

International Journal of Language Education  
Volume 8, Number 1, 2024, pp. 162-184  
ISSN: 2548-8457 (Print) 2548-8465 (Online)  
Doi: <https://doi.org/10.26858/ijole.v8i1.63833>

## **Language Choices and Digital Identity of High School Student Text Messages in the New Capital City of Indonesia: Implication for Language Education**

**Devi Ambarwati Puspitasari**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
& Universitas Gadjah Mada, Indonesia  
*Email: devi018@brin.go.id*

**Yenny Karlina**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
*Email: yenn010@brin.go.id*

**Hernina**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
*Email: hernina@brin.go.id*

**Kurniawan**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
*Email: kurn025@brin.go.id*

**Sutejo**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
*Email: sutejo@brin.go.id*

**Agus Sri Danardana**

National Research and Innovation Agency of Republic Indonesia, Indonesia  
*Email: agus.sri.danardana@brin.go.id*

*Received: 1 June 2022*

*Reviewed: 1 June 2023-1 December 2023*

*Accepted: 1 February 2024*

*Published: 30 March 2024*

### **Abstract**

This research investigates the language choices and digital identities represented in text messages exchanged by high school students in the New Capital City of Indonesia (Ibu Kota Nusantara/IKN) and the implications of these linguistic practices for language teaching. Using a corpus of messages transmitted via WhatsApp collected from 100 high school students, this study produced 2.1 million tokens of 83,414 word types. This study uses corpus linguistic analysis to investigate the distinctive features of language usage, lexical changes, and communication trends in the digital discourse of IKN high school students. Data analysis utilizes features from AntConc corpus tools, including word lists, collocations, concordances, and N-grams. Concerning the features uncovered the subtle nuances that contribute to the unique linguistic

identity of the student authors. This research highlights how digital language patterns attribute individual identity within the context of IKN high school students. The data analysis discovers three linguistic patterns in the electronic texts produced by IKN High School students, reflecting their general digital identity as language users in IKN, namely (1) lexical choice, (2) orthographic selection, and (3) lexical bundles. Furthermore, the study reveals the complex construction of digital identity through language, with students negotiating social positions, connections, and personal identities via their linguistic choices in electronic texts. The research findings contribute to the implications of language teaching in the IKN area and a deeper understanding of how students within the IKN area express themselves through language in the virtual realm, thus shaping their digital identities.

Keywords: Authorship attribution; corpus linguistics; digital text, IKN, WhatsApp

## Introduction

In recent years, digital communication technology has transformed how people engage and communicate, particularly among adolescents and young adults. The increased adoption of social media platforms, messaging applications, and mobile devices has enabled constant and instantaneous communication, influencing users' linguistic patterns and digital identities worldwide. Within the Indonesian setting, the New Capital City of Indonesia (Ibu Kota Nusantara/IKN) has risen as a center of urban development, and technical innovation has increased the reliance on digital communication among its residents, especially high school students. This study investigates the language choices and digital identities reflected in text messages exchanged by high school students in IKN, particularly emphasizing the implications of these linguistic practices for language education.

In recent years, digital communication technology has transformed how people engage and communicate, particularly among adolescents and young adults. The increased adoption of social media platforms, messaging applications, and mobile devices has enabled constant and instantaneous communication, influencing users' linguistic patterns and digital identities worldwide. Within the Indonesian setting, the New Capital City of Indonesia (Ibu Kota Nusantara/IKN) has risen as a center of urban development, and technical innovation has increased the reliance on digital communication among its residents, especially high school students. This study investigates the language choices and digital identities reflected in text messages exchanged by high school students in IKN, particularly emphasizing the implications of these linguistic practices for language education.

IKN is a distinct sociolinguistic environment defined by growing urbanization, cultural variety, and digital connectedness. As the focal point of Indonesia's ambitious urban development project, IKN represents progress and modernity, attracting a varied population of citizens across the country. Within this dynamic urban setting, high school students play a critical role in defining IKN's linguistic and digital culture by navigating the intricacies of identity-building and social engagement in digital spaces.

A digital identity is a digital depiction of an individual or institution. It refers to the distinct set of qualities, attributes, and information that can be used to identify and authenticate a person or entity online (Biró, 2020; Masiero, 2023). Digital identities are utilized to engage with digital systems, access online services, and build confidence in the digital environment. Personal information, particularly personal data, is called digital identity. It demonstrates that a person's identity in the contemporary digital era depends on biographical data, fingerprint information, and the authorship attribution they have made using technological devices (Biró, 2020; Belvisi et al.,

2020). In other words, language choice and its inherent linguistic features can be used to recognize digital identity.

In text messaging, language choice includes various linguistic features such as syntax, lexical choice, and discourse style, all contributing to digital identity formation. High school students at IKN use a variety of languages in their text messages, ranging from formal Indonesian to regional dialects, slang, and hybrid forms of language affected by digital communication. These language preferences are influenced not just by sociolinguistic characteristics like age, gender, and social standing but also by the features and conventions of digital communication platforms.

The concept of digital identity through language and its features encompasses the distinctive characteristics that make up an individual's online persona, as expressed through written communication. This concept is because online communication often lacks non-verbal cues present in face-to-face interactions, such as tone of voice and body language. As a result, the language individuals use becomes even more crucial in conveying their intentions, emotions, and personality traits. Analyzing linguistic features, such as word choice, sentence structure, and emotive expressions, can provide the nuances of online interactions and the dynamics between individuals in digital spaces. In other words, text and its linguistic features carry on the unique and identifiable aspects of an individual's online presence as their digital identity. It includes how people construct their identities and transmit their personalities, connections, feelings, and opinions using digital language, such as emails, social media posts, text messages, blog entries, or comments (Grant, 2022; McMenamin, 2001). Somehow, individuals strategically craft their online personas, and these identity performances impact social relationships and interactions. Hence, understanding the concept of digital identity through text and language is essential for comprehending the intricacies of online interaction, how individuals strategically craft their online writing personas, and the impact of the writing styles on identity performances.

Today's importance of digital communication is not emphasized, as it is adolescents' primary medium for social contact, self-expression, and identity performance. In the setting of IKN, where digital technologies are firmly integrated into daily life, comprehending high school students' language choices and digital identities in text messages is critical for gaining insights into their communication behaviors and cultural practices. This study investigates the complex interaction between language, digital technology, and identity building in IKN by investigating the linguistic features, communicative strategies, and identity indicators in high school students' text messages.

Furthermore, this study attempts to investigate the implications of these linguistic practices for language instruction and develop pedagogical approaches relevant to students' communicative needs and digital realities in IKN. This study seeks to advance our understanding of the role of language in digital communication and its implications for language instruction in urban settings by combining ideas from corpus linguistic analysis and sociolinguistic investigation. This study aims to shed light on the intricate interplay of language, technology, and identity in IKN by examining language choices and digital identities in high school student text messages.

## **Literature review**

### **Language choice**

The relationship between language choice and digital communication has received much scholarly attention in recent years, indicating that digital technologies are becoming increasingly crucial in communication practices and identity building. Language choice, defined as the selection of language elements and variations in communicative encounters, is essential in shaping

digital identity in online environments. Some research has examined the impact of language choice on digital communication platforms like social media, messaging apps, and online forums. These studies emphasize the dynamic and complex character of language choice in online interactions, with people using various linguistic elements to express themselves and communicate with others. For example, (Biró, 2020; Sagredos, 2019) discovered that language choice varied depending on factors such as social context, audience, and communicative goals, with participants using a variety of linguistic styles and registers to convey meaning and establish rapport.

Furthermore, research has investigated the relationship between language choice and digital identity, revealing how linguistic practices in online interactions help to shape and negotiate identity. (Biró, 2020; Mehmood et al., 2023; Vazquez-Calvo & Thorne, 2022) Conducted research on language choice in multilingual online communities, demonstrating how individuals utilize language to signal group affiliations, cultural identities, and social conventions, defining their digital identities inside virtual communities. Similarly, several studies investigated how people utilize linguistic elements like code-switching, slang, and emoticons to construct and perform various aspects of their identities in digital settings (Febrianti et al., 2022; Rustono, 2016; Utami et al., 2021).

