

**IDENTITAS LINGUAL *SHORT MESSAGE SERVICE (SMS) SOBIS:*
ANALISIS LINGUISTIK KORPUS DENGAN
APLIKASI ANTCONC**



KARTIKA

F011201028

DEPARTEMEN SASTRA INDONESIA

FAKULTAS ILMU BUDAYA

UNIVERSITAS HASANUDDIN

MAKASSAR

2024



Optimized using
trial version
www.balesio.com

**IDENTITAS LINGUAL IDENTITAS LINGUAL *SHORT MESSAGE SERVIS*
(SMS) *SOBIS*: ANALISIS LINGUISTIK KORPUS DENGAN APLIKASI
*ANTCONC***

**KARTIKA
F011201028**



**DEPARTEMEN SASRA INDONESIA
FAKULTAS ILMU BUDAYA
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**



Optimized using
trial version
www.balesio.com

**IDENTITAS LINGUAL IDENTITAS LINGUAL *SHORT MESSAGE SERVIS*
(SMS) *SOBIS*: ANALISIS LINGUISTIK KORPUS DENGAN APLIKASI
*ANTCONC***

KARTIKA

F011201028

Skripsi

sebagai salah satu syarat untuk mencapai gelar sarjana

Departemen Sastra Indonesia

pada

DEPARTEMEN SASTRA INDONESIA

FAKULTAS ILMU BUDAYA

UNIVERSITAS HASANUDDIN

MAKASSAR

2024



SKRIPSI

IDENTITAS LINGUAL IDENTITAS LINGUAL *SHORT MESSAGE SERVIS* (SMS) *SOBIS*: ANALISIS LINGUISTIK KORPUS DENGAN APLIKASI *ANTCONC*

KARTIKA

F011201028

Skripsi,

telah dipertahankan di depan Panitia Ujian Sarjana pada 4 Desember 2024 dan dinyatakan telah memenuhi syarat kelulusan



pada

Departemen Sastra Indonesia
Fakultas Ilmu Budaya
Universitas Hasanuddin
Makassar

Mengesahkan:
Pembimbing tugas akhir,



id, M.Hum.
92031032

Mengetahui:
Ketua Program Studi,

Dr. Munira Hasjim, S.S., M.Hum.
NIP 197105101998032001

PERNYATAAN KEASLIAN SKRIPSI DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, skripsi berjudul "Identitas Lingual *Short Message Servis (SMS) Sobis: Analisis Linguistik Korpus dengan Aplikasi AntConc*" adalah benar karya saya dengan arahan dari pembimbing Dr. Ikhwan M. Said, M.Hum. Karya ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka skripsi ini. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini adalah karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut berdasarkan aturan yang berlaku.

Dengan ini saya melimpahkan hak cipta (hak ekonomis) dari karya tulis saya berupa skripsi ini kepada Universitas Hasanuddin.

Makassar, 6 Desember 2024



METERAI
TEMPEL
KARTIKA
00DA3AMX085601817 111201028



Optimized using
trial version
www.balesio.com

UCAPAN TERIMA KASIH

Puji dan syukur penulis panjatkan doa atas berkat dan rahmat Allah Swt. Tidak lupa juga salawat serta salam penulis ucapkan kepada Baginda Rasulullah saw. Skripsi berjudul "Identitas Lingual *Short Message Servis (SMS) Sobis: Analisis Linguistik Korpus dengan Aplikasi AntConc*" bertujuan memenuhi persyaratan ujian skripsi guna memperoleh gelar Sarjana Sastra pada Departemen Sastra Indonesia, Fakultas Ilmu Budaya, Universitas Hasanuddin.

Penyusunan skripsi ini tentu tidak lepas dari berbagai hambatan. Namun, semua itu teratasi berkat uluran tangan dari berbagai pihak berupa bantuan, dorongan, dan juga bimbingan kepada penulis. Penulis menyadari adanya berbagai kekurangan dan kesalahan dalam skripsi ini. Sehubungan dengan hal itu, penulis selalu membuka diri untuk menerima masukan dan kritikan dari berbagai pihak sebagai upaya penyempurnaan skripsi ini.

Penulis mengucapkan terima kasih atas semua bantuan yang telah diberikan, baik secara langsung maupun tidak langsung selama menyusun skripsi ini. Untuk itu penulis mengucapkan terima kasih kepada semua pihak yang terlibat sebagai berikut.

1. Dr. Ikhwan M.Said, M.Hum., selaku Pembimbing Skripsi dan Pembimbing Akademik yang dengan kesabaran hatinya telah memberikan bimbingan, meluangkan waktu di tengah kesibukan beliau, memberikan kritik, saran, motivasi yang tidak pernah putus, dan mengarahkan penulis dalam penulisan skripsi ini.
2. Prof. Dr. Lukman, M.S., selaku Penguji I atas segala bimbingan dan saran untuk kelengkapan skripsi dan juga arahan serta motivasinya sehingga penulis dapat menyelesaikan perbaikan skripsi dengan baik.
3. Prof. Dr. Asriani Abbas, M.Hum., selaku Penguji II atas segala bimbingan dan saran untuk kelengkapan skripsi ini, serta telah meluangkan waktunya untuk memberikan arahan dalam penulisan skripsi ini.
4. Dr. Munira Hasjim, S.S., M.Hum. selaku Ketua Departemen Sastra Indonesia Fakultas Ilmu Budaya, Universitas Hasanuddin dan Rismayanti, S.S., M. Hum. selaku Sekretaris Departemen Sastra Indonesia Fakultas Ilmu Budaya, Universitas Hasanuddin.
5. Seluruh Dosen di Departemen Sastra Indonesia dan Dosen di Fakultas Ilmu Budaya Universitas Hasanuddin yang telah banyak memberikan bimbingan dan ilmu pengetahuan selama penulis mengikuti perkuliahan sehingga dapat



memotivasi dan mendukung, serta membimbing penulis dalam mengarungi kehidupan ini.

8. Saudara kandung ibu penulis AstutiYanti, Amd.Keb., dan Rivayana, yang selalu mendukung dan memberi bantuan pada saat masa-masa sulit yang dilalui penulis, memberikan banyak dukungan, dan juga doa baik untuk penulis.
9. Teman baik sekaligus kerabat jauh yang penulis temui sejak perkuliahan di Kota Makassar, Abdul Muqaddim Rizal, terima kasih selalu percaya dan mendukung mimpi-mimpi penulis.
10. Teman-teman seperjuangan meraih gelar sarjana, Dyah Muliasari, Andi Tenri Pada, Nur Natasya Salsabila, Nadilla Najamuddin, Nurhamdani, Muh.Alief, Fadilla Umma Syam, Muhammad Akhdan Abizar Anwar, Rian Arman, Akhry Fadly, Ardianyah Badaruddin, Andi Audia Faiza Nazli I, Reyhan Prasanto, Muhizar Nugraha, dan Putri Ayu Lestari yang saling membantu dan memberikan semangat selama masa perkuliahan.
11. Ilham Wasi, S.S, M.Hum., yang telah memotivasi dan menjadi panutan bagi penulis, baik dalam karir jurnalistik maupun dalam penulisan skripsi ini.
12. Keluarga besar IMSI KMFIB-UH, yang telah menjadi rumah yang hangat untuk tumbuh dan berkembang, terima kasih sudah menjadi wadah bagi penulis selama perkuliahan, terima kasih atas semua pengalaman berharga yang telah dilewati.
13. Teman-teman Adaptasi 2020, yang telah memberikan banyak hal seru dan berkesan selama perkuliahan. Terima kasih telah menjadi keluarga baru yang pertama kali menerima penulis di Departemen Sastra Indonesia.
14. Mace-mace Di Tallo, Khususnya kepada Mace Zaenab dan Mace Syamsiah, yang telah menjadi orang tua bagi penulis selama di Kota Makassar. Semoga keadilan segera tiba mengetuk pintu rumah kalian.
15. Anak-anak di Bara-Baraya, yang sudah mengajak penulis melihat tanah tempat bermain kalian. Setelah menyelesaikan skripsi ini, penulis akan melanjutkan liputan tentang perjuangan kalian.
16. Seluruh korban kekerasan dan pelecehan seksual, di mana pun 'aman' itu berada, pastikan bahwa kita tidak pernah berhenti menciptakan risau di kepala.

Penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan dan kesalahan. Dengan segala kerendahan hati, penulis menerima dengan baik segala bentuk masukan dan kritik untuk diperbaiki pada masa mendatang.

Makassar, 6 Desember 2024



Kartika

DAFTAR ISI

	Halaman
JUDUL	ii
HALAMAN PENGANTAR.....	iii
HALAMAN PENGESAHAN.....	iv
HALAMAN PERSETUJUAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN SKRIPSI	v
UCAPAN TERIMA KASIH.....	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xi
DAFTAR ISTILAH	xii
DAFTAR SINGKATAN.....	xiii
ABSTRAK	xiv
<i>ABSTRACT</i>	xv
BAB I. PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	2
1.3 Batasan Masalah	3
1.4 Rumusan Masalah	3
1.5 Tujuan Penelitian	4
1.6 Manfaat Penelitian	4
BAB II. TINJAUAN PUSTAKA.....	5
2.1 Landasan Teori	5
2.1.1 Linguistik Umum	5
2.1.1.1 Kata	6
2.1.1.2 Frasa	6
2.1.1.3 Klausa.....	7
2.1.2 Linguistik Terapan	7
orpus	8
ic	9
<i>Language Processing (NLP)</i>	21
.....	23
Sobis	24



2.2	Hasil Penelitian yang Relevan	25
2.3	Kerangka Pikir	26
BAB III METODE PENELITIAN		28
3.1	Jenis dan Pendekatan Penelitian.....	28
3.2	Tempat dan Waktu Penelitian.....	28
3.3	Sumber Data	28
3.3.1	Populasi dan Sampel.....	29
3.4	Metode dan Teknik Pengumpulan Data.....	29
3.5	Metode Analisis Data	29
3.5.1	Pengumpulan Data	29
3.5.2	Pereduksian Data	29
3.5.3	Pentranskripsian Data	30
3.5.4	Penginputan Data	30
3.5.5	Penyajian Data	31
3.5.6	Penyimpulan dan Pemverifikasian.....	31
BAB IV. HASIL DAN PEMBAHASAN.....		32
4.1	Identifikasi Pola Lingual (kata, frasa, dan klausa) <i>Sobis</i> Melalui Pemanfaatan <i>AntCont</i>	32
4.1.1	Identifikasi Kata yang Paling Sering Digunakan oleh <i>Sobis</i> Menggunakan Alat Kata (<i>Tools Word</i>)	32
1.	Kata 'anda'	33
2.	Kata 'info'	34
3.	Kata 'selamat'	34
4.	Kata 'klik'	34
5.	Kata 'dptkan'	34
4.1.2	Pemindaian N-Gram untuk Menemukan Pola Frasa atau Klausa Berulang <i>Sobis</i>	34
1.	Frasa 'nomor coin pulsa'.....	35
2.	Frasa 'bit ly kejutang'	35
3.	Klausa 'tu info klik'	36
	a m dptkan'	36
	ar jadi pulsa'	36
	lan Pengelompokan Pola SMS <i>Sobis</i> menggunakan alat Key Word KWIC)	37



