan, dalam urutan defaultnya.
ere	nya menampilkan baris yang cocok dengan kondisi. Jika dihilangkan, semua baris akan ditampilkan.
oup by	enggabungkan nilai di seluruh baris.
vot	engubah nilai yang berbeda dalam kolom menjadi kolom baru.
der by	engurutkan baris menurut nilai dalam kolom.
mit	embatasi jumlah baris yang ditampilkan.
fset	elewati sejumlah baris pertama tertentu.
bel	enetapkan label kolom.
rmat	emformat nilai dalam kolom tertentu menggunakan pola pemformatan yang diberikan.
tions	enetapkan opsi tambahan.

Sumber :<https://developers.google.com/chart/interactive/docs/querylanguage?hl=id>, 2024

header

- [OPSIONAL] - Jumlah baris header di bagian atas data. Jika dihilangkan atau ditetapkan ke -1, nilai diterka berdasarkan isi data.

4. Pembahasan

Pembangunan purwarupa KTTBB pada makalah ini melalui tiga tahapan, yaitu tahap penyusunan data, tahap penyusunan fungsi, serta tahap visualisasi. Pada tahap penyusunan data, data primer dan data sekunder dibedakan dalam dua lembar kerja

(Sheet). Data primer berisikan leksikon bahasa Bali beserta berbagai informasinya dalam suatu format tabular. Sedangkan data sekunder berupa daftar kalimat bahasa Bali yang telah diekstraksi dari data korpus Majalah Suara Saking Bali.

	A	D	C	D
1	Arti	Entri	Kelas Kata	Tingkat Tutur
2	saya	tiang	n	ama
3	saya	titiang	n	aso
4	saya	cang	n	andap
5	kamu	cai	n	andap
6	kamu	ci	n	andap
7	dia	ia	n	andap
8	dia	ida	n	asi
9	dia	dane	n	asi
10	siana	sira	pro	ami

Gambar 1. Data Primer

	A	D	C	D	E	F	G
1	Ada anak uyeng-uyengan, misi utah-utah ulian bas liunan nginum tuak, ento madan punyah.						
2	Yen suba punyah, makejang tepukin soleh, makejang tepukin tawah.						
3	Di papunyanan tepuk kone gumine jag maplincer kija kaden.						
4	Tusing karoan dija kangin kauh.						
5	Salingke kangin kauh, menek tuun suba tusing pasti.						
6	Sirahe ane sabilang wai tongosne baduur, saget prajani bisa nungkalik beten macelep ka jelinjingane.						
7	Jelinjingan tusing pelih apa-apa ento bakat dengik aji mantran tuak.						
8	Ane madan mantran tuak, tuah misi aksara Uh ajak eh.						
9	Uhhhhhh Ehhhhhhkkkk...!!! Nyak mesib cara munyin modre ane tusing nyidayang baan ngundukang aji papineh kejoka						

Gambar 2. Data Sekunder

Penyusunan data pada purwarupa KTTBB ini merupakan bentuk adaptasi teori pemerolehan bahasa dalam bidang psikolinguistik, yaitu teori Innateness Hypothesis (Chomsky, 1965). Teori Keberadaan Bawaan ini berpendapat bahwa manusia dilahirkan dengan kemampuan bawaan untuk memperoleh bahasa, termasuk kosakata. Tata bahasa universal yang dimiliki manusia sejak lahir membantu mereka mengasosiasikan kata-kata dengan struktur linguistik yang sesuai, sehingga mereka dapat memahami dan menyimpan kata-kata dengan cepat dalam memori. Manusia dianggap memiliki perangkat bawaan yang memungkinkan mereka untuk secara otomatis memetakan kata-kata dari bahasa yang didengar atau didokumentasikan.

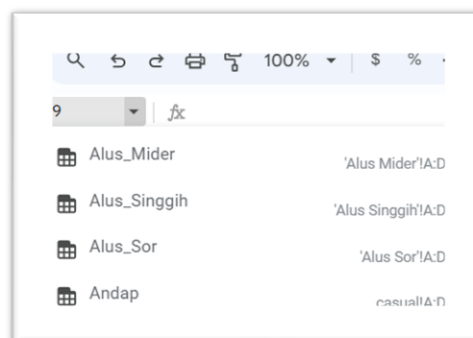
Pemetaan yang dilakukan pada purwarupa KTTBB adalah pemetaan tingkat tutur leksikon bahasa Bali.