Besides, the rise of digital communication tools has spread new linguistic practices and conventions, challenging long-held concepts of language use and identity. (Banga et al., 2018; Jambi et al., 2021; Belvisi et al., 2020) The constraints of text messaging, such as character limits and less formality, have created unique language elements and communication styles among users. These findings underline the dynamic and adaptable character of language choice in digital communication and its importance in determining digital identity in today's online environments.

Everyone has a unique writing style impacted by education, cultural background, personal experiences, and preferences (Aked et al., 2018; Grieve, 2023; Nini, 2018). This writing style is evident in the language, sentence structures, punctuation, and tone. Linguistic analysis can show these stylistic trends (Grieve, 2023), which then become part of an author's profile as part of an individual's digital identity. Understanding digital identity via text and linguistic features is helpful for various applications, such as authorship analysis research and digital forensics. It provides insights into how people navigate and express themselves digitally (Smith et al., 2021; Whitley & Schoemaker, 2022) and how online language use impacts their identities.

Cyberspace in Indonesia is unique, considering that various language variations emerge, mingle, and create unique language phenomena. The writing style of Indonesians has a tremendous potential for distinctiveness in Indonesian literature. Aside from letters, other linguistic features, such as words and phrases, will contribute to a highly personalized writing style. Regional languages (Ramadhani, 2018; Rustono, 2016) and term trends in cyberspace (Sukma et al., 2021; Tur, 2019) can potentially affect personal writings in Indonesia. More specifically, regarding the specificity of language choice in Indonesia, the uniqueness of the language phenomenon currently often in the spotlight is the language used in IKN.

Language choice among teenagers has changed significantly since the introduction of digital communication platforms like social media and messaging applications. In IKN, where fast urbanization and technology breakthroughs are transforming sociocultural landscapes, it is critical to investigate how high school students use language and lexical variants in text messages. Previous research reveals the complex interplay between language choice, digital communication, and digital identity. Researchers can acquire valuable insights into the intricate interplay between language, technology, and identity-building in the digital age by investigating how individuals negotiate their identities through linguistic practices in digital contexts. Understanding these

processes is critical for understanding educational practices, policy decisions, and theoretical frameworks surrounding digital communication and identity creation.

#### Language-expressed digital identities

Linguistic analysis has emerged as a critical method for identifying digital identities. Researchers investigated the language characteristics that distinguish people in the digital domain (Bouncken & Barwinski, 2021; Masiero, 2023; Mulyono et al., 2023). Stylometric studies have revealed several writing styles distinguished by word choice, sentence structure, and emotional tone (Langlois, 2021; Belvisi et al., 2020). N-gram analysis has aided in this attempt by discovering word or phrase sequences that serve as linguistic fingerprints (Cicres & Queralt, 2019; Grieve et al., 2019; Belvisi et al., 2020). Concordance technologies have enabled academics to investigate the contextual use of language, exposing how specific phrases or idiomatic expressions are used inside writers' digital identities (Alamri, 2023; Chen & Flowerdew, 2018).

Language-expressed digital identities reflect sociolinguistic dynamics. Researchers have discovered how people modify their language to conform to the standards of distinct online groups and subcultures (Biró, 2020; Pérez-Sabater, 2015), highlighting the impact of social affiliations on linguistic choices (Adams et al., 2022; Li, 2019; Whitley & Schoemaker, 2022). Furthermore, the digital environment has allowed for the coexistence of various languages, the interaction of languages, and the evolution of digital dialects (Mandal, 2019; Puspitasari, 2022; Cardoso, Aeni, & Muthmainnah, 2023). These sociolinguistic variances reflect an individual's identity and broader cultural and societal forces.

Language research on digital identities highlights the complex interplay between linguistic traits, sociolinguistic dynamics, and the transformational power of the digital environment. Understanding the intricate interplay between language and identity is critical as the digital landscape evolves. Researchers investigate the linguistic fingerprints, sociolinguistic influences, and multimodal presentations that make up the colorful tapestry of digital identities, providing vital insights into how people navigate and express themselves in the digital era.

#### Cultural and linguistic phenomenon in IKN

Indonesia's new capital city, IKN, can reshape Indonesia's landscape and set an example for urban development around the world with careful planning and strategic vision. IKN emphasizes modernism and sustainability while honoring Indonesia's varied cultural heritage (Hasibuan & Aisa, 2020; Limas et al., 2022). The city's urban planning will include traditional Indonesian architecture and design (Fristikawati et al., 2022), reflecting the country's rich history and regional diversity. This combination of cultural legacies will give the new capital a sense of identity, pride, and continuity.

Considering Indonesia's cultural richness, residents in the IKN area are likely to speak various regional languages and dialects, depending on their ethnic backgrounds. Local languages spoken by various ethnic groups around the country include Javanese, Sundanese, and many others. These regional languages, which have cultural and historical significance in their communities, frequently coexist with Indonesian. However, it is essential to remain cognizant that language usage patterns can change over time, and establishing a new capital city may impact regional language dynamics. Indonesian's position as the official language will likely remain in place, but local languages may thrive as a means of cultural expression and community identification. It has resulted in a linguistic issue that significantly impacts the IKN population,

particularly teenagers who represent language and culture successors and are most closely connected with technology.

The emergence of the digital age has revolutionized communication, socialization, and self-expression, profoundly impacting how people interact and engage with one another. This transformation is particularly evident among teenagers, who have embraced many digital platforms as integral tools for social connection and self-expression. The language they employ in electronic writing, such as social media posts, text messages, and online chats, shapes their digital identities significantly. Understanding how teenagers, especially high school students, develop their digital identities through language is an increasingly important topic. This topic is even more attractive for teenagers in areas that are intensively facing significant changes in this digital and technological era, namely high school students in IKN.

### Corpus linguistics' role in identifying an author profile through N-gram features

Corpus linguistics involves the collection of extensive language data. There are different sorts of corpora. However, corpora for authorship analysis are typically constructed primarily (Baker et al., 2006; Shiroda et al., 2023). A particular corpus, i.e., authorship corpus, is a corpus of data that includes written texts produced by the target authors (Grieve, 2023). This corpus can encompass various types of texts, such as essays, emails, social media posts, or any content authored by the individuals under investigation. By employing multiple qualities supplied by corpus analysis tools, corpus linguistics plays a critical role in recognizing an individual's digital identity through the electronic text they create (Belvisi et al., 2020). These characteristics, including word lists, collocations, concordances, and n-grams, allow academics to delve into various writers' unique language patterns and styles.

Furthermore, word lists generated by corpus technologies display the frequency of specific words in each text or set text. Researchers can uncover common language markers that distinguish one individual from another by analyzing the most used terms. These identifiers include personal pronouns, common language, and even specialized topic-related phrases unique to the author's digital identity. On the other hand, collocation analysis identifies terms that appear together frequently in an author's works. These collocations might provide information about a person's linguistic habits, such as using unique idiomatic expressions, domain-specific terminology, or even colloquialisms that add to their digital identity. Besides, examining how an author utilizes specific terms or phrases in different situations can be recognized by taking concordance analyses. Concordances enable the investigation of the contextual use of a term or phrase in an author's texts (Ruan, 2020; Shiroda et al., 2023). Concordances can provide insight into writing style, objectives, and digital identity nuances. They can also reveal an author's usage of slang, idiomatic language, or specific phrases that may be linked with various digital communities or subcultures.

In addition, another feature that is also very useful for getting a more specific picture of a writer's profile pattern is the N-gram (Banga et al., 2018; Grieve et al., 2019; Belvisi et al., 2020). N-grams (n-word sequences) can be used as an authorial signature, especially when studying more extensive sequences such as trigrams or quad grams. It can detect distinctive language patterns characteristic of the author's digital identity by evaluating the frequency and distribution of N-grams. N-grams reveal an author's writing style, word selection, and adherence to specific registers. These linguistic indications help to identify an author's digital identity.