1.	HADIAH PANIK_MENDESAK	38
2.	HADIAH LINK_MENCURIGAKAN TATA_BAHASA_SALAH	38
3.	HADIAH TATA_BAHASA_SALAH.....	38
4.	PANIK_MENDESAK	39
5.	INFO_PINJOL, TATA_BAHASA_SALAH	39
6.	PANIK_MENDESAK NO_TELEPON.....	40
4.2	Daftar Penanda Lingual Identitas <i>Sobis</i>	42
4.2.1	Kelompok Ke-1	43
4.2.2	Kelompok Ke-2	44
4.2.3	Kelompok ke-3	45
4.2.4	Kelompok ke-4	46
4.2.5	Kelompok ke-5	47
4.2.6	Kelompok ke-6	47
BAB V. PENUTUP		48
5.1	Simpulan	48
5.2	Saran	49
DAFTAR PUSTAKA.....		50
LAMPIRAN		51



DAFTAR GAMBAR

Gambar 2. 1 Alat KWIC pada <i>AntConc</i>	10
Gambar 2. 2 Alat Plot pada <i>AntConc</i>	11
Gambar 2. 3 Alat View pada <i>AntConc</i>	13
Gambar 2. 4 Alat Cluster pada <i>AntConc</i>	14
Gambar 2. 5 Alat N-Gram pada <i>AntConc</i>	15
Gambar 2. 6 Alat Collocate pada <i>AntConc</i>	16
Gambar 2. 7 Alat Word pada <i>AntConc</i>	18
Gambar 2. 8 Alat Keyword pada <i>AntConc</i>	19
Gambar 2. 9 Alat Wordcloud pada <i>AntConc</i>	20
Gambar 3. 1 Sembilan alat pada aplikasi <i>AntConc</i>	30
Gambar 4. 1 Tampilan analisis data alat Word pada Aplikasi <i>AntConc</i>	33
Gambar 4. 2 Wordcloud data laporan SMS <i>Sobis</i>	32
Gambar 4. 3 Diagram jumlah kelompok dan ekspresi umum SMS <i>Sobis</i>	41
Gambar 4. 4 alat kolokasi pada Aplikasi <i>AntConc</i>	42



DAFTAR ISTILAH

- Barcode* : Dalam konteks linguistik atau komputasi, dapat merujuk pada pola berulang atau identifikasi unik, tetapi perlu klarifikasi lebih lanjut untuk makna khusus.
- Cluster* : Kelompok elemen yang serupa, biasanya dalam data atau analisis linguistik, seperti pengelompokan kata dengan pola tertentu.
- Cybercrime* : Tindak kejahatan yang dilakukan menggunakan teknologi komputer atau internet, seperti penipuan online atau phishing.
- Entri : Item atau unit data tunggal dalam basis data, korpus, atau teks.
- Executable* : File atau program yang dapat dieksekusi langsung oleh komputer untuk menjalankan fungsi tertentu.
- File View* : Fitur dalam perangkat lunak yang memungkinkan pengguna melihat isi file atau struktur data.
- Fraud* : Penipuan atau tindakan curang untuk mendapatkan keuntungan, sering ditemukan dalam cybercrime.
- Hosting phishing* : Server atau layanan yang digunakan untuk menjalankan serangan phishing, misalnya situs web palsu untuk mencuri data.
- Machine learning* : Teknologi yang memungkinkan komputer mempelajari data dan membuat prediksi atau pengambilan keputusan berdasarkan pola.
- Malware* : Perangkat lunak berbahaya yang dirancang untuk merusak, mencuri, atau mengganggu sistem komputer.
- Plot* : Visualisasi data dalam bentuk grafik atau diagram.
- Purposive sampling* : Teknik pengambilan sampel berdasarkan tujuan tertentu, sering digunakan dalam penelitian kualitatif.
- Pyinstaller* : Alat Python untuk mengubah skrip menjadi file executable mandiri.
- Python* : Bahasa pemrograman tingkat tinggi yang populer untuk berbagai aplikasi, termasuk analisis data.
- Qt : Framework pengembangan aplikasi GUI (antarmuka pengguna grafis) yang sering digunakan dengan Python.
- Range* : Kisaran atau distribusi nilai dalam data atau analisis.
- Shobiz* : Kependeakan dari show business, merujuk pada industri hiburan.
- SQLite : Database ringan yang sering digunakan dalam aplikasi kecil atau prototipe.
-  : Unit terkecil dalam analisis teks, seperti kata, simbol, atau karakter.
-  : Alat atau perangkat lunak yang digunakan untuk mendukung analisis atau proses lainnya.

DAFTAR SINGKATAN

AILA	:	Asosiasi Internasional Linguistik Terapan, organisasi global yang berfokus pada penelitian linguistik terapan.
APJII	:	Asosiasi Penyelenggara Jasa Internet Indonesia, organisasi yang menaungi penyelenggara jasa internet di Indonesia.
BERT	:	BiDirectional Encoder Representations from : Transformer, model pembelajaran mesin berbasis transformer yang digunakan dalam NLP untuk memahami konteks kata dalam teks.
CTRL	:	Kemungkinan merujuk pada tombol Control pada keyboard komputer
KWIC	:	format tampilan kata kunci dalam analisis korpus yang menampilkan kata dalam konteks penggunaannya di teks.
Min. Freq	:	Minimum Frequency, frekuensi minimum yang ditentukan dalam analisis data atau korpus untuk menyaring kata atau elemen yang jarang muncul.
NLP	:	Natural Language Processing, cabang kecerdasan buatan yang memproses dan menganalisis bahasa manusia secara alami menggunakan teknologi komputer.
SMS	:	Short Message Service, layanan pengiriman pesan singkat melalui jaringan telepon.
TESOL	:	Teaching English as a Second Language, program atau metode pengajaran bahasa Inggris kepada non-penutur asli.



ABSTRAK

KARTIKA, *Identitas Lingual Short Message Service (SMS) Sobis: Analisis Linguistik Korpus dengan Aplikasi AntConc* (dibimbing oleh **Ikhwan M.Said**).

Berbeda dengan pendekatan matematis berbasis pembelajaran mesin yang lazim digunakan oleh para ahli statistik untuk mendeteksi penipuan, penelitian ini menawarkan analisis linguistik untuk memahami lebih dalam pola dan penanda lingual pada teks SMS *Sobis*. Tujuan penelitian ini adalah mengidentifikasi identitas lingual SMS *Sobis* serta menyajikan daftar penanda lingual *Sobis*. Dari 100 SMS aduan, dapat diidentifikasi 2.546 token (kata) dan 882 kata unik. Analisis dilanjutkan dengan mengolah frasa dan klausa menggunakan alat **N-Gram** dilanjutkan dengan pengklasifikasian menggunakan alat **KWIC**, sehingga menghasilkan 73 tipe N-Gram dengan 338 token yang menunjukkan pola komunikasi berulang *Sobis*.

Pengelompokan terhadap 73 tipe N-Gram menghasilkan enam kelompok penanda lingual *Sobis* yang merepresentasikan strategi lingual tertentu berdasarkan analisis kata, frasa, dan klausa. Selanjutnya, dilakukan analisis menggunakan alat Collocation terhadap keenam kelompok tersebut, sehingga diketahui bahwa keenam kelompok menunjukkan keakraban dengan kata-kata tertentu. Kata-kata tersebut muncul bersamaan sebagai kolokasi tiap kelompok, yaitu kelompok ke-1 memiliki 13 daftar penanda lingual yang digunakan untuk memberi iming-iming hadiah dengan cara mendesak penerima pesan, kelompok ke-2 memiliki 48 daftar penanda lingual yang dengan ciri khusus terdapat kesalahan penulisan atau tata bahasa dan ada perintah untuk mengklik tautan, kelompok ke-3 memiliki 46 penanda lingual dengan memberi iming-iming hadiah diikuti tata bahasa yang salah, kelompok ke-4 memiliki 12 penanda lingual dengan perintah desakan, kelompok ke-5 memiliki tiga penanda lingual dengan penanda dominan menyebut sebuah nama perusahaan simpan pinjam, dan kelompok ke-6 memiliki delapan penanda lingual yang digunakan untuk mendesak diikuti arahan untuk menghubungi sebuah nomor telepon.

Kata kunci: *Sobis*, linguistik korpus, *AntConc*, kolokasi



ABSTRACT

KARTIKA, *Lingual Identity of Sobis Short Message Service (SMS): A Corpus Linguistic Analysis Using AntConc Application* (supervised by **Ikhwan M. Said**).

Unlike the mathematical approach based on machine learning commonly used by statisticians to detect fraud, this study offers a linguistic analysis to gain a deeper understanding of the patterns and linguistic markers in Sobis SMS texts. The aim of this research is to identify the linguistic identity of Sobis SMS and present a list of its linguistic markers. From 100 complaint SMS messages, 2,546 tokens (words) and 882 unique words were identified. The analysis continued with processing phrases and clauses using the N-Gram tool, followed by classification using the KWIC tool, resulting in 73 types of N-Grams comprising 338 tokens that reveal recurring communication patterns of Sobis.

The grouping of these 73 N-Gram types resulted in six categories of Sobis linguistic markers that represent specific linguistic strategies based on the analysis of words, phrases, and clauses. Furthermore, a Collocation analysis was conducted on these six groups, revealing their association with certain keywords. These words frequently appeared together as collocations within each group. The findings for each group are as follows, group 1, consists of 13 linguistic markers used to offer rewards with a sense of urgency directed at the recipient; group 2, Includes 48 linguistic markers characterized by spelling or grammatical errors, often accompanied by instructions to click a link; group 3, contains 46 linguistic markers that combine reward offers with incorrect grammar. group 4; comprises 12 linguistic markers emphasizing urgency in instructions; group 5, contains three linguistic markers predominantly referencing the name of a financial lending company; group 6 Includes eight linguistic markers used to urge recipients while providing instructions to contact a specific phone number.

Keywords: Sobis, corpus linguistics, AntConc, collocation



BAB I PENDAHULUAN

1.1 Latar Belakang

Penggunaan pesan singkat atau *Short Message Service* (SMS) sebagai sarana teknologi, informasi, dan komunikasi telah berkembang pesat dalam beberapa tahun terakhir. SMS tidak hanya digunakan untuk keperluan pribadi, tetapi juga untuk keperluan transaksi bisnis, perbankan, dan layanan publik. Sebagai contoh, pada tahun 2020, pemerintah melalui Keputusan Menteri Kesehatan Nomor HK.01.07/Menkes/12757/2020 mulai mengirimkan pesan kepada para penerima vaksin Covid-19 menggunakan SMS (Dzulfaroh dan Ingried 2021, hlm. 2-10). Keterjangkauan menjadi faktor utama sehingga SMS digunakan sebagai saluran komunikasi paling dasar dengan total pengguna lebih dari 2,4 miliar atau hampir 75 persen pelanggan seluler dunia (Mulyawan 2023, hlm. 4).