Pada tahap penyusunan data lanjutan, fungsi KUERI dengan klausul SELECT dan WHERE digunakan untuk memetakan DATA PRIMER berdasarkan tingkat tutur bahasa Balinya. Terdapat lima jenis data primer yang dihasilkan dan termuat dalam lima lembar kerja (Sheet) berbeda, yaitu data leksikon andap, alus_sor, alus_singgih, serta alus_mider. Adapun fungsi KUERI yang digunakan untuk masing-masing lembar kerja antara lain:

Tabel 2. Kueri Penyusunan Data Primer

NAMA LEMBAR	KUERI
Andap	=QUERY('DATA PRIMER'!A:D, "select A,B,C,D where D contains 'andap' ")
Alus_Sor	=QUERY('DATA PRIMER'!A:D, "select A,B,C,D where D contains 'Aso' ")
Alus_Singgih	=QUERY('DATA PRIMER'!A:D, "select A,B,C,D where D contains 'Asi' ")
Alus_Mider	=QUERY('DATA PRIMER'!A:D, "select A,B,C,D where D contains 'Ami' ")

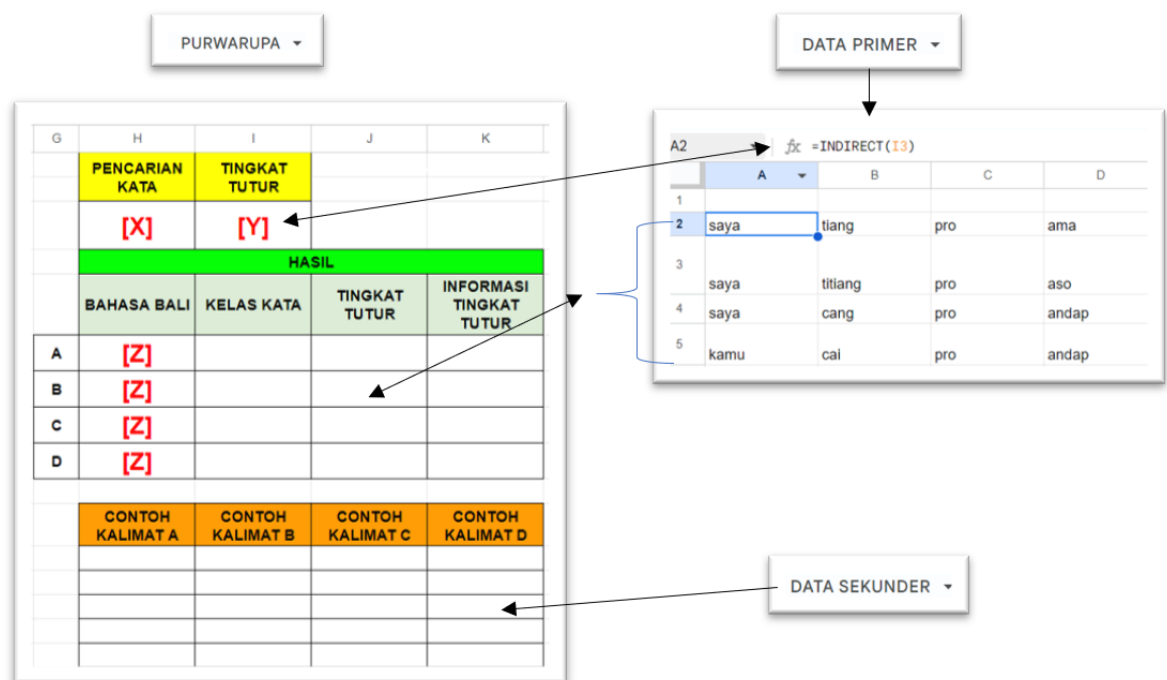
Apabila kita berasumsi bahwa kata kunci atau string klausul WHERE dalam pemilahan tingkat tutur adalah X, maka kode KUERI pada kolom sebelumnya dapat dinarasikan sebagai perintah “Carilah pada lembar kerja DATA PRIMER, yaitu kolom A,B,C,D dengan ketentuan bahwa kolom D mengandung string/rangkaian karakter X”. Kelebihan penggunaan fungsi KUERI dalam pemilahan data berdasarkan tingkat tutur ini memiliki keunggulan dalam hal integrasi data. Integrasi yang dimaksud adalah segala perubahan pada DATA PRIMER akan berdampak pada masing-masing data berdasarkan tingkat tutur. Semua data primer tersebut dibuatkan rentang bernama [Named Range] untuk memudahkan tahap penyusunan fungsi seperti berikut :



Gambar 3. Rentang Bernama [Named Range] untuk setiap data tingkat tutur. Rentang Bernama ini

akan memudahkan proses indeks [=INDIRECT]

Tahap penyusunan fungsi disesuaikan dengan kebutuhan pengguna KTTBB. Tahap penyusunan fungsi dalam purwarupa KTTBB ini mengadopsi salah satu teori psikolinguistik yang menjelaskan bagaimana manusia memilih kosakata dalam kamus mental yang disebut Teori Pemilihan Dua Tahap (*Two-Stage Model*) (Dell, 1986). Menurut teori ini, pemilihan kosakata dalam kamus mental dilakukan melalui dua tahap utama, yaitu tahap pemilihan lema, serta tahap pemilihan fonologis. Implementasi pemilihan lema dan bentuk fonologis pada KTTBB ini diasosiasikan dalam pembuatan purwarupa mesin pencarian terhadap lema kamus pada DATA PRIMER serta contoh penggunaannya pada DATA SEKUNDER. Berikut adalah penyusunan fungsi purwarupa KTTBB :