In brief, corpus linguistics and its associated tools allow scholars to examine electronic texts for linguistic traits that define an author's digital identity. Word lists, collocations, concordances, plots, and N-grams provide helpful information about a person's writing style,

theme choices, and language usage. These technologies aid in detecting an individual's digital identity by revealing the linguistic idiosyncrasies that distinguish their electronic messages. Accordingly, this research project investigates the complicated relationship between digital identity and linguistic patterns in electronic writing. This study, which focuses on IKN students, investigates the authorship characteristics of a corpus of WhatsApp messages.

The role of corpus linguistic analysis on the implications of language teaching

Corpus linguistic analysis influences the implications of language instruction by giving valuable insights into language use, variation, and acquisition. Corpus linguistic analysis allows educators to get a thorough understanding of actual language usage in a variety of situations (Febrianti et al., 2022; Thao & Khoi, 2022), genres (Barlas & Stamatatos, 2021; Budiwiyanto, 2023), and registers (Budiwiyanto, 2023; Wahyuningsih, 2022). Educators can detect frequent patterns, collocations, and lexical differences in natural language use by analyzing vast volumes of texts (Puspitasari et al., 2023). This empirical research is the basis for creating language teaching materials and curricula representing the linguistic realities of learners' everyday conversations. The corpus-linguistic analysis also enables the investigation of language variance and diversity (Puspitasari et al., 2024), which is critical for establishing inclusive language teaching techniques. Corpus data enables educators to identify linguistic traits shared by various social groups, regions, and cultural backgrounds. By adding a variety of linguistic examples into language instruction, educators can legitimize students' language repertoires, encourage linguistic diversity, and develop a sense of belonging among students from varied backgrounds.

Furthermore, corpus linguistic analysis promotes the development of learner-centered methods for language training. Instructors can use learner corpora to uncover prevalent errors, challenges, and developmental patterns in students' language production (Cinato & Verdiani, 2023; Pham, 2022). This data informs targeted instructional interventions and remedial procedures designed to address the unique requirements of individual students. Furthermore, corpus data can be leveraged to build learner-friendly resources that are relevant, authentic, and engaging, increasing student motivation and learning results (Lee & Park, 2023; Oktavianti et al., 2022). Corpus linguistic analysis adds to improved language competency assessment and evaluation techniques. Educators can create valid and reliable assessment systems that accurately measure learners' communication competence by analyzing the corpora of written and spoken language samples. The corpus-based assessment identifies significant language elements and criteria for evaluation, resulting in more objective and standardized assessment processes (Thao & Khoi, 2022). Corpus data can also be used to build diagnostic examinations, formative assessments, and feedback mechanisms that help learners develop their language skills over time.

As a result, corpus linguistic analysis provides empirical data, promotes linguistic diversity, supports learner-centered approaches, and improves language assessment processes, all impacting language instruction. Using corpus data, educators can create pedagogically sound and evidence-based language teaching approaches that effectively suit the different demands of today's multilingual and multicultural learners.

## **Research method**

### **Research design**

This research employed a mixed-method approach, combining quantitative and qualitative elements to comprehensively understand the phenomenon under investigation (Busse, 2021). Quantitatively, it involves using electronic tools like AntConc as the corpus toolkit to count words

and tokens efficiently (Khairas, 2019; Pham, 2022). This approach allows for objective measures and assessment of word frequency and distribution (Hasan, 2021; Hernina et al., 2023; Karlina, 2023; Puspitasari, 2022). This research aimed to explore language variances among high school students.

### Participants

The data was gathered from 100 high school students' WhatsApp messages. These individuals represented a broad demographic, allowing for a thorough examination of language variations among the sample population. The research period took seven days, ensuring an adequate corpus of texts for examination.

### The procedure of data collection

During seven days, the data-collecting technique involved documenting or collecting WhatsApp messages from the participants. Before data collection, participants were told about the study's objectives and given their consent. Ethical approval was obtained to ensure compliance with privacy guidelines, focusing primarily on language changes within the texts and respect for each participant confidentially. Information regarding respondent details and text data appears in Table 1 below.

Table 1. Data source summary

Respondent Category (author/authorship)	Specifications	Total (%)
Gender	Male	42%
	Female	58%
Age	15 years old	2%
	16 years old	22%
	17 years old	41%
	18 years old	35%
language mastery of the respondents	Bahasa Indonesia	100%
	Local language ( <i>Bahasa Daerah</i> )	72%
	Foreign language (English)	100%
place of residence of the respondent (author)	Kabupaten Penajam Paser Utara	50%
	Kota Balikpapan	45%
	Other regions of the IKN area	5%

### Data analysis of corpus

The AntConc corpus tool provides extensive linguistic analysis features and employs corpus analysis (Khairas, 2019). The WhatsApp texts corpus of these 100 unique authors contains 2.1 million tokens and 83,414-word types. Information related to the corpus appears in Table 2 below. The primary focus is on analyzing word lists, lexical collocations, and the order and arrangement of words within texts. Using a corpus linguistic approach, it explores the distinctive features of language usage within the electronic discourse of IKN high school students. Data was analyzed using features from the AntConc corpus tool, namely word lists, collocations, concordances, and n-grams, which were used in a thorough examination, providing insights into language variances among high school students. N-gram tracing is carried out at character and word levels to identify authorship attribution markers for each text.

Table 2. Description of corpus data

Data Information	Total
Respondents/Authors	100
Total text /messages	300
Word type (from all tokens)	65.814
Word type (average in one text)	2.000 - 4.000
Token (all entries)	2.086.165
Average token in a text-set	50.000 - 100.000

### The legitimacy of corpus

Corpus linguistics is a well-known paradigm in linguistic research that takes a systematic and empirical approach to investigating language use. Corpus analysis is legitimate because it follows stringent methodological requirements, such as using representative data samples, transparent analytical techniques, and ethical considerations. Corpus linguistics allows us to find patterns and trends in language use that might otherwise go undetected. Furthermore, using corpus tools like AntConc improves the efficiency and accuracy of language analysis, resulting in reliable conclusions. Corpus linguistics provides a valid and reliable paradigm for studying language variations and patterns in digital communication environments.

## Results

### Digital identities of IKN students

The WhatsApp text data collected from high school students at IKN (*Ibu Kota Nusantara*) provides unique insights into the digital identities that emerge in this dynamic urban milieu. Several digital identities may be discerned from the text data, illustrating the diverse nature of communication among IKN's younger generation. First, WhatsApp text data may disclose identities shaped by linguistic preferences and language differences. High school students frequently demonstrate linguistic originality and innovation by introducing slang, acronyms, and colloquial terms into their digital communications. These linguistic options reflect individual preferences and help build unique group identities within the IKN community. This research identified the linguistic markers that characterize social groupings and subcultures among high school students by evaluating language use patterns.

Through this research data, digital identities in WhatsApp text messages prove that they are linked to cultural affinities and manifestations of identity. High school students at IKN come from various cultural backgrounds, representing many areas, nationalities, and traditions throughout Indonesia. Their WhatsApp communication represents cultural values, norms, and practices, revealing how cultural identities are negotiated and articulated digitally. Emoticons, memes, and references to cultural events or symbols may all function as cultural identification markers in text data. Below are some examples of data that reflect the diversity of identities reflected in the text.

- (1) *Mau plg kah sdh?* [Do you want to go home yet?]
- (2) *Tuntas semua kah itu?* [Is it all finished?]
- (3) X: *Minta uang sm ayah* [Ask your father for money]  
Y: *Iye Bu* [oke, mom]

Datum (1) and Datum (2) are examples of speech that use the particle *kah*, where this particle is a characteristic particle characteristic of languages in Kalimantan, such as Banjar and

Bugis. Moreover, in datum (1), there is a composition of the phrases *kah* and *sdh*, which are characteristic of the local language. Based on the two data, it could uncover the identity of the speakers from the local areas in Kalimantan. In datum example (3), the word *eye* means "yes." *Iye* is a typical word from the Bugis language and identical to the city of Makassar. Hence, the speaker's identity can be recognized from word choices taken from the local language, even though these markers are composed in Indonesian.