Meskipun SMS memberikan kemudahan, peningkatan aktivitas penggunaan SMS beberapa tahun terakhir cenderung memunculkan masalah bagi pengguna seperti penipuan (*fraud*). Kementerian Kominfo mencatat, selama bulan Agustus sampai dengan pertengahan November 2023, terdapat 958 laporan kasus penyalahgunaan telepon dan SMS untuk penipuan online (Kominfo 2023, hlm. 2). Hal ini berbanding lurus dengan hasil survei Asosiasi Penyelenggara Jasa Internet Indonesia (APJII) yang mengumumkan adanya peningkatan penetrasi internet di Indonesia yang menyentuh angka 79,5 persen dengan total pengguna internet Indonesia tahun 2024 mencapai 221.563.479 jiwa dari total populasi 278.696.200 jiwa penduduk Indonesia tahun 2023 (APJII 2024, hlm 1). Selain itu, serangan digital di Indonesia meningkat pada kuartal I 2024 (Januari-Maret). Berdasarkan data lembaga Southeast Asia Freedom of Expression Network (SAFEEnet), terdapat 61 insiden keamanan digital di Indonesia pada kuartal I 2024. Rincian jumlah insiden tersebut adalah 13 insiden pada Januari, 20 insiden pada Februari, dan 27 insiden pada Maret 2024 (SAFEEnet 2024, hlm. 1-2).

Fenomena SMS penipuan (*fraud*) oleh sebagian masyarakat sipil khususnya di daerah Sulawesi Selatan dikenal dengan istilah *Sobis*. Istilah *Sobis* tercatat dalam penelitian Ahmad (2022, hlm. 12) sebagai kependekan dari 'sosial bisnis'. Istilah ini cukup populer digunakan oleh masyarakat untuk menggambarkan tindakan penipuan dengan menggunakan perantara internet, telepon, dan SMS. Beberapa penelitian lain yang membahas tentang fenomena *Sobis* diantaranya penelitian Wardana, dkk (2021), Ulfah, dkk (2022), Asrul (2022), dan Fausiah (2023).

Salah satu cara untuk menanggulangi permasalahan SMS *Sobis* adalah



akan teknik klasifikasi pembelajaran mesin (*machine learning*). Teknik semacam ini sebenarnya sudah cukup banyak dilakukan dengan berbagai metode algoritma statistik, seperti Support Vector Naïve Bayes (NB), C4.5, PART, dan K-Nearest Neighbor (KNN) (hlm. 635). Perbandingan beragam metode algoritma tersebut telah

dilakukan oleh Irawan dan kawan-kawan pada tahun 2021. Mereka mencoba menyaring dan memisahkan SMS penipuan dan bukan penipuan.

Teknik klasifikasi SMS penipuan dan bukan penipuan melalui penerapan metode algoritma semacam itu kemudian diadopsi oleh penelitian ini ke dalam pengembangan model linguistik komputasional. Dalam hal ini melibatkan pengumpulan dan analisis data bahasa (korpus) berupa SMS *Sobis*.

Istilah linguistik komputasional atau linguistik korpus digunakan untuk menggambarkan kajian linguistik terapan yang mengandalkan data bahasa (korpus) untuk menjawab pertanyaan tentang berbagai istilah bahasa, seperti deskripsi suatu ragam bahasa, teori-teori bahasa, proses pemerolehan bahasa, dan penggunaan bahasa dalam konteks tertentu. Oleh karena itu, linguistik korpus dipilih dalam penelitian ini, sehingga upaya untuk memodelkan metode algoritma pada penelitian berdasarkan data kebahasaan (SMS *Sobis*) tidak hanya sebatas representasi 'matematis' saja, melainkan juga pemahaman lebih jauh tentang sebuah struktur bahasa ataupun identitas lingual seorang penutur (*Sobis*).

Linguistik korpus dianggap sebagai metode yang sangat efektif untuk menerapkan konteks sebuah kata atau ekspresi (Ermanto,dkk 2022, hlm. 16). Oleh karenanya, penerapan linguistik korpus akan sangat berguna untuk mengembangkan algoritma atau metode analisis korpus yang dapat memberikan informasi berupa penanda atau identitas lingual *Sobis*.

Pemanfaatan linguistik korpus dalam penelitian ini dilakukan dengan bantuan aplikasi *AntConc*. *AntConc* adalah sebuah aplikasi pencarian korpus yang diciptakan oleh Laurence Anthony dari Universitas Waseda, Jepang. Prihantoro (2022) dalam bukunya menyebutkan *AntConc* sebagai salah satu daftar program analisis korpus yang paling banyak digunakan. Aplikasi ini sangat ringan dan sederhana, serta memiliki kecepatan yang baik untuk mengolah dan mencari data bahasa. Penggunaan *AntConc* dalam penelitian ini bertujuan untuk mempermudah mengamati struktur linguistik dan kecenderungan kolokasi pada sebuah SMS untuk muncul di SMS lainnya dalam frekuensi yang lebih tinggi. Kolokasi menjadi fitur penting yang terdapat di aplikasi *AntConc* untuk membantu penelitian ini mengidentifikasi kata-kata yang secara alami sering digunakan *Sobis* melakukan penipuan.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan di atas, penelitian ini akan mengkaji identitas lingual *Short Message Service* (SMS) *Sobis* melalui pemanfaatan linguistik korpus menggunakan aplikasi *AntConc*.



masalah yang muncul dalam konteks penelitian ini dapat sebagai berikut.

anda atau identitas lingual *Short Message Service* (SMS) *Sobis* ketahui agar terhindar dari penipuan.

2. Terdapat cara untuk mengidentifikasi *Short Message Service (SMS) Sobis* selain menggunakan algoritma statistik, yaitu dengan pemanfaatan ilmu bahasa, yaitu linguistik korpus.
3. Terdapat hubungan timbal balik antara linguistik korpus dan kolokasi yang dapat membantu mengidentifikasi *Short Message Service (SMS) Sobis*.
4. Terdapat pola bahasa yang sering digunakan dalam SMS *Sobis* yang mungkin berbeda dari SMS biasa. Identifikasi pola ini penting untuk memahami strategi bahasa yang digunakan dalam pesan penipuan.
5. Perlu diketahui sejauh mana pendekatan linguistik korpus dapat menjadi alat yang efektif dalam mengidentifikasi pesan penipuan dibandingkan metode lain yang ada.
6. Terdapat kendala dalam mengumpulkan dan menganalisis data bahasa sms *Sobis*, termasuk aksesibilitas data, validitas, dan teknik pengumpulan yang akurat untuk memastikan hasil analisis yang representatif.
7. SMS *Sobis* mungkin memiliki ciri-ciri bahasa yang berbeda berdasarkan lokasi atau latar belakang dialek penutur, yang berpotensi memengaruhi hasil identifikasi linguistik.
8. Ada kemungkinan bahwa beberapa SMS penipuan lebih kompleks dan sulit diidentifikasi hanya dengan analisis linguistik dasar, sehingga perlu diketahui keterbatasan pendekatan ini dan potensi untuk mengombinasikan metode lain.

1.3 Batasan Masalah

Berdasarkan identifikasi masalah yang ditemukan, penelitian ini dibatasi hanya pada:

1. cara mengidentifikasi identitas lingual *Short Message Service (SMS) Sobis* melalui pemanfaatan linguistik korpus menggunakan aplikasi *AntConc*;
2. daftar identitas lingual *Short Message Service (SMS) Sobis* melalui pemanfaatan linguistik korpus menggunakan aplikasi *AntConc*.

1.4 Rumusan Masalah

Berdasarkan batasan masalah di atas, maka dapat dirumuskan masalah penelitian sebagai berikut.

1. Bagaimana pengidentifikasian identitas lingual *Short Message Service (SMS) Sobis* ?
2. Apa saja daftar lingual *Short Message Service (SMS) Sobis*?



1.5 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, diharapkan dapat dicapai tujuan penelitian sebagai berikut.

1. Memaparkan hasil mengidentifikasi identitas lingual *Short Message Service* (SMS) *Sobis* melalui pemanfaatan linguistik korpus menggunakan aplikasi *AntConc*.
2. Menyajikan daftar lingual *Short Message Service* (SMS) *Sobis* yang ditemukan melalui pemanfaatan linguistik korpus menggunakan aplikasi *AntConc*

1.6 Manfaat Penelitian

Secara garis besar, manfaat yang diperoleh dari penelitian ini terbagi menjadi dua, yakni manfaat teoretis dan manfaat praktis.

1. Secara teoretis penelitian ini diharapkan dapat menambah wawasan perkembangan ilmu pengetahuan mengenai bahasa secara komputasional dalam hal ini linguistik korpus dan kajian terapannya kepada para pendidik dan peneliti selanjutnya.
2. Secara praktis penelitian ini diharapkan memberikan informasi kepada masyarakat luas tentang penanda lingual atau identitas lingual *Short Message Service* (SMS) *Sobis* yang mengindikasikan bentuk penipuan.



BAB II TINJAUAN PUSTAKA

2.1 Landasan Teori

Landasan teori untuk penelitian ini mencakup konsep-konsep dan teori-teori yang meliputi, (1) Linguistik Umum, (2) Linguistik Terapan, (3) Linguistik Korpus, (4) Kolokasi, dan (5) Terminologi *Sobis*

2.1.1 Linguistik Umum

Secara umum, banyak orang yang memahami linguistik sebagai ilmu yang mempelajari bahasa, atau dengan kata lain ilmu yang menjadikan bahasa sebagai objek kajiannya sebagaimana yang dipaparkan oleh Chaer (2014). Meski banyak keilmuan lain menggunakan bahasa sebagai objek kajiannya, linguistik tetaplah ilmu yang memperlakukan bahasa sebagai bahasa. Kata linguistik berasal dari kata berbahasa Latin *lingua* yang berarti 'bahasa'. Orang yang ahli dalam ilmu linguistik disebut sebagai linguis.

Ilmu linguistik dikenal juga sebagai linguistik umum (*general linguistics*). Hal ini berarti, ilmu linguistik tidak hanya mengkaji sebuah bahasa saja, melainkan seluk beluk bahasa pada umumnya. Dengan kata lain, bahasa sebagai alat interaksi sosial manusia. Terlepas dari sebanyak apapun perbedaan tiap bahasa di dunia ini, pasti ada persamaannya. Itulah yang disebut universalitas bahasa.

Sebagai alat interaksi sosial manusia, bahasa merupakan sebuah sistem yang bersifat sistematis sekaligus sistemis. Sifat sistematis dari bahasa merujuk pada keteraturan dan keterorganisasian dalam penggunaannya. Ini berarti bahwa bahasa memiliki aturan dan pola yang konsisten dalam pembentukan dan penggunaan kata, frasa, serta kalimat. Sementara itu, sifat sistemis menunjukkan bahwa bahasa bukanlah sebuah sistem tunggal, melainkan terdiri dari beberapa subsistem, yaitu fonologi, morfologi, sintaksis, dan semantik.