Gambar 4. Skema Penyusunan Fungsi KTTBB

Apabila dibuat persamaan bahwa nilai [PENCARIAN KATA] adalah [X], nilai pilihan [TINGKAT TUTUR] adalah [Y], serta hasil pencarian bahasa Bali adalah [Z], maka sintaksis umum dari Purwarupa KTTBB pada gambar di atas adalah :

Tampilkan hasil pencarian DATA PRIMER berdasarkan kata kunci rangkaian karakter [X] dan kelompok [Y], serta tampilkan DATA SEKUNDER yang mengandung rangkaian karakter [Z] .

Konsep dasar yang membedakan Purwarupa KTTBB dengan kamus bahasa Bali

(Daring/Luring) lainnya adalah adanya fitur pemilihan tingkat tutur pada hasil pencarian yang direalisasikan oleh nilai Y, serta penyajian contoh penggunaan berdasarkan hasil pencarian Z. Fitur ini akan memungkinkan pengguna dalam menyeleksi tingkat tutur kosakata berdasarkan konteks penggunaan lema yang dicari (misalnya untuk menyusun suatu kalimat dengan tingkat tutur tertentu) secara lebih efisien. Konsep ini dapat dikustomisasi lebih lanjut sesuai keinginan pekamus, seperti seleksi hasil pencarian berdasarkan kelas kata, maupun variabel lainnya. Penyusunan Purwarupa KTTBB ini menggunakan beberapa fungsi, antara lain :

Tabel 3. Fungsi-Fungsi dalam Penyusunan KTTBB

FUNGSI	TUJUAN	SINTAKSIS
FILTER	Menyeleksi dan menyajikan hasil pencarian DATA PRIMER berdasarkan kata kunci rangkaian karakter X. Menghasilkan Z	=FILTER("Rentang DATA PRIMER" ; "Kolom Kata Kunci" = "X")
QUERY	Menyeleksi dan menyajikan hasil pencarian DATA SEKUNDER berdasarkan kata kunci rangkaian karakter Z.	=QUERY("Rentang DATA SEKUNDER"; "SELECT * WHERE DATA SEKUNDER contains Z ")
IFS	Membuat suatu rangkaian kondisi berdasarkan hasil Z terhadap penyediaan informasi terkait tingkat tutur bahasa Bali.	=IFS ('Z'='Andap'; 'Deskripsi Andap'; 'Z'='Ami; 'Deskripsi Ami'; 'Z'='Asi; 'Deskripsi Asi'; 'Z'='Aso; 'Deskripsi Aso'; 'Z'='Ami; 'Deskripsi Ami')

5. Kesimpulan

Penggunaan perangkat lunak komputer dalam memproses data kebahasaan membuka peluang dan potensi baru dalam berbagai kajian maupun penerapan teori linguistik. Salah satunya adalah dalam pembuatan purwarupa Kamus Tingkat-Tutur Bahasa Bali menggunakan formula KUERI pada aplikasi pengolah data Google Sheet. Setidaknya ada dua kendala yang sering ditemui peneliti bahasa dalam hal kolaborasi antara komputasi dan linguistik. Kendala pertama yang umum ditemui adalah kendala teknis, terlebih pada perangkat lunak yang memerlukan kompetensi bahasa pemrograman. Hal tersebut dapat diatasi dengan menggunakan perangkat lunak populer Google sheet yang memiliki beberapa kelebihan, antara lain tidak berbayar, mudah diakses, kolaboratif, serta menunjang fungsi-fungsi penting seperti KUERI. Fungsi KUERI pada Google Sheet memungkinkan pengguna untuk mengolah data dengan berbagai kustomisasi

perintah berdasarkan teori linguistik yang relevan melalui klausul-klausul seperti SELECT dan WHERE. Pembuatan Purwarupa KTTBB dalam makalah ini hanyalah contoh kecil dan pengantar awal dari potensi penggunaan komputer dalam menjawab berbagai permasalahan linguistik berbasis data.

6. Daftar Pustaka

- Anthony, L. (2005). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. *IPCC 2005. Proceedings. International Professional Communication Conference, 2005.*, 729–737.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Cunha, J., Fernandes, J. P., Mendes, J., Pereira, R., & Saraiva, J. (2013). Querying model-driven spreadsheets. *2013 IEEE Symposium on Visual Languages and Human Centric Computing*, 83–86.
- Dell, G. S. (1986). A Spreading Activation Theory of Retrieval in Sentence Production. *Psychological Review*.
- Mitkov, R. (2004). *The Oxford Handbook of Computational Linguistics*. Oxford University Press.
- Putra, I. W. E. W. (2024). *Tukang Korpus: Data Processing Tool*. Retrieved from <https://tukangkorporus.shinyapps.io/Shiny/>.