WhatsApp communication data provided insight into the social identities and interpersonal relationships formed by IKN high school students. Digital communication systems such as WhatsApp promote social interaction and networking by allowing students to connect, communicate, and collaborate with peers within and outside their immediate social networks. The text data reflects patterns of socialization, peer influence, and group dynamics, providing insight into the creation of friendships, cliques, and social hierarchies among IKN high school students.

Furthermore, the WhatsApp text data used in this study provide insights into how high school students develop and curate their digital identities in IKN's urban setting. As citizens of a fast-rising capital city, these students are immersed in a digitally mediated world marked by technological innovation, urbanization, and globalization. Their digital identities identify influences from the challenges of urban life, as well as their involvement in global youth culture and digital trends. The WhatsApp text data acquired from IKN high school students provides a valuable resource for understanding this urban milieu's various and complicated digital identities. By examining online communication's linguistic, cultural, social, and urban components, researchers can better understand how high school students at IKN navigate and negotiate their identities in the digital age.

#### Lexical choice of IKN students

Word patterns, or lexical trends, are crucial for understanding linguistic variance among IKN high school students. By studying variations in word frequency over time, this study used massive language corpora and computational linguistic tools to find linguistic patterns and differences within and across groups (Pescuma et al., 2023). Examining word trends revealed the students' linguistic quirks and preferences, providing valuable insights into their language use. This approach allows us to identify common lexical choices (Sugianto & Hasby, 2023), developing linguistic trends (Fajri, 2020; Beyer, 2014; Puspitasari, 2022), and variances (Puspitasari, 2023) in language usage among IKN high school students. By focusing on word patterns, this study gave a nuanced understanding of language variance throughout the student population, adding to our knowledge of the linguistic landscape in IKN and informing language teaching strategies tailored to the needs of students in this diverse urban context. An overview of the lexical choices of high school students in the IKN area can be seen from the analysis of word choices with the highest frequency, shown in table 3 below.

Table 3. Top five-word list

Rank	Word	Freq.
1	<i>aku</i>	20.230
2	<i>aq</i>	15.536
3	<i>q</i>	14.246
4	<i>di</i>	13.671
5	<i>yg</i>	12.477

*Aku* is one of the most often used terms. In Indonesian, this word is the personal pronoun "I" for English, demonstrating IKN students' strong emphasis on self-expression and individuality.

Meanwhile, the word *di* means "at" or "in" in English. The frequency of contextual references and geographical dimensions in IKN communication highlights their significance, demonstrating that students engage in location-based conversations and expressions. Besides, the abbreviation *yg* means *yang*, meaning "that" or "which" in English. The abbreviation shows IKN high school students' penchant for employing shortened language patterns, which is common in online communication.

Similarly, the widespread use of the term *di* meaning "at" or "in" exemplifies the linguistic diversity in IKN student communication. The frequent use of contextual allusions and spatial dimensions in IKN discourse emphasizes the significance of location-based interactions and expressions in the student community. This language trend shows that students actively participate in debates about their immediate surroundings and environments, including spatial allusions in their communication patterns. "di" shows a particular linguistic trait that distinguishes IKN students' language use, reflecting their distinct communication preferences and interactional norms.

Another prominent characteristic of language diversity seen among IKN high school students is using acronyms like *yg* for *yang* or "that" in English. The frequent use of this abbreviation demonstrates students' proclivity for using shorter language patterns commonly found in online communication. This language change reflects the effect of digital communication platforms such as WhatsApp, emphasizing brevity and efficiency in text-based conversations. Using *yg* reveals the student's ability to navigate digital communication venues and familiarity with current communication trends. Overall, the abundance of linguistic variations, such as abbreviations, enriches and complexifies IKN students' language repertoires, demonstrating the dynamic character of language use in this varied and digitally connected community.

Peers and individual communication styles impact the frequent use of the first-person pronoun *aku* among teens in IKN. This finding has important implications for language instruction, particularly in understanding the socio-cultural elements influencing language use among young learners. Language educators can add lessons about peer influence and communication styles into their curriculum to help students understand how linguistic trends spread and are reinforced within social groupings. Educators can empower students by engaging them in critical thoughts on their language choices and the impact of peer dynamics.

Individuals naturally communicate their ideas and emotions with first-person pronouns in various languages, including Bahasa Indonesia. Language teachers can use this natural propensity by creating instructional activities that encourage students to engage in self-referential communication and personal expression. Educators can improve students' linguistic fluency and communication competency by allowing them to practice using first-person pronouns in various communicative contexts while promoting autonomy and self-expression in language learning.

Despite IKN's international nature, cultural customs and personal preferences play a role in generating linguistic variance. Linguistic educators can use this knowledge to create culturally responsive teaching methods that acknowledge students' linguistic backgrounds and preferences. Recognizing the impact of cultural influences on language use allows educators to establish inclusive learning environments in which students feel valued and respected for their linguistic identities. This strategy encourages linguistic diversity and intercultural understanding while improving the language learning experience for all students.

The media and contemporary culture have had a limited impact on IKN linguistic differences, particularly in terms of using personal pronouns. Language instructors can use this discovery to combat prejudices and misconceptions about language evolution and variation.

Educators may foster a more nuanced knowledge of linguistic diversity and language evolution by teaching students to examine media depictions of language and culture critically. This critical language awareness can enable students to influence language norms and practices in their communities, resulting in a more inclusive and dynamic linguistic landscape in IKN.

#### Orthographic selection

N-gram tracing is essential in electronic communications on platforms such as WhatsApp consequences for language variety and training. N-grams, or sequences of N-contiguous words, offer valuable insights into language patterns, communication styles, and linguistic behaviors in text data. For language teachers, this provides an opportunity to investigate real-world language use and introduce students to actual communication methods. Educators can identify linguistic patterns and variances in electronic texts by analyzing N-grams. This analysis allows them to create teaching resources that match students' language realities in digital environments.

At the word level, N-gram tracing allows educators to investigate how words appear in context, revealing information on language diversity, slang usage, and linguistic advances. N-gram tracing also helps language teachers expose children to various language variants and registers, preparing them to navigate varied linguistic environments efficiently. Incorporating N-gram analysis into language instruction can improve students' language awareness and competency by exposing them to actual language patterns and phrases used in digital communication.

Similarly, at the character level, N-gram tracing provides a unique perspective on language variances and linguistic characteristics exclusive to chat participants. By examining character-level N-grams, educators can identify specific linguistic traits such as abbreviations, emoticons, and non-standard spellings often used in electronic communication. This knowledge can enhance language teaching techniques by emphasizing the importance of digital literacy skills and increasing students' critical awareness of language variance. Educators can incorporate discussions on character-level N-grams into language lessons to help students comprehend how language adapts and evolves in digital environments.

N-gram tracing is a practical approach for identifying language changes and linguistic activities in electronic communications such as WhatsApp discussions. By introducing N-gram analysis into language teaching, educators may enrich students' learning experiences, raise language awareness, and provide them with the skills they need to speak effectively in various linguistic contexts. This strategy promotes the development of communicative competence and digital literacy skills, which are critical for success in today's interconnected world.

A pattern of words frequently used by IKN high school students was discovered in the data. Each participant, however, writes each word differently. For example, the word *aku* [I/me/my], which topped the N-gram tracing ranking as the most common word, is spelled in five different ways. The five observed variations are *aku*, *aqua*, *ak*, *aq*, and *q*. However, each writer employs these variations regularly, as shown in table 4.