Ferdinand de Saussure (1857-1913) dikenal sebagai bapak linguistik modern berdasarkan pandangannya pada buku *Course de Linguistique Generale*. Buku ini terbit dua tahun sepeninggal de Saussure, yaitu pada tahun 1915. Buku ini memuat pandangan-pandangan de Saussure selama memberikan kuliah di Universitas Jenawa tahun 1906. Beberapa konsep yang terdapat dalam buku itu antara lain: (1) telaah sinkronik dan diakronik, (2) perbedaan *langue* dan *parole*, (3) perbedaan signifiant dan signifie, dan (4) hubungan sintagmatik dan paradigmatis.

Dewasa ini, penelitian tentang bahasa telah dilakukan secara masif di berbagai aspek, sehingga menyebabkan perkembangan linguistik yang signifikan kembali bahwa bahasa sangat erat kaitannya dengan kegiatan maka cabang ilmu ini pun menjadi sangat luas. Perkembangan signifikan ini sejalan dengan konsep filsafat ilmu itu sendiri. Kelahiran iralisme adalah bentuk ketidakpuasan atas linguistik tradisional



Perubahan kembali terjadi dengan lahirnya linguistik transformasional yang membawa pendekatan baru dan berbeda dari struktural sebelumnya. Terbitnya buku Noam Chomsky berjudul *Syntactic Structure* pada tahun 1957 menjadi awal dari era transformasi dilanjutkan dengan buku keduanya yang berjudul *Aspect of Theory of Syntax* pada tahun 1965. Chomsky kemudian dikenal dengan model tata bahasa generatifnya, yaitu *Transformational Generative Grammar*.

2.1.1.1 Kata

Dalam lingkup linguistik, kata didefinisikan sebagai unit terkecil yang memiliki makna atau fungsi dalam suatu bahasa. Chaer (2014, hlm. 219) menjelaskan bahwa kata adalah satuan terbesar (satuan terkecilnya morfem) dalam tataran morfologi, sedangkan dalam tataran sintaksis kata merupakan satuan terkecil, yaitu dalam hubungannya dengan unsur-unsur pembentuk satuan sintaksis yang lebih besar, yaitu frasa, klausa, dan kalimat. Dengan demikian, kata menjadi satuan leksikal yang dapat berdiri sendiri sebagai ungkapan lengkap, meskipun bergabung dengan unit lain untuk membentuk struktur yang lebih besar.

Kata sebagai pengisi satuan sintaksis dibedakan menjadi dua, yaitu kata penuh (*fullword*) dan kata tugas (*functionword*). Kata penuh adalah kata yang secara leksikal memiliki makna, mempunyai kemungkinan untuk mengalami proses morfologi, merupakan kelas terbuka dan dapat berdiri sebagai sebuah satuan turunan. Sedangkan yang disebut tugas adalah kata yang secara leksikal tidak mempunyai makna, tidak mengalami proses morfologi, merupakan kelas tertutup, dan dalam pertuturan dia tidak dapat berdiri sendiri. Adapun kata penuh tersendiri diantaranya termasuk kata nomina, verba, adjektiva, adverbialia, dan numeralia. Sedangkan yang termasuk kata tugas, yaitu kata-kata yang termasuk preposisi dan konjungsi.

Secara morfologi sendiri, kata adalah satuan gramatikal bebas dan terkecil. Maksudnya bahwa suatu kata tidak dapat disegmentasikan lagi menjadi yang lebih kecil tanpa merusak makna. Adapun dengan bebas berarti satuan tersebut berdiri di dalam kalimat atau penuturan. Kata-kata yang termasuk kata penuh memang mempunyai kecenderungan yang mutlak, atau hampir mutlak, sehingga dapat menjadi pengisi fungsi-fungsi sintaksis. Sedangkan kata tugas mempunyai kebebasan yang terbatas. Hal ini sesuai dengan namanya, yaitu kata tugas, di selalu terikan dengan yang ada di belakangnya (preposisi) dengan yang ada di depannya (posposisi), serta dengan yang dirangkaikannya (konsungsi).



merupakan kombinasi dua atau lebih kata yang membentuk unit frasa tidak memiliki predikat utama atau non-predikatif. Menurut n. 222), istilah frasa sendiri memiliki banyak pengertian. Frasa akan sebagai satuan sintaksis yang satu tingkat berada di atas

masalah yang berkaitan dengan bahasa dalam kehidupan sehari-hari. Telah dikatakan di awal bahwa objek linguistik adalah bahasa dan linguistik terapan merupakan salah satu subdisiplin ilmu linguistik. Oleh karena itu, menurut Pateda (1991, hlm. 27) dengan sendirinya objek linguistik terapan adalah bahasa. Namun, sasarannya bukanlah bahasa itu sendiri, melainkan bahasa dalam kaitannya dengan kebutuhan praktis.

Berbeda dengan linguistik teoretis yang lebih menekankan pada pemahaman dan analisis struktural bahasa, Sebagaimana yang turut dituliskan oleh Romadhan dan kawan-kawan (2023, hlm 12) bahwa linguistik teoritis berkaitan dengan cabang ilmu linguistik seperti fonologi, morfologi, sintaksis, dan lain-lain, sedangkan linguistik terapan mengaplikasikan pengetahuan linguistik ke dalam konteks nyata seperti pendidikan, penerjemahan, pengajaran, perencanaan, pengembangan, dan pemeliharaan bahasa.

Contoh bidang yang termasuk dalam linguistik terapan adalah pengajaran bahasa asing (seperti Teaching English as a Second Language atau TESOL), analisis wacana dalam konteks sosial atau politik, dan pengembangan materi pembelajaran bahasa. Linguistik terapan juga mencakup penelitian tentang akuisisi bahasa kedua, penerjemahan, interpretasi, serta isu-isu kebijakan bahasa. Linguistik terapan bahkan telah digunakan dalam penyusunan beberapa program komputer dan digunakan dalam penerjemahan dengan bantuan komputer. Berdasarkan hal tersebut, maka dapat disimpulkan bahwa linguistik terapan berperan dalam menjembatani antara teori linguistik dan kebutuhan praktis masyarakat dalam berbagai konteks berbahasa.

Linguistik terapan menjadi bidang yang berkaitan erat dengan masalah dan isu sosial nyata, di mana bahasa memainkan peran sentral. Oleh sebab itu, bidang ini memanfaatkan berbagai model dan kerangka kerja linguistik. Seperti yang dijelaskan oleh Llinares dan Tom (2017, hlm. 27) bahwa ahli linguistik terapan harus memiliki pengetahuan yang luas dan serba bisa sebab mereka perlu menggunakan berbagai teori linguistik yang berbeda. Keanekaragaman bidang ini terlihat dari banyaknya subbidang, subspecialisasi, dan bidang terkait yang ada. Bukti dari keragaman ini dapat dilihat dalam berbagai buku pegangan dan referensi ensiklopedia yang telah diterbitkan dalam sepuluh tahun terakhir. Indikasi lain dari keberagaman ini telah dipaparkan oleh Peniro dan Jorde (2019, hlm. 1-13) bahwa hingga tahun 2005 Asosiasi Internasional Linguistik Terapan (AILA) mempunyai 24 'Komisi Ilmiah' berbeda yang mewakili berbagai bidang Linguistik Terapan.



Linguistik Korpus

Peniro dan Jorde (2019, hlm. 5-6) linguistik korpus adalah bidang yang menggabungkan elemen linguistik dengan ilmu komputer untuk menganalisis dan memproses bahasa manusia secara komputasional. Linguistik korpus semakin populer saat ini di kalangan akademisi

dan praktisi di berbagai bidang ilmu. Tidak hanya pada disiplin ilmu komputer dan sains, tetapi juga ilmu sosial seperti bahasa. Peniro dan Jorde juga menganggap linguistik korpus telah muncul sebagai bidang penting dalam linguistik terapan selama dekade terakhir. Tren linguistik terapan mencangkup pengakuan bahwa linguistik korpus telah memberikan kontribusi yang lebih luas dalam mengatasi permasalahan bahasa di dunia nyata.

Penyelesaian masalah kebahasaan melalui linguistik korpus melibatkan pemilihan dan penggunaan algoritma, representasi data, penggunaan abstraksi, hingga pengujian hipotesis, sebagaimana yang diistilahkan oleh Wing sebagai “*Computational Thinking*” pada tahun 2006. Hal ini menjadi semakin penting mengingat pada era digital sekarang ini, teknologi dan komputer telah menjadi bagian integral kehidupan manusia, Christi dan Widyawanti (2023, hlm. 12594). Adapun tujuan penerapan linguistik korpus adalah untuk menerapkan teknologi untuk mengkaji pola penggunaan bahasa dengan mempelajari kumpulan besar teks yang ada secara alami dalam penggunaan bahasa yang sebenarnya. Dalam hal ini, linguistik korpus dapat digunakan sebagai metode untuk memberikan bukti secara sistematis penggunaan verba pelaporan dalam konteks tertentu (Switchan dan Nathaya 2020, hlm. 4).

Dalam konteks linguistik korpus, pemanfaatannya menjadi sebuah metode daripada aspek bahasa yang perlu penjelasan atau pendeskripsian. Dengan demikian, linguistik korpus dianggap sebagai disiplin ilmu empiris, sebab bertujuan untuk menarik simpulan berdasarkan analisis data yang ada. Sebagai metode, penelitian linguistik korpus menggunakan data bahasa yang terkumpul dari suatu sumber yang disebut korpus (jamak: korpora). Istilah ini berasal dari bahasa Latin yang berarti tubuh (Zufferey 2020, hlm. 14). Adapun pengolahan korpus seringkali didasarkan pada sejumlah besar data baik lisan maupun tulisan (MCEnergy & Hardie 2012, hlm. 45). Analisis korpus dapat mencangkup berbagai aspek bahasa seperti kata, frasa, kalimat, maupun wacana. Melalui analisis korpus yang dibangun, peneliti dapat melihat frekuensi kemunculan suatu kata atau frasa dengan kata lainnya, atau bagaimana suasana semantis suatu kata ditampilkan melalui pencarian kolokasinya (Ermanto dan kawan-kawan 2022, hlm. 16).