Table 4. Word Variant of *aku*

Word Choice	N-gram	Token (variation)	Found in... Text Set/s	Freq. in Corpus
<i>aku</i>	n-1	q	19	14.246
[I/me/my]	n-2	ak	14	13.752
		aq	22	15.539
	n-3	aku	31	20.230
		aqu	1	52

The importance of authorial style and individual writing patterns at the character level determines authorship. According to McMenamain (2001), the uniqueness of an author's writing style, as demonstrated by the variety and frequency of characters used, maybe a marker for identifying the author. Researchers such as Beyer (2014), Grieve et al. (2019), and Belvisi et al. (2020) have noted that the author's distinctive writing style is determined by their habitual usage and recurrence of specific grammatical patterns.

N-units at the character level include grammatical patterns and word choice since authors intentionally choose specific words and use structured writing. It means that linguistic elements at the character level, such as word choice and writing structure, might provide important insights into an author's writing style and be used as authorship markers. By examining N-units at the character level, researchers can delve deeper into textual linguistic intricacies, providing a more complete comprehension of authorial styles and writing processes.

Word abbreviations are popular and frequently used among IKN youths, according to data from word frequency, N-gram, survey, and interview. Abbreviations allow users to transmit information quickly and efficiently, especially in digital communication where character limits may exist. Shorter forms save time and effort, making communication easier and faster. Teenagers can foster a sense of in-group identity by using abbreviations to indicate membership in a specific social group or to exclude individuals who do not understand the code.

In terms of lexical usage, abbreviations are also idiosyncratic characteristics. In language and linguistics, an idiosyncratic feature is a distinct or individual aspect of a person's language use. It refers to an individual's personal preferences, habits, and choices when using words or expressions that may diverge from the norms or conventions of the greater linguistic community. The discovery of idiosyncratic lexical features in the Corpus is consistent with previous research (Belvisi et al., 2020; Grieve et al., 2019; McMenamain, 2001), that unique lexical usage can provide insights into a small community's personality, background, or experiences, and may become markers of their distinct linguistic style.

Abbreviations are another example of unique lexical usage. In language and linguistics, an idiosyncratic characteristic is a distinct or unique aspect of a person's language usage. It refers to an individual's personal preferences, habits, and decisions when using words or expressions that deviate from the norms or conventions of the broader linguistic community. The discovery of idiosyncratic lexical features in the Corpus is consistent with previous research (Belvisi et al., 2020; Grieve et al., 2019; McMenamain, 2001), which found that unique lexical usage can provide insights into a small community's personality, background, or experiences, as well as serve as markers of their distinct linguistic style.

#### The most frequent lexical bundle

A "lexical bundle" encapsulates words that consistently appear together within a specific language or linguistic context (Fajri & Okwar, 2020; Petchprasert, 2021). These bundles represent recurring combinations of words frequently in spoken and written language. They are considered vital linguistic units due to their significant impact on language fluency, comprehension, and communication efficiency, as highlighted by Baker et al. (2008). Lexical bundles let language users communicate more effectively by providing ready-made phrases and expressions that convey meaning efficiently. By incorporating often-occurring word combinations, speakers and writers can improve the coherence and cohesion of their conversation, resulting in more transparent and more coherent communication. Furthermore, lexical bundles help to improve language proficiency

by speeding up language production and comprehension, allowing people to express themselves more fluently and better understand the speech or writing of others.

Studying lexical bundles also offers valuable insights into language acquisition and proficiency development. Researchers can identify vital linguistic features that characterize proficient language use by analyzing lexical bundles' prevalence and usage patterns in different contexts and registers. This information can inform language teaching and learning practices, helping educators design instructional materials and activities that target acquiring and mastering essential lexical bundles. Overall, exploring lexical bundles enriches our understanding of language structure, usage, and proficiency, offering practical implications for language teaching, communication, and linguistic research.

Lexical bundles range from two to many long words and can be found in spoken and written language (Baker et al., 2006; Baker, 2008). Because they are employed and recognized as a single chunk rather than separate words, they are frequently referred to as multi-word chunks or formulaic sequences (Petchprasert, 2021). Because these bundles are ubiquitous, native speakers often interpret them as cohesive units rather than evaluating each word individually. The statistics reveal that lexical bundles are becoming popular in WhatsApp texts for IKN high school students. The data are shown in table 5.

Table 5. The pattern of lexical bundle

N-gram (word level)	Lexical Bundle
N-2	function word + <i>kah</i> , <i>nah</i> <i>aja kah</i> [just do it] <i>itu nah</i> [that's it]
	Verb + <i>sudah</i> <i>minum sudah</i> [just drink] <i>taruh sudah</i> [put it down]
	<i>Lah</i> + Noun/Pronoun <i>Lah coeg</i> [oh shit] <i>Lah kamu</i> [yeah, you]

These bundles have a range of functions, including expressing thoughts and concepts and emphasizing points; specific bundles are linguistic particles in the Indonesian version of IKN. The data analysis also found noteworthy tendencies with word particles in the primary variety of Indonesian used by IKN high school students. The employment of particles at the end of words like *lah*, *kah*, and *nah* adds nuance and emphasis to the speech.

The particle *kah* appears 5,629 times in the corpus study, underscoring its importance in Indonesian language usage. 'Kah' is widely used to form yes or no queries or to request agreement or affirmation, indicating its usefulness in fostering participatory discourse. This study has far-reaching consequences for language teaching, particularly in pragmatics and discourse approaches. Language educators can take advantage of the frequent use of *kah* to teach students about question-building and conversational methods in Indonesian. By emphasizing its use in seeking confirmation or agreement, educators can assist students in understanding how to engage in

dialogue and elicit replies from interlocutors effectively. Furthermore, adding examples of *kah* in language training materials can help students improve their communicative competence and pragmatic awareness, allowing them to traverse varied communication circumstances confidently.

The inclusion of *kah* in the corpus is consistent with earlier studies on authorship attribution, highlighting the importance of distinctive terms and patterns in language use. This idea can help educators improve language teaching practices by encouraging them to expose pupils to various linguistic registers and styles. Educators can help students develop their ability to grasp and produce language in various contexts and genres by exposing them to different linguistic elements and discourse conventions. The study of *kah* and its implications for language teaching emphasizes the value of using actual linguistic data in language learning. Educators can improve students' language and pragmatic competence by incorporating corpus-based findings into their teaching techniques, providing them with the tools they need to communicate effectively in real-world circumstances.

## Discussion

Intercourse between language choices and digital identity

The findings of this study provide insight into the complex relationship between language choice and digital identity among high school students in IKN. Several major themes emerged from examining student text messages, emphasizing how language choice affects the development and negotiation of digital identity inside online communication networks. The investigation revealed a prominent theme: the range of linguistic choices high school pupils use in text messages. The corpus analysis revealed diverse linguistic features, including formal Indonesian, regional dialects, slang, and hybrid language influenced by Internet communication. This variation reflects the varied character of digital identity, as students use various linguistic tools to express themselves and engage with others in online environments.

Furthermore, the study revealed language use patterns that reflect students' negotiation of social identities and group memberships. For example, frequent usage of slang, emoticons, and abbreviated forms of language may indicate affiliation with specific peer groups or subcultures within IKN. Similarly, code-switching between languages can indicate cultural hybridity and linguistic inventiveness as students traverse numerous linguistic and cultural identities in digital spaces.

Additionally, the study found that students use language choice to portray themselves and manage their impressions. Students can shape how others perceive them in digital settings by adopting specific linguistic styles or vocabulary. This research emphasizes the performative character of digital identity, with language choice used as a tool for people to create and portray their chosen online personas. The data analysis results also demonstrate the dynamic and developing nature of language choice in digital communication, with students tailoring their linguistic practices to the affordances and norms of online platforms. For example, shorter forms of language and informal registers may be more common in text messaging than in formal written communication, reflecting the casual and spontaneous character of digital exchanges between peers.

In general, the findings of this study assist in comprehending the complicated interplay between language choice and digital identity among high school students in IKN. By investigating the linguistic elements and communicative patterns found in text messages, this research is in line with (Biró, 2020; Mehmood et al., 2023; Vazquez-Calvo & Thorne, 2022) that language influences the development and negotiation of digital identity in online communication platforms. These

findings are significant for language education because they emphasize combining digital literacy skills and critical language awareness into language teaching approaches to prepare students for effective communication in digital environments.