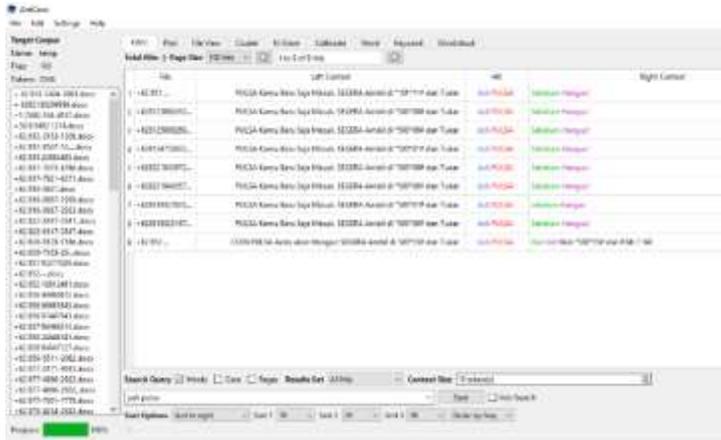
2.1.3.1 *AntConc*

AntConc adalah perangkat lunak lintas *platform* gratis yang digunakan untuk melakukan penelitian korpus, memperkenalkan metode korpus, dan melakukan pembelajaran bahasa berbasis data. *AntConc* diciptakan oleh Laurence University of Waseda, Jepang. Alat ini dikembangkan dengan *Python* menggunakan kompiler *Pyinstaller* untuk menghasilkan *executable* untuk berbagai *platform*. *AntConc* menggunakan SQLite sebagai databasenya.



Menurut Anthony (2023), terdapat sembilan *tools* atau alat yang dapat diakses di *AntConc*. Caranya dengan mengklik 'tab' di jendela alat, menggunakan CTRL+TAB untuk beralih antar alat, atau menggunakan kombinasi tombol CTRL + Nomor Alat (misalnya, CTRL +1 untuk KWIC, CTRL +2 untuk Plot) untuk memilih alat tertentu. Berikut ini sembilan alat tersebut dipaparkan:

1) KWIC



Gambar 2. 1 Alat KWIC pada AntCont

Alat ini menampilkan hasil pencarian dalam format konkordansi atau *Key Word In Context* (KWIC). Hal ini memungkinkan untuk melihat bagaimana kata dan frasa biasa digunakan dalam kumpulan teks.

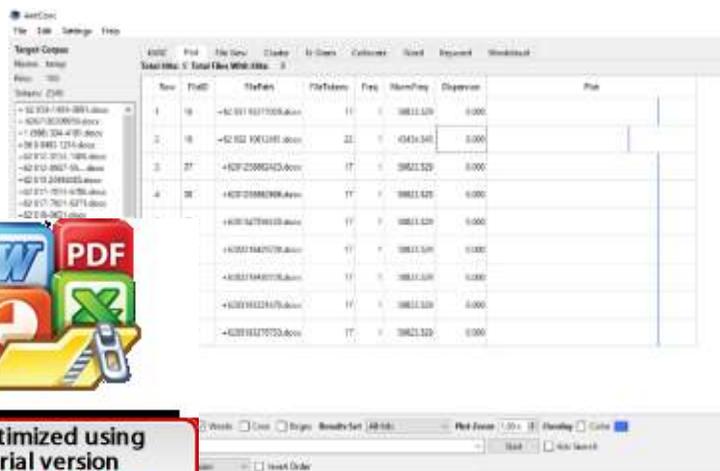
Berikut ini langkah-langkah untuk menghasilkan serangkaian baris konkordansi dari korpus dan fitur utama pada alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File atau buat korpus cepat dengan memilih "Open File(s) as Quick Corpus" dari menu file. File yang terdapat dalam korpus ditampilkan di bingkai kiri jendela utama di bawah "Target Corpus".
 2. Masukkan permintaan pencarian di kotak pencarian. Lihat bagian 'PILIHAN PENCARIAN' dalam dokumen ini penjelasan tentang pilihan istilah pencarian "Kata", "Kapital", dan "Regex".
 3. Pilih ukuran kumpulan hasil yang akan disajikan menggunakan widget kotak kombo "Kumpulan Hasil".
 4. Pilih jumlah kata yang akan ditampilkan di kedua sisi istilah pencarian an "Ukuran Konteks" widget kotak putar.
- 1) "Mulai" untuk memulai pencarian dan tunggu hasilnya



6. Gunakan "Opsi Sortir" untuk mengatur ulang baris konkordansi berdasarkan ID baris, nama file, atau posisi file.kata. Widget pertama memungkinkan Anda untuk mengurutkan baris konkordansi dengan cepat menurut kata-kata di sebelah kanan atau kiri kata tengah, atau memilih tanpa pengurutan, atau menggunakan pengurutan khusus. Tiga widget berikutnya memungkinkan Anda untuk memilih parameter pengurutan: 1L, 2L... adalah kata-kata di sebelah kiri kata target, 'C' adalah kata tengah, dan 1R, 2R... adalah kata-kata di sebelah kanan kata tengah. Widget terakhir memungkinkan Anda untuk mengurutkan hasil menurut frekuensi pola yang ditentukan oleh parameter pengurutan (opsi "Urutkan menurut frekuensi") atau menurut abjad (opsi "Urutkan menurut nilai"). Opsi default "Urutkan menurut frekuensi" sangat disarankan karena akan memungkinkan Anda untuk dengan mudah mengidentifikasi pola yang paling sering muncul dalam korpus target. Setelah menyesuaikan opsi pengurutan, klik tombol "Mulai" untuk membuat ulang baris konkordansi.
7. Jumlah total baris konkordansi yang dihasilkan (Total Hits) ditampilkan di bagian atas jendela alat. Jika tidak ada hasil yang ditemukan, peringatan akan ditampilkan di layar.
8. Mengklik dua kali pada sel mana saja di jendela hasil akan menyebabkan perangkat lunak melompat ke alat File (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat hasil tersebut persis seperti yang muncul dalam konteks di file asli.
9. Jika Anda ingin memfilter hasilnya, pilih baris yang diinginkan, lalu tekan tombol "Hapus" untuk menghapus baris yang dipilih atau tekan "SHIFT+Hapus" untuk membiarkan baris yang dipilih dan menghapus semua baris lainnya.
10. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini

2) Plot



Row	FileID	FileName	FileFormat	Page	WordFreq	Dispersion	Plot
1	18	-623319217008.docx	11	1	38823.529	0.000	
2	18	-6233219612881.docx	22	1	43424.348	0.000	
3	27	+62725992423.docx	17	1	38823.529	0.000	
4	28	+62725992423.docx	17	1	38823.529	0.000	

Optimized using trial version
www.balesio.com

Plot pada AntCont

Alat ini menunjukkan hasil pencarian konkordansi yang diplot dalam format 'barcode', dengan panjang teks dinormalisasi dengan lebar bar dan setiap pukulan ditampilkan sebagai garis vertikal di dalam bilah. Hal ini memungkinkan untuk melihat posisi kemunculan hasil pencarian di masing-masing teks korpus. Contoh penggunaan Plot Tool adalah untuk menentukan di mana kata-kata konten tertentu muncul dalam makalah teknis, atau di mana aktor atau karakter cerita muncul melalui drama atau novel.

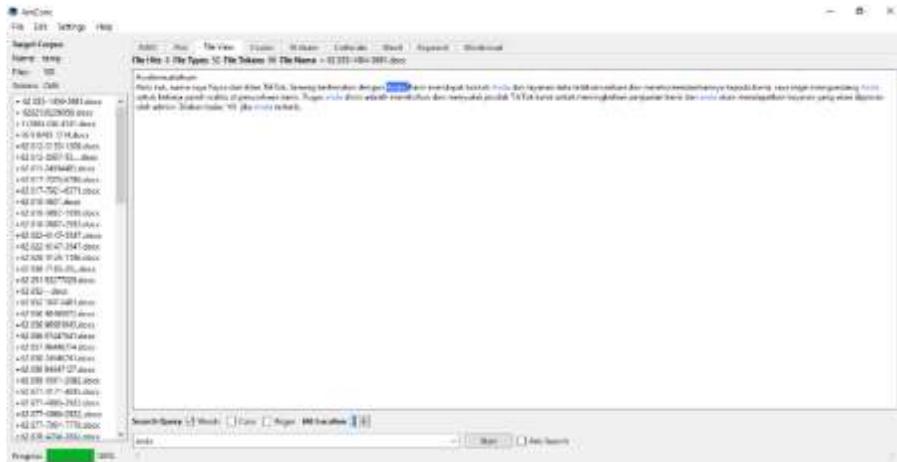
Berikut ini langkah-langkah menghasilkan serangkaian hasil plot dari korpus dan menunjukkan fitur utama alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File. Atau, buat korpus cepat dengan memilih "Open File(s) as Quick Corpus" dari menu file. File yang terdapat dalam korpus ditampilkan di bingkai kiri jendela utama di bawah "Target Corpus".
2. Masukkan permintaan pencarian di kotak pencarian. Lihat bagian 'PILIHAN PENCARIAN' dalam dokumen ini untuk penjelasan tentang pilihan istilah pencarian "Kata", "Kapital", dan "Regex".
3. Pilih ukuran kumpulan hasil yang akan disajikan menggunakan widget kotak kombo "Kumpulan Hasil".
4. Gunakan widget "Plot Zoom" untuk mengontrol ukuran plot dan tingkat detail yang akan ditampilkan.
5. Klik tombol "Mulai" untuk memulai pencarian dan tunggu hasilnya ditampilkan.
6. Gunakan opsi "Urutkan berdasarkan" untuk mengatur ulang plot menurut berbagai parameter yang ditampilkan.
7. Jumlah total hit dan jumlah total plot ditampilkan di bagian atas jendela alat. Bila tidak ada ditemukan, peringatan akan ditampilkan di layar.
8. Mengklik dua kali pada sel mana saja di jendela hasil akan menyebabkan perangkat lunak melompat ke alat File (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat hasil tersebut persis seperti yang muncul dalam konteks di file asli.
9. Dengan mencentang opsi "Overlay" dan memilih warna yang sesuai (dengan mengklik kotak warna), hasil yang ada dapat dilapis dengan hasil baru untuk pencarian yang berbeda. Ini memungkinkan Anda melihat bagaimana kueri pencarian yang berbeda saling terkait dan/atau tumpang tindih.

lanjutan tersedia dengan alat ini. Beberapa preferensi menu ia dengan alat ini alat.



3) File View



Gambar 2. 3 Alat View pada AntConc

Alat ini menunjukkan isi teks individual. Hal ini memungkinkan untuk menyelidiki secara lebih rinci hasil yang dihasilkan di alat *AntConc* lainnya.

Berikut ini langkah-langkah menghasilkan tampilan berkas asli dan menunjukkan fitur utama alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File. Atau, buat korpus cepat dengan memilih "Open File(s) as Quick Corpus" dari menu file. File yang terdapat dalam korpus ditampilkan di bingkai kiri jendela utama di bawah "Target Corpus".
2. Klik dua kali berkas dalam daftar "Target Corpus" di sebelah kiri jendela utama untuk melihat isinya. Atau, pilih file dalam daftar "Target Corpus" dan klik "Start" di antarmuka alat. Alat File akan dipilih secara otomatis, dan isi file akan ditampilkan.
3. Untuk menyorot hasil kueri penelusuran di layar, masukkan kueri penelusuran dan klik "Mulai". Lihat bagian terkait dalam dokumen ini yang menjelaskan opsi istilah penelusuran "Kata", "Kapitalisasi", dan "Regex". Kata dalam berkas yang cocok dengan kueri akan disorot secara otomatis.
4. Gunakan widget "Lokasi Hit" untuk melompat ke hit yang berbeda dalam file. Atau, gunakan keyboard pintasan untuk sistem operasi Anda (lihat bagian PINTASAN).
5. Beberapa lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga dapat diakses dengan alat ini.