Language's role as a digital identity means

In the digital age, where much of our lives occur in virtual places, language is a powerful tool for establishing and communicating our digital identities. As the significant medium of communication in the digital realm, language conveys the intricacies of our personalities, backgrounds, connections, and aspirations. The findings of this study demonstrate the multidimensional significance of language in building digital identities, including linguistic analysis, sociolinguistic elements, and the tremendous impact of language on the virtual personas we create. A person's digital identity is built around a unique linguistic DNA (Grant, 2022). Our digital persona is defined by how we express ourselves through words, sentences, and phrases. Language analysis, typically driven by tools such as stylometry, investigates the tiny language features that distinguish each of us in the digital world.

Lexical and orthographic selection, two key strategies of corpus linguistic research (Baker et al., 2008), delve deeply into an individual's writing style (Grieve, 2023). It examines sentence structure, word choice, punctuation, and emotional tone. These elements all contribute to a person's linguistic fingerprint. The research data demonstrates that a person's phrase pattern and word choice will generate a distinct digital identity, which extends to numerous digital writings.

N-gram tracing, encouraged by corpus linguistics, extends language analysis by investigating sequences of N words (Grieve et al., 2019). Unigrams (individual words) or longer sequences such as bigrams or trigrams have been used in this research. This research discovers linguistic patterns that become typical of an individual's digital identity by studying the frequency and distribution of these sequences. Specific words or word combinations serve as a linguistic signature, distinguishing oneself from others. This method helps determine authorship and recognize the linguistic traces that an author leaves behind.

Frequency analysis is another fundamental part of corpus linguistic analysis (Baker et al., 2008) that identifies the most frequently used terms in an individual's texts (Hernina et al., 2023). The research data demonstrates that personal pronouns are frequently used in vocabulary; even topic-specific keywords are typically included among these famous words (Suhardi, et.al., 2023). The predominance of these terms is used to create a word cloud of a person's digital identity. The frequency of specific words gives linguistic markers that distinguish people. It is identical to discovering the linguistic foundations of one's digital identity.

Digital identities are about more than just the individual (Masiero, 2023; Farida, Supardi, & Muchtar, 2023; Barbe, et.al., 2023). They are also about the communities and cultures in which they exist. The use of language reflects sociolinguistic characteristics that shape our online personalities (Whitley & Shoemaker, 2022). IKN high school students frequently participate in online communities, forums, and social communications. These groups have their own linguistic rules, jargon, and discourse practices. Participating in these online forums necessitates altering one's language to conform to these norms, constituting a defining component of an individual's digital identity. The research data reveals that this community uses local-specific vocabulary and idioms, which form their digital identity within the IKN community. Data has shown that their choice of words in digital communication shows their regional identity, such as using typical Kalimantan particles and Bugis language vocabulary (see Data 1-3).

Multilingualism and language contact are encouraged by the digital domain. Individuals may transition between languages depending on the situation and audience. Language transitions from native to second languages reflect a person's linguistic adaptability and cultural affinities. Multilingualism is an essential aspect of digital identity since it demonstrates linguistic versatility. Language reflects components of a person's socioeconomic and cultural identity. Language usage can provide information about education, employment, socioeconomic class, and cultural background. Moreover, contextual flexibility characterizes language in the digital domain. Different communication venues and modes necessitate different linguistic tactics. Language on professional networks is frequently formal, emphasizing accomplishments, abilities, and experience. Social media sites, on the other hand, foster more colloquial, emotional, and casual language. This adaptability highlights a person's verbal skill in building their digital identity.

Language is more than just a method of communication in the digital age; it is also a potent tool for expressing one's digital identity. Individuals are distinguished by their linguistic DNA, which reveals their writing style and preferences. Sociolinguistic dimensions highlight the importance of online communities, multilingualism, and cultural aspects—language's transformative potential impacts how people adjust their verbal expression across various digital platforms and modalities. The language of digital identity constantly expands, encapsulating the subtleties of our personalities and affinities in the broad digital universe.

#### Implication for language teaching

This research tries to present important implications for language training, especially in the setting of high school education in Indonesia's future capital city (IKN). The study provides valuable insights into language education methods and curriculum development by analyzing the language choices and lexical differences found in text messages exchanged by high school students. This study's findings suggest that educators must notice and incorporate the linguistic diversity and variances in students' internet communication into language instruction. High school students in IKN use a variety of linguistic styles, including slang, colloquialisms, and abbreviations that may differ from formal written or spoken language. Language instructors can use these insights to create teaching resources that mirror the linguistic realities of their students' daily conversation, making language learning more relevant and engaging.

This study also emphasizes the significance of increasing linguistic awareness and language variation competency among high school pupils. Educators can use the findings to urge students to critically reflect on their language use and grasp how language choices differ based on context, audience, and goal. By cultivating metalinguistic awareness, language educators can help students navigate numerous language registers and adjust their communication abilities to varied social and cultural contexts. Furthermore, this study emphasizes integrating digital literacy skills into language education programs. In today's digital world, text messaging and digital communication are essential components of students' daily lives. Language teachers can capitalize on this by incorporating digital communication tools like WhatsApp into language learning exercises. Educators can improve students' digital literacy by studying and discussing text messages they send while encouraging linguistic competence and critical thinking.

This study highlights the importance of language education in developing inclusive and culturally relevant pedagogies. High school students at IKN come from various cultural and linguistic backgrounds, and their digital communication reflects that. Language educators can use the study's findings to create inclusive instructional approaches that celebrate linguistic diversity, affirm students' language repertoires, and encourage intercultural understanding and empathy

among students. Therefore, this study has significant consequences for language instruction and education. Language educators can enrich students' learning experiences by embracing linguistic diversity, promoting metalinguistic awareness, incorporating digital literacy skills, and fostering inclusive pedagogies.

## Conclusion

The new capital city (Ibu Kota Nusantara/IKN) exemplifies a distinct blend of socio-cultural features, fusing traditional rural values with metropolitan influences. As the new capital develops and urbanization transforms the environment, IKN's socio-cultural fabric remains in constant motion—the convergence and symbiosis of traditional rural values and urban dynamics. Analyzing the language patterns in high school students' electronic texts at IKN yields valuable information regarding their digital identities. Three distinct patterns emerge: a strong emphasis on self-expression via the frequent use of 'aku' [I/me/my], personalized writing styles distinguished by grammatical deviations and the use of abbreviated words, and the incorporation of unique regional words and phrases such as 'iye' and 'aja kah' in WhatsApp communication. These patterns represent the community's changing language dynamics, demonstrating a sophisticated interplay between traditional linguistic features and contemporary digital communication activities set to continue as defining characteristics of this changing environment.

Interestingly, the study demonstrates a propensity among students not to use local language at the vocabulary level, except for distinctive particles. Instead, linguistic elements in digital communication serve as indicators of individual digital identity, reflecting each student's unique writing style, language choice, and communicative preferences. These linguistic signals provide vital insights into how IKN students navigate and shape their identities in the digital environment, emphasizing the importance of language choice in developing digital identity.

Language choice is an essential factor in developing digital identity among IKN students. Students actively navigate and shape their identities in digital communication by making conscious decisions about the language they use and how they show themselves online. The observed patterns of language use in digital communication demonstrate the dynamic nature of digital identity development, in which language choice is critical. Students' linguistic repertoire, including their vocabulary, syntax, and stylistic preferences, contributes to developing their online image. By studying these language cues, this study gained a better understanding of the varied nature of digital identity and the complicated processes by which it is negotiated and formed online.

Drawing on these findings, educators must notice and incorporate the linguistic diversity and variability observed in students' internet interactions into language training. By identifying and embracing these linguistic variations, educators can develop more inclusive and prosperous language teaching techniques that reflect students' communicative reality in the digital age. Furthermore, this study emphasizes additional research, arguing for deeper sociolinguistic analysis and larger datasets to investigate the broader range of language variances seen in IKN. Continued research and exploration can lead to a more thorough understanding of linguistic variety in IKN, paving the door for improved language education initiatives and socio-cultural integration within the community.