Kanan") atau dapat dimulai atau diakhiri dengan istilah kueri pencarian ("Di Sebelah Kiri/Kanan").

7. Jumlah total tipe klaster ("Tipe Klaster") dan jumlah total gabungan semua token klaster ("Token Klaster") ditampilkan di bagian atas jendela alat. Jika tidak ada hasil yang ditemukan, peringatan akan ditampilkan di layar.
8. Klik dua kali pada sel mana saja di jendela hasil untuk menyebabkan perangkat lunak melompat ke alat KWIC (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat garis konkordansi untuk klaster tersebut di seluruh korpus.
9. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini.

5) N-Gram

Type	Rank	Freq	Range
isi ty	1	25	25
oort auka	2	11	11
u info	2	11	11
anda di	4	9	9
ata hater	4	9	9
di dan	4	9	9
jak polri	4	9	9
ta anda	4	9	9
ngara anda	4	9	9
akanat narasi	6	9	9
laku jak	6	9	9
nya raja	12	9	9
laku klu	12	9	9
ji di	12	9	9
laku klu	12	9	9
masak ngara	12	9	9
manan-man	12	9	9
polri kama	12	9	9

Gambar 2. 5 Alat N-Gram pada AntCont

Alat ini memindai seluruh korpus untuk semua cluster berukuran 'N' (misalnya, cluster 2 kata, cluster 3 kata,...). Hal ini memungkinkan untuk menemukan ekspresi umum dalam korpus.

Berikut langkah-langkah menghasilkan sekumpulan N-Gram dan menunjukkan fitur utama alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File.

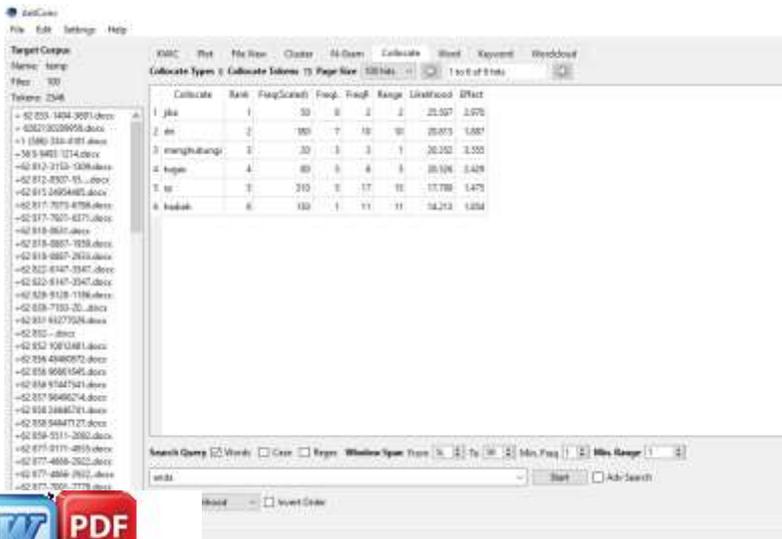


korpus cepat dengan memilih "Open File(s) as Quick Corpus" le. File yang terdapat dalam korpus ditampilkan di bingkai kiri ia di bawah "Target Corpus".

ai parameter untuk memfilter jumlah n-gram yang akan ukuran n-gram (jumlah kata), slot terbuka (jumlah slot dalam

- n-gram yang dapat mengambil beberapa nilai), frekuensi n-gram minimum, dan rentang n-gram minimum (jumlah file).
3. Klik tombol "Mulai" untuk memulai pencarian dan tunggu hingga hasilnya ditampilkan. Jika kueri pencarian dimasukkan, hanya ngram yang cocok dengan kueri tersebut yang akan ditampilkan,
 4. Gunakan opsi "Urutkan berdasarkan" untuk mengatur ulang urutan hasil.
 5. Jumlah total tipe n-gram ("tipe N-Gram") dan jumlah total gabungan semua token n-gram ("Token N-Gram") ditampilkan di bagian atas jendela alat. Jika tidak ada hasil yang ditemukan, peringatan akan ditampilkan di layar.
 6. Klik dua kali pada sel mana saja di jendela hasil untuk menyebabkan perangkat lunak melompat ke alat KWIC (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat garis konkordansi untuk n-gram tersebut di seluruh korpus.
 7. Untuk entri yang berisi slot terbuka, Shift + Klik dua kali pada entri "Jenis" di jendela hasil untuk menampilkan varian yang dapat masuk ke slot terbuka melalui "Penampil Slot Terbuka" dan dua statistik terkait yang menampilkan tingkat variasi untuk slot tersebut. Nilai *_TT adalah rasio jenis/token untuk slot tersebut, dan nilai *_ent adalah nilai Entropi untuk slot tersebut.
 8. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini.

6) Collocate



Collocate	Rank	Freq(Search)	Freq. Range	Likelihood	Effect
1 jika	1	50	8 2 2	25.027	2.076
2 dan	2	100	7 18 30	25.013	1.087
3 menghambat	3	33	3 3 1	20.102	3.100
4 begin	4	80	5 4 5	18.104	3.429
5 itu	5	210	3 17 11	17.788	1.475
6 tidak	6	130	1 11 11	14.213	1.054

locate pada AntConc



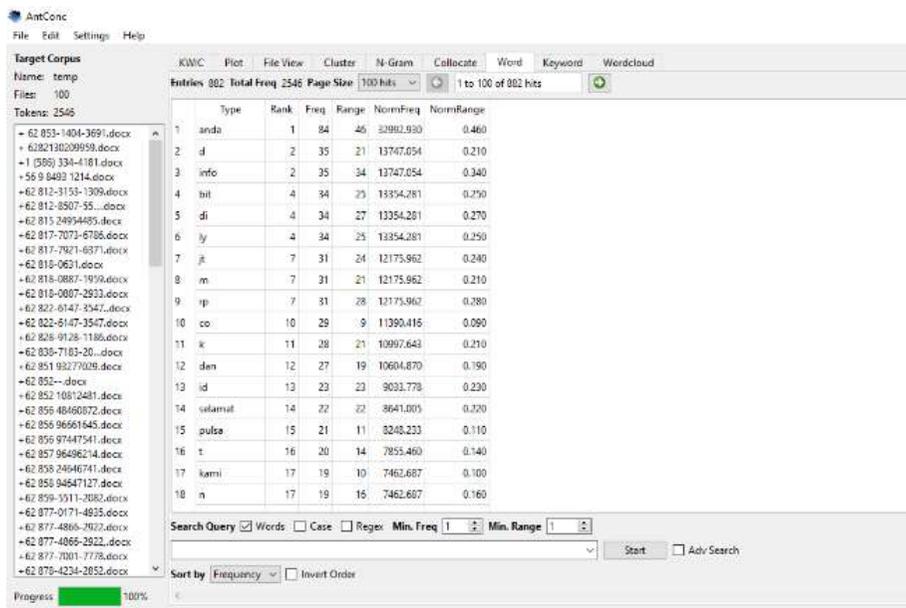
Alat ini menampilkan kata-kata yang sering muncul dalam jarak tertentu dari istilah pencarian (yaitu, berkolokasi). Hal ini memungkinkan untuk menemukan kata mana yang muncul bersamaan dengan kata lain dalam korpus.

Berikut ini langkah-langkah untuk menghasilkan serangkaian kolokasi dan menunjukkan fitur utama alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File. Atau, buat korpus cepat dengan memilih "Open File(s) as Quick Corpus" dari menu file. File yang terdapat dalam korpus ditampilkan di bingkai kiri jendela utama di bawah "Target Corpus".
2. Pilih berbagai parameter untuk memfilter jenis kolokasi yang akan ditampilkan: rentang jendela (mungkin posisi di sebelah kiri dan kanan istilah pencarian, tempat kolokasi dapat muncul), frekuensi kolokasi minimum, dan rentang kolokasi minimum (jumlah file).
3. Masukkan permintaan pencarian di kotak pencarian. Lihat bagian 'PILIHAN PENCARIAN' dalam dokumen ini untuk penjelasan tentang pilihan istilah pencarian "Kata", "Kapital", dan "Regex".
4. Klik tombol "Mulai" untuk memulai pencarian dan tunggu hasilnya ditampilkan.
5. Gunakan opsi "Urutkan berdasarkan" untuk mengatur ulang urutan hasil.
6. Jumlah total tipe kolokasi ("Jenis Kolokasi") dan jumlah total gabungan dari semua tipe kolokasi token ("Token Collocate") ditampilkan di bagian atas jendela alat. Bila tidak ada hasil yang ditemukan, peringatan akan ditampilkan di layar.
7. Klik dua kali pada sel mana saja di jendela hasil untuk menyebabkan perangkat lunak melompat ke alat KWIC (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat garis konkordansi untuk kolokasi tersebut di seluruh korpus.
8. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini.



7) Word



Gambar 2. 7 Alat Word pada AntCont

Alat ini menghitung semua kata dalam korpus dan menyajikannya dalam daftar yang terurut. Ini memungkinkan untuk menemukan kata mana yang paling sering muncul dalam korpus

Berikut ini langkah-langkah untuk menghasilkan daftar kata dan menunjukkan fitur utama alat ini.

1. Pilih korpus menggunakan "Corpus Manager" yang tersedia dari menu File. Atau, buat korpus cepat dengan memilih "Open File(s) as Quick Corpus" dari menu file. File yang terdapat dalam korpus ditampilkan di bingkai kiri jendela utama di bawah "Target Corpus".
2. Klik tombol "Mulai" untuk memulai pemrosesan dan tunggu hingga hasilnya ditampilkan. Jika kueri pencarian dimasukkan, hanya kata-kata yang cocok dengan kueri tersebut yang akan ditampilkan. Lihat bagian 'PILIHAN PENCARIAN' dalam dokumen ini untuk penjelasan tentang opsi istilah pencarian "Kata", "Kapital", dan "Regex".
3. Gunakan opsi "Urutkan berdasarkan" untuk mengatur ulang urutan hasil.
4. Jumlah total jenis kata ("Jenis kata") dan jumlah total gabungan semua

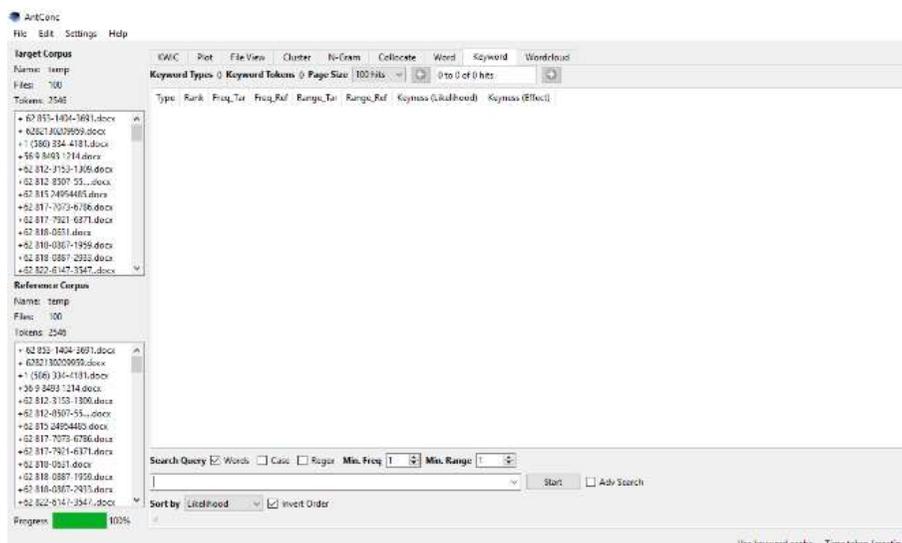


"Jenis kata") Token") ditampilkan di bagian atas jendela alat. la hasil yang ditemukan, peringatan akan ditampilkan di layar. li pada sel mana pun di jendela hasil untuk menyebabkan nak melompat ke alat KWIC (lihat bagian yang relevan dari

dokumen ini) di mana Anda dapat melihat garis konkordansi untuk kata itu di seluruh korpus.

6. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini

8) Keyword



Gambar 2. 8 Alat Keyword pada AntCont

Alat ini menunjukkan kata-kata yang sering muncul dalam korpus target dibandingkan dengan kata-kata dalam korpus referensi berdasarkan ukuran statistik (misalnya, 'kata kunci'). Kata-kata ini dapat dianggap sebagai ciri korpus sasaran. Pengaturan juga dapat diubah untuk menampilkan kata-kata yang jarang muncul dalam korpus target dibandingkan dengan korpus referensi (misalnya, 'kata kunci negatif').

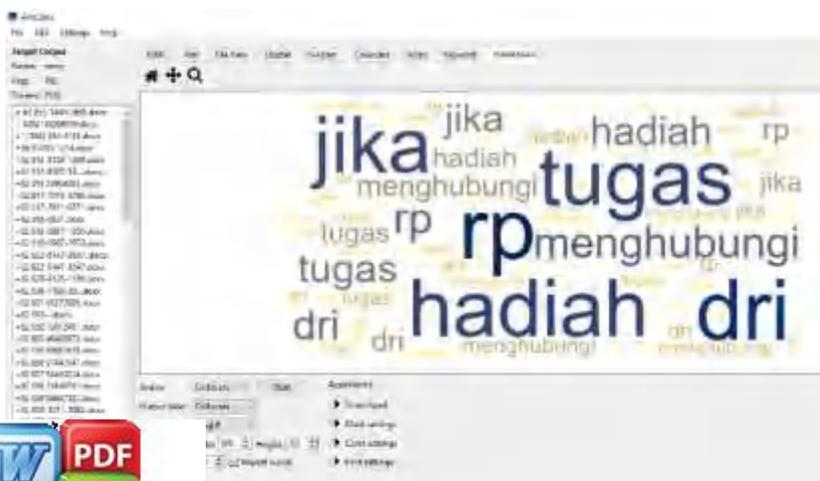
Berikut ini langkah-langkah untuk menghasilkan daftar kata kunci dan menunjukkan fitur utama alat ini.

1. Buat korpus cepat dengan memilih opsi "Buka File sebagai Korpus Cepat" dari menu file. Atau, pilih opsi "Corpus Manager" dari menu file dan pastikan opsi "Target Corpus" dipilih. Kemudian, pilih salah satu korpus yang tersedia atau buat korpus Anda sendiri dari file mentah atau daftar petunjuk di bawah bagian Corpus Manager pada halaman untuk mengetahui cara melakukannya). Korpus ini kemudian akan ditampilkan di bingkai kiri atas jendela utama sebagai korpus target untuk analisis Anda. File yang akan ditampilkan di bingkai kiri atas jendela utama sebagai korpus Target".



2. Pilih korpus referensi dengan membuka opsi "Corpus Manager" dari menu file dan memeriksa opsi "Opsi Referensi". Selanjutnya, seperti pada langkah 1, pilih salah satu korpus yang tersedia atau buat korpus Anda sendiri dari file mentah atau daftar kata. File yang terdapat dalam korpus akan ditampilkan di bingkai kiri bawah jendela utama di bawah "Korpus Referensi".
3. Klik tombol "Mulai" untuk memulai pemrosesan dan tunggu hingga hasilnya ditampilkan. Jika kueri pencarian dimasukkan, hanya kata-kata yang cocok dengan kueri tersebut yang akan ditampilkan. Lihat bagian 'PILIHAN PENCARIAN' dalam dokumen ini untuk penjelasan tentang opsi istilah pencarian "Kata", "Kapital", dan "Regex".
4. Gunakan opsi "Urutkan berdasarkan" untuk mengatur ulang urutan hasil.
5. Jumlah total jenis kata kunci ("Jenis kata kunci") dan jumlah total gabungan semua token kata ("Token Kata Kunci") ditampilkan di bagian atas jendela alat. Bila tidak ada hasil yang ditemukan, peringatan akan ditampilkan di layar.
6. Klik dua kali pada sel mana saja di jendela hasil untuk menyebabkan perangkat lunak melompat ke alat KWIC (lihat bagian yang relevan dari dokumen ini) di mana Anda dapat melihat baris konkordansi untuk kata kunci tersebut di seluruh korpus.
7. Pencarian lanjutan tersedia dengan alat ini. Beberapa preferensi menu juga tersedia dengan alat ini

9) Wordcloud



Wordcloud pada AntCont

Alat ini memvisualisasikan hasil yang dihasilkan oleh alat KWIC, File, Cluster, N-Gram, Collocation, Word, dan Kata Kunci serta "Scratchpad" teks biasa dalam bentuk 'awan kata'. Wordclouds seringkali digunakan sebagai visualisasi yang estetik, di mana kata-kata diletakkan di area tampilan atau topeng gambar 'bertema' dan diberi ukuran sesuai dengan properti (misalnya, frekuensi kata). Kehati-hatian harus diberikan saat menggunakan awan kata untuk analisis linguistik, karena visualisasinya memerlukan distorsi ukuran kata agar sesuai dengan area tampilan.

Berikut ini langkah-langkah untuk menghasilkan awan kata atau *Wordcloud* dan menunjukkan fitur utama alat ini.

1. Pilih "Sumber" untuk word cloud. Ini bisa berupa "Scratchpad" dari teks biasa (dapat diakses sebagai opsi pertama dalam daftar "Appearance", atau output dari KWIC, File, Cluster, N-Gram, Collocate, Word, dan Alat kata kunci).
2. Pilih properti dari sumber untuk ditampilkan sebagai "label" dan "nilai" di word cloud. Untuk sumber KWIC, File, dan Scratchpad, properti "Type" (Word) dan "Freq" (Frekuensi) dipilih secara otomatis.
3. Pilih ukuran gambar. Ini akan menentukan berapa banyak kata yang dapat ditempatkan dalam gambar. Perhatikan bahwa gambar akan secara otomatis diskalakan ke ukuran tampilan jendela.
4. Pilih jumlah kata maksimum yang akan ditampilkan di awan kata. Bergantung pada jumlah item dari sumber dan pengaturan lainnya (misalnya, ukuran font minimum), nilai ini mungkin tidak tercapai.
5. Centang kotak "Gunakan kata henti" untuk menghapus kata henti dari kumpulan kata.
6. Centang kotak "Ulangi kata" untuk mengisi ruang yang tersisa sebanyak mungkin di gambar awan kata dengan kata-kata yang ada. Bergantung pada pengaturan lain (misalnya, ukuran font minimum), tidak semua ruang akan digunakan.
7. Klik "Mulai" untuk membuat awan kata.

2.1.3.2 *Natural Language Processing (NLP)*

Dalam linguistik korpus, Natural Language Processing (NLP) merujuk pada penerapan teknologi dan metode komputasi untuk menganalisis, memproses, dan memahami korpus. Artinya, NLP membantu untuk mengotomatiskan analisis dan pemrosesan korpus tersebut. Mengutip Dunn (2022, hlm. 4) "*Corpus analysis can be scaled up by incorporating computational methods from natural language processing,*" bahwa analisis korpus dapat diperluas dan ditingkatkan dengan bingkai metode komputasional dari pemrosesan bahasa alami ini memungkinkan analisis otomatis berbagai macam fenomena skala yang besar dan generalisasi yang lebih luas.



Menurut Kyle (2021, hlm 1-16), bidang NLP cukup luas untuk mencakup berbagai proses. Dalam penelitian ini, NLP berhubungan langsung dengan kajian linguistik korpus, sehingga fokusnya adalah analisis korpus secara otomatis, analisis frekuensi dan distribusi kata, menemukan pola linguistik seperti N-Gram (kombinasi kata yang muncul berurutan), KWIC (kata kunci dalam konteks teks), serta kolokasi (kata-kata yang muncul secara bersamaan). Adapun salah satu aspek terpenting NLP dalam penelitian ini adalah ekspresi umum penutur. Ekspresi umum adalah frasa atau ungkapan yang sering digunakan dalam kehidupan sehari-hari dan memiliki makna yang dikenal luas oleh penutur bahasa tersebut. Dalam konteks penelitian ini, penting untuk mengidentifikasi frasa yang digunakan oleh pelaku penipuan (*Sobis*). Berikut ini beberapa jenis frasa yang sering muncul dalam SMS penipuan:

1) Frasa yang memicu rasa panik atau mendesak

Contoh:

- (1) "Segera hubungi..."
- (2) "Jangan lewatkan kesempatan ini..."
- (3) "Segera lakukan pembayaran..."

2) Frasa yang menawarkan hadiah atau keuntungan besar

Contoh:

- (1) "Selamat! Anda memenangkan..."
- (2) "Dapatkan bonus besar..."
- (3) "Anda terpilih sebagai pemenang..."

3) Frasa yang meminta informasi pribadi atau keuangan

Contoh:

- (1) "Kirimkan nomor rekening Anda..."
- (2) "Mohon verifikasi data diri Anda..."
- (3) "Masukkan kode OTP berikut..."

4) Frasa yang mengandung kesalahan ejaan atau tata bahasa

Contoh:

- (1) "Kami adlah Bank nasional..."
- (2) "Segra balas pesan ini..."
- (3) "Anda memenangkan hadia luar biasa..."

5) Frasa yang menyertakan *link* atau tautan mencurigakan

Contoh:

- (1) "Klik *link* berikut untuk verifikasi..."
- (2) "Klik *link* ini untuk klaim hadiah Anda..."



.ly/penipuan..."

nyebutkan nomor telepon untuk dihubungi

- (1) "Hubungi kami di nomor berikut..."
- (2) "Untuk informasi lebih lanjut, silakan telepon..."
- (3) "SMS kami ke nomor ini..."