## Declaration of conflicting interest

The authors declare that there is no conflict of interest in this work.

## Funding acknowledgements

The research received no external funding.

## References

- Adams, Z., Ludwiczak, A., Sharma, D., & Osman, M. (2022). Verbal interaction in a social dilemma. *Rationality and Society*, 34(3). <https://doi.org/10.1177/10434631221094555>
- Aked, J. P., Canter, D., Sanford, A. J., & Smith, N. (2018). Approaches to the scientific attribution of authorship. In *Offender Profiling Series: Vol II: Profiling in Policy and Practice (Vol. 2)*. <https://doi.org/10.4324/9781315245942-13>
- Alamri, B. (2023). A multidimensional comparative analysis of MENA and international English research article abstracts in applied linguistics. *SAGE Open*, 13(1). <https://doi.org/10.1177/21582440221145669>
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
- Baker, P., Gabrielatos, C., Khosravini, M., Krzyzanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse and Society*. <https://doi.org/10.1177/0957926508088962>
- Banga, R., Bhardwaj, A., Peng, S. L., & Shrivastava, G. (2018). Authorship attribution for online social media. In *Social Network Analytics for Contemporary Business Organizations*. <https://doi.org/10.4018/978-1-5225-5097-6.ch008>
- Barbe, K. ., Sakkir, G. ., Syafitri, E. ., & Dollah, S. . (2023). Using Illustration Images to Enhance Junior High School Students' Writing Skills. *Journal of Language Learning and Assessment*, 1(1), 1–8. Retrieved from <https://e-journal.naifaderu.org/index.php/jlla/article/view/74>
- Barlas, G., & Stamatatos, E. (2021). A transfer learning approach to cross-domain authorship attribution. *Evolving Systems*, 12(3). <https://doi.org/10.1007/s12530-021-09377-2>
- Belvisi, N.M.S., Muhammad, N., & Alonso-Fernandez, F. (2020). Forensic authorship analysis of microblogging texts using n-grams and stylometric features. 2020 8th International Workshop on Biometrics and Forensics, IWBF 2020 - Proceedings. <https://doi.org/10.1109/IWBF49977.2020.9107953>
- Beyer, K. (2014). Urban language research in South Africa: achievements and challenges. *Southern African Linguistics and Applied Language Studies*, 32(2).
- Biró, E. (2020). Linguistic identities in the digital space. *Acta Universitatis Sapientiae, Philologica*, 11(2). <https://doi.org/10.2478/ausp-2019-0011>
- Bouncken, R., & Barwinski, R. (2021). Shared digital identity and rich knowledge ties in global 3D printing—A drizzle in the clouds? *Global Strategy Journal*, 11(1). <https://doi.org/10.1002/gsj.1370>
- Budiwiyanto, A. (2023). The use of lexical bundles is an online comprehensive dictionary of Indonesian (KBBI Daring). *Lexicography*, 10(1), 4–22. <https://doi.org/10.1558/lexi.25177>
- Busse, B. (2021). Practices of discursive urban place-making in Brooklyn, New York: (Hidden) digital and embodied discourse. *Text and Talk*, 41(5–6). <https://doi.org/10.1515/text-2021-0003>
- Cardoso, L. M. ., Aeni, N. ., & Muthmainnah, M. (2023). Mobilizing Metacognitive Strategies Through Zoom for EFL Classrooms: An Innovative Practice Amidst Covid 19. *Journal of*

- Language Learning and Assessment*, 1(1), 19–25. Retrieved from <https://e-journal.naifaderu.org/index.php/jlla/article/view/77>
- Chen, M., & Flowerdew, J. (2018). Introducing data-driven learning to PhD students for research writing purposes: A territory-wide project in Hong Kong. *English for Specific Purposes*. <https://doi.org/10.1016/j.esp.2017.11.004>
- Cicres, J., & Queralt, S. (2019). An n-gram-based approach to the automatic classification of schoolchildren's writing. *Vigo International Journal of Applied Linguistics*, 16. <https://doi.org/10.35869/vial.v0i16.93>
- Cinato, L., & Verdiani, S. (2023). Spoken language corpora for teaching German language. A proposal for CLIL in German. *RiCognizioni*, 10(19). <https://doi.org/10.13135/2384-8987/7452>
- Fajri, M. S. Al, & Okwar, V. (2020). Exploring a diachronic change in the use of English relative clauses: a corpus-based study and its implication for pedagogy. *SAGE Open*, 10(4). <https://doi.org/10.1177/2158244020975027>
- Farida, F., Supardi, S., & Muchtar, J. (2023). Web based Instruction for Immersion Class in Alauddin State Islamic University of Makassar. *Journal of Language Learning and Assessment*, 1(2), 57–61. Retrieved from <https://e-journal.naifaderu.org/index.php/jlla/article/view/73>
- Febrianti, Y., Azzahrah, F., Yannuar, N., & L. Zen, E. (2022). Multilingualism in social media: Indonesian K-pop fans language choices on Twitter. *KLAUSA (Kajian Linguistik, Pembelajaran Bahasa, Dan Sastra)*, 6(1). <https://doi.org/10.33479/klausu.v6i01.457>
- Fristikawati, Y., Alvander, R., & Wibowo, V. (2022). Pengaturan dan penerapan sustainable development pada pembangunan Ibukota Negara Nusantara [Arrangement and implementation of sustainable development in the development of the National Capital of Indonesia (IKN)]. *Jurnal Komunitas Yustisia*, 5(2).
- Grant, T. (2022). *The idea of progress in forensic authorship analysis*. Cambridge University Press. <https://doi.org/10.1017/9781108974714>
- Grieve, J. (2023). Register variation explains stylometric authorship analysis. *Corpus Linguistics and Linguistic Theory*, 19(1). <https://doi.org/10.1515/cllt-2022-0040>
- Grieve, J., Clarke, I., Chiang, E., Gideon, H., Heini, A., Nini, A., & Waibel, E. (2019). Attributing the Bixby Letter using n-gram tracing. *Digital Scholarship in the Humanities*, 34(3), 493–512. <https://doi.org/10.1093/lc/fqy042>
- Hasan, N. H. (2021). Pengaplikasian AntConc pada korpus bahasa Melayu Ambon [The application of AntConc on Ambon Malay language corpus]. *Kandai*, 17(2). <https://doi.org/10.26499/jk.v17i2.2605>
- Hasibuan, R. R. A., & Aisa, S. (2020). Dampak dan resiko perpindahan ibu kota terhadap ekonomi di Indonesia [The impact and risks of moving the capital city on the economy in Indonesia]. *AT-TAWASSUTH: Jurnal Ekonomi Islam*, 5(1). <https://doi.org/10.30829/ajei.v5i1.7947>
- Hernina, H., Karlina, Y., & Puspitasari, D. A. (2023). The Indonesian terms of disease names: A corpus linguistic study. *OKARA: Jurnal Bahasa dan Sastra*, 17(1), 14–31. <https://doi.org/10.19105/ojbs.v17i1.8571>
- Jambi, K. M., Khan, I. H., Siddiqui, M. A., & Alhaj, S. O. (2021). Towards authorship attribution in Arabic short-microblog text. *IEEE Access*, 9, 128506–128520. <https://doi.org/10.1109/ACCESS.2021.3112624>
- Karlina, Y. (2023). Corpus-assisted discourse study (Cads): Women with disabilities in Kompas.com's news reporting. *RETORIKA: Jurnal Bahasa, Sastra, dan Pengajarannya*, 16