7) Frasa yang mengandung ancaman atau intimidasi

Contoh:

- (1) "Jika Anda tidak merespons, akun Anda akan diblokir..."
- (2) "Anda akan dikenakan denda jika tidak segera membayar..."
- (3) "Kami akan mengambil tindakan hukum..."

2.1.4 Kolokasi

Dalam morfologi bahasa Inggris, prefiks adalah huruf atau kelompok huruf yang dilekatkan pada awal kata yang sebagian menunjukkan maknanya (Nordquist 2015, hlm. 1). Seperti prefiks 'co-' dalam kata *collocation* (kolokasi dalam bahasa Indonesia). prefiks 'co-' berasal dari bahasa Latin yang berarti 'bersama' atau 'dengan' (*with, together*). Awalan ini mengacu pada gagasan bahwa kata-kata tersebut berpasangan atau muncul bersama-sama dalam penggunaan bahasa alami (Membean 2024, hlm. 1). Kata-kata seperti *cooperate* (bekerja bersama), *coordinate* (mengatur bersama), dan *coexist* (hidup bersama) menggunakan awalan ini untuk menunjukkan tindakan atau keadaan yang melibatkan lebih dari satu pihak bekerja bersama atau saling terhubung.

Menurut Chaer (2002, hlm. 113), kolokasi (berasal dari bahasa Latin *colloco*) yang berarti 'di tempat yang sama dengan' menunjuk kepada hubungan sintagmatik yang terjadi antarkata atau unsur leksikal. Misalnya, pada kalimat *tiang layar perahu nelayan itu patah dihantam badai, lalu perahu itu digulung ombak, dan tenggelam beserta isinya*. Kita dapat kata-kata *layar, perahu, nelayan, badai, ombak, dan tenggelam* yang merupakan kata-kata dalam satu kolokasi. Sebagaimana yang telah dikemukakan sebelumnya, bahwa kata-kata yang berkolokasi dapat ditemukan bersama atau berada bersama dalam satu tempat atau satu lingkungan.

Istilah kolokasi yang disebutkan oleh Chaer (2002, hlm. 112) merujuk kepada pembicaraan tentang penggolongan medan makna, yang mana salah satunya membicarakan tentang istilah 'set'. Begitu pun yang disampaikan oleh Pateda (2001, hlm. 257) tentang medan makna bahwa medan makna merupakan kelompok kata yang maknanya saling terjalin. Untuk kolokasi sendiri, makna kata tertentu berkenaan dengan keterikatan kata tersebut dengan kata lain yang merupakan kolokasinya. Misalnya, kata *tampan, cantik, dan indah*. Ketiga kata tersebut sama-sama bermakna denotatif 'bagus'. Namun, kata *tampan* memiliki makna [+laki-laki], sedangkan kata *cantik* memiliki komponen atau ciri makna [+wanita]; dan kata *indah* memiliki komponen atau ciri makna [-manusia]. Selain itu, ada bentuk kata-kata *pemuda tampan, gadis cantik, dan*



pemandangan indah, sedangkan bentuk **pemudah indah*, **gadis tampan*, dan **pemandangan cantik* tidak dapat diterima.

Jika kolokasi merujuk pada hubungan sintagmatik karena sifatnya yang linear, maka set merujuk pada hubungan paradigmatis, sebab kata atau unsur yang berada dalam satu set dapat saling menggantikan satu sama lain dalam suatu konteks. Misalnya, dalam kalimat yang sama, kata *kucing* bisa diganti dengan *anjing* atau *burung* tanpa mengubah struktur kalimat, tetapi mengubah maknanya. Suatu set biasanya berupa sekelompok unsur leksikal dari kelas yang sama yang tampaknya merupakan satu kesatuan. Setiap unsur leksikal dalam suatu set dibatasi oleh tempatnya dalam hubungan dengan anggota-anggota dalam set tersebut. Misalnya pada kata *remaja* merupakan tahapan pertumbuhan antara *kanak-kanak* dengan *dewasa* dan kata *sejuk* merupakan suhu di antara *dingin* dengan *hangat*. Dengan demikian, apabila dibagangkan kata-kata yang berada dalam satu set dengan kata *remaja* dan *sejuk* adalah sebagai berikut:

SET (paradigmatik)	bayi	dingin
	kanak-kanak	<i>sejuk</i>
	<i>remaja</i>	hangat
	dewasa	panas
	manula	terik

Pengelompokan kata berdasarkan kolokasi dan set dapat memberikan gambaran jelas mengenai teori medan makna sebagaimana yang telah dikemukakan oleh Chaer dan Pateda. Kolokasi dan set tidak dapat dipisahkan karena keduanya merupakan bagian penting dari bagaimana kita memahami dan menggunakan bahasa, meskipun keduanya terkadang tumpang tindih dan batas-batasnya seringkali menjadi kabur.

2.4.5 Terminologi Sobis

Istilah Sobis atau Passobis merupakan istilah lokal di Sulawesi Selatan yang digunakan sebagai julukan kepada seseorang atau sekelompok orang tertentu yang melakukan aktivitas penipuan online. Terminologi itu lalu dimaknai oleh Ulfa dan kawan-kawan (2023) berasal dari istilah “sok bisnis” yang merujuk pada tindakan kejahatan dunia maya (cybercrime). Berdasarkan temuan Fausiah (2023, hlm. 23) Sobis pertama kali muncul pada tahun 2000-an di kalangan pemuda yang sebelumnya telah merantau ke berbagai daerah di luar pulau Sulawesi, yaitu Jakarta, Palembang, Kalimantan, hingga luar negeri, seperti Malaysia dan Singapore.



n beberapa tahun, istilah ini kemudian berkembang pesat. Istilah ini serupa dengan terminologi 'showbiz' dalam penelitian (30) yang merupakan istilah berbahasa Inggris yang terdiri dari kata jkatan dari kata business) atau dalam istilah lainnya disebut social ini kemudian cenderung disamakan saat diucapkan padahal

keduanya sangat berbeda. Secara harfiah, show biz berarti pertunjukan bisnis yang terasosiasi ke dalam suatu manajerial perusahaan pertunjukan atau hiburan (Akbarrrr94 2023, hlm. 2). Sobis sesungguhnya merupakan pihak kriminal yang mewakili aktivitas penipuan online lewat berbagai media, mulai dari GSM, Blackberry, Facebook, investasi, hingga modus operasi yang beragam.

2.2 Hasil Penelitian yang Relevan

Identifikasi penelitian terdahulu yang relevan dengan penelitian ini adalah penelitian yang telah mencoba mendeteksi SMS penipuan menggunakan berbagai metode seperti penerapan teknik klasifikasi pembelajaran mesin, algoritma, software statistik, maupun kombinasi metode lainnya. Misalnya pada penelitian Sjarif dan kawan-kawan (2020, hlm. 635). Sjarif mengusulkan penggunaan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan SMS Spam. Ia mengidentifikasi adanya peningkatan drastis jumlah pertumbuhan SMS Spam dalam beberapa tahun terakhir dan terbatasnya jumlah alat klasifikasi Spam hingga sering terjadi penipuan berbasis digital.

Selain itu, penerapan teknik klasifikasi pembelajaran mesin untuk mendeteksi SMS penipuan juga dilakukan oleh Guptaa (2023, hlm. 241). Guptaa mencatat, lebih dari 260 miliar SMS terkirim setiap harinya, sekitar 57 miliar di antaranya hanyalah SMS Spam. SMS seperti ini berbahaya sebab bisa jadi berisi situs *hosting phishing* atau *malware* yang dibuat untuk melakukan pencurian dan memanipulasi data pribadi, seperti informasi rekening bank.

Analisis SMS Spam atau pesan penipuan memang telah banyak dilakukan dan sedang berkembang dikalangan para peneliti. Begitu pun dalam bidang ilmu bahasa, analisis pesan penipuan menggunakan metode linguistik komputasional telah membantu mengidentifikasi pola bahasa, karakteristik teks, atau tanda-tanda linguistik lainnya yang dapat mengungkapkan potensi penipuan. Oyeyemi dan Adebola (2023) mengungkap solusi yang lebih efektif untuk mengatasi masalah SMS Spam meskipun sudah banyak teknik penyaringan SMS Spam menggunakan teknik pembelajaran mesin. Studi Oyeyemi dan Adebola memperkenalkan pendekatan baru yang memanfaatkan *Natural Language Processing (NLP)* dan model pembelajaran mesin *BiDirectional Encoder Representations from Transformer (BERT)*.

Di Indonesia sendiri, penelitian terkait hal ini juga mulai berkembang. Dwiansyaputra dan kawan-kawan (2021, hlm. 56) melalui penelitiannya melihat fenomena serupa, mereka menggunakan teknik klasifikasi pembelajaran mesin untuk menyaring SMS Spam secara otomatis. Begitu pula dengan Amaia dan kawan-kawan (2023), serta Maharani (2021, hlm. 28). Pemanfaatan linguistik korpus juga penelitian sebelumnya oleh Almucharomah (2023, hlm. 778). rusaha melihat frekuensi kemunculan kata 'corona' dengan likasi *AntConc* sehingga diperoleh sebanyak 2.993 kali kata itu idi topik utama pembicaraan masyarakat Jerman sepanjang tahun terbaru terkait pemanfaatan linguistik korpus dengan instrumen



AntConc dilakukan pada tahun 2023 oleh Simanjuntak. Ia berusaha menyelidiki pola representasi korpus pidato Presiden Joko Widodo. Ia menggunakan metode campuran yang memadukan analisis kuantitatif kualitatif untuk diolah oleh alat analisis korpus sehingga dihasilkan simpulan bahwa Jokowi menggunakan pola berbicara yang konsisten.

2.3 Kerangka Pikir

Penelitian ini merupakan penelitian yang berada di bawah naungan linguistik korpus. Cabang ilmu linguistik terapan ini berfokus pada gabungan ilmu bahasa dan perihal komputasional atau komputer. Pemanfaatan linguistik korpus pada penelitian ini menggunakan sebuah aplikasi perangkat lunak bernama *AntConc*. Aplikasi ini memiliki sembilan fitur, yaitu Key Word in Context (KWIC); Plot; File View; Cluster; N-Gram; Collocate; Word; Keyword; dan Wordcloud. Berdasarkan pertimbangan representasi data, hanya empat fitur yang digunakan sebagai landasan pengidentifikasian dan penyajian penanda lingual teks SMS *Sobis*. Pertama fitur Word, digunakan untuk mengetahui jumlah dan frekuensi entri yang paling sering muncul; fitur N-Gram, digunakan untuk menentukan ekspresi umum *Sobis*; fitur KWIC, digunakan untuk mengklasifikasikan ekspresi umum *Sobis* berdasarkan polanya; dan fitur Collocate, digunakan untuk mengetahui daftar kata atau kelompok kata yang muncul bersamaan berdasarkan informasi ekspresi umum pada teks SMS *Sobis*.



Berikut bagan kerangka pikir penelitian ini.