- (2), 160-171 <https://doi.org/10.26858/retorika.v16i2.45222>
- Khairas, E. E. (2019). Using antconc software as english learning media: The students' perception. *Epigram*, 16(2). <https://doi.org/10.32722/epi.v16i2.2234>
- Langlois, J. (2021). When linguistics meets computer science: stylometry and professional discourse. *Training, Language and Culture*, 5(2). <https://doi.org/10.22363/2521-442X-2021-5-2-51-61>
- Lee, S., & Park, H. I. (2023). Changing patterns of interactive metadiscourse in English teaching articles. *English Teaching(South Korea)*, 78(2), 83–102. <https://doi.org/10.15858/engtea.78.2.202206.83>
- Li, H. (2019). Corpus approaches to contemporary british speech: sociolinguistics studies of the spoken BNC2014. Voclav Brezina, Robbie Love and Karin Aijmer (eds.). *Digital Scholarship in the Humanities*, 34(1), 230–232. <https://doi.org/10.1093/llc/fqy086>
- Limas, C., Setyaningsih, O., Putriani, O., & Fauzi, I. (2022). Konsep smart port di Ibu Kota Negara (IKN) Indonesia [Smart port concept in Indonesia's National Capital City (IKN)]. *Jurnal Penelitian Transportasi Laut*, 23(2). <https://doi.org/10.25104/transla.v23i2.1847>
- Mandal, A. (2019). Digital humanities. Year's Work in Critical and Cultural Theory. <https://doi.org/10.1093/ywcct/mbz020>
- Masiero, S. (2023). Digital identity as platform-mediated surveillance. In *Big Data and Society* (Vol. 10, Issue 1). <https://doi.org/10.1177/20539517221135176>
- McMenamin, G. R. (2001). Style markers in authorship studies. *Speech, Language and the Law*, 8(2). <https://doi.org/10.1558/sll.2001.8.2.93>
- Mehmood, A., Irfan, I., & Hussain, M. S. (2023). Social media influence on Pakistani youth's linguistic expressions: A predictive study of resultant changes. *Journal of Education and Social Studies*, 4(1). <https://doi.org/10.52223/jess.20234113>
- Mulyono, B., Affandi, I., Suryadi, K., & Darmawan, C. (2023). Online civic engagement through social media: An analysis of Twitter's big data. *Cakrawala Pendidikan*, 42(1), 12–26. <https://doi.org/10.21831/cp.v42i1.54201>
- Nini, A. (2018). An authorship analysis of the Jack the Ripper letters. *Digital Scholarship in the Humanities*, 33(3), 621–636. <https://doi.org/10.1093/llc/fqx065>
- Oktavianti, I. N., Triyoga, A., & Prayogi, I. (2022). Corpus for language teaching: Students' perceptions and difficulties. *PROJECT (Professional Journal of English Education)*, 5(2). <https://doi.org/10.22460/project.v5i2.p441-455>
- Pérez-Sabater, C. (2015). Discovering language variation in WhatsApp text interactions. *Onomazein*, 31(1). <https://doi.org/10.7764/onomazein.31.8>
- Pescuma, V. N., Serova, D., Lukasek, J., Sauermann, A., Schäfer, R., Adli, A., Bildhauer, F., Egg, M., Hülk, K., Ito, A., Jannedy, S., Kordoni, V., Kuehnast, M., Kutscher, S., Lange, R., Lehmann, N., Liu, M., Lütke, B., Maquate, K., ... Knoeferle, P. (2023). Situating language register across the ages, languages, modalities, and cultural aspects: Evidence from complementary methods. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.964658>
- Petchprasert, A. (2021). Utilizing an automated tool analysis to evaluate EFL students' writing performances. *Asian-Pacific Journal of Second and Foreign Language Education*, 6(1). <https://doi.org/10.1186/s40862-020-00107-w>
- Pham, Q. H. P. (2022). A corpus-based approach to grammar instruction in IELTS writing classes. *RELC Journal*, 53(3). <https://doi.org/10.1177/0033688220952170>

- Puspitasari, D. A. (2022). Corpus-based speech act analysis on the use of word 'lu' in cyber bullying speech. *Proceedings of the 1st Konferensi Internasional Berbahasa Indonesia Universitas Indraprasta PGRI, KIBAR 2020*, 28 October 2020, Jakarta, Indonesia, 1–10. <https://doi.org/10.4108/eai.28-10-2020.2315314>
- Puspitasari, D. A., Fakhurroja, H., & Sutrisno, A. (2023). Identify fake author in Indonesia crime cases: A forensic authorsip analysis using n-gram and stylometric features. *2023 International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS)*, 1–6. <https://doi.org/10.1109/ICADEIS58666.2023.10271069>
- Puspitasari, D. A., Hernina, Okitasari, I., & Karlina, Y. (2022). A corpus linguistic analysis of the words of aku and saya in Indonesian children's essays. In *Proceedings of the Sixth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2022)*. [https://doi.org/10.2991/978-2-494069-91-6\\_20](https://doi.org/10.2991/978-2-494069-91-6_20)
- Puspitasari, D. A., Karlina, Y., Hernina, H., Kurniawan, K., & Mulyo, B. M. (2024). Word trends in digital communication of high school students in the new capital of Indonesia (IKN). In *7th International Conference on Language, Literature, Culture, and Education (ICOLLITE 2023)* (pp. 101–108). Atlantis Press. [https://doi.org/10.2991/978-94-6463-376-4\\_14](https://doi.org/10.2991/978-94-6463-376-4_14)
- Ramadhani, A. R. (2018). Lingua franca in the linguistic landscape of gresik kota baru (gkb). *ETNOLINGUAL*, 2(2). <https://doi.org/10.20473/etno.v2i2.10569>
- Ruan, Z. (2020). Metadiscourse use in l2 student essay writing: A longitudinal cross-contextual comparison. *Chinese Journal of Applied Linguistics*, 42(4). <https://doi.org/10.1515/CJAL-2019-0028>
- Rustono, T. C. (2016). Akulturasi budaya dalam pilihan bahasa pedagang etnis Tionghoa pada ranah perdagangan di kota Salatiga [Cultural acculturation in the language choices of ethnic Chinese traders in the realm of trade in the city of Salatiga]. *Seloka - Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 5(1).
- Sagredos, C. (2019). The representation of sex work in the Greek Press: A corpus-assisted critical discourse analysis. *Journal of Language and Sexuality*, 8(2). <https://doi.org/10.1075/jls.18012.sag>
- Shiroda, M., Fleming, M. P., & Haudek, K. C. (2023). Ecological diversity methods improve quantitative examination of student language in short constructed responses in STEM. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.989836>
- Smith, B. E., Pacheco, M. B., & Khorosheva, M. (2021). Emergent bilingual students and digital multimodal composition: A systematic review of research in secondary classrooms. *Reading Research Quarterly*, 56(1). <https://doi.org/10.1002/rrq.298>
- Sugianto, S., & Hasby, M. A. (2023). The language variations used among speakers of Sasak Lombok, West Nusa Tenggara. *Journal of Languages and Language Teaching*, 11(1). <https://doi.org/10.33394/jollt.v11i1.6071>
- Suhardi, N. A. ., Muliati, A., Sakkir, G., & Villarama, J. A. . (2023). Increasing Students' Vocabulary Using Fairy Tales at SMKN 6 Makassar. *Journal of Language Learning and Assessment*, 1(2), 73–80. Retrieved from <https://e-journal.naifaderu.org/index.php/jlla/article/view/101>
- Sukma, B. P., Puspitasari, D. A., Afyani, S. A., Okitasari, I., Palupi, D., Kusumawardani, F., Husnul, K., & Prayoga, R. A. (2021). Cyberbullying speech patterns among Indonesian students. *Jurnal Bahasa, Sastra, Seni Dan Pengajarannya*, 49(2).
- Thao, T. Q., & Khoi, N. M. (2022). The deployment of English lexical bundles in applied linguistics research articles by Vietnamese researchers. *Ho Chi Minh City Open University*

- Journal of Science - Social Sciences*, 12(2), 75–84.  
<https://doi.org/10.46223/hcmcoujs.soci.en.12.2.2227.2022>
- Tur, A. P. A. (2019). Patterns of linguistics features in private chats of social media account leading someone to be a victim of a cyber crime. *LESIKA*, 13(1).  
<https://doi.org/10.30595/lks.v13i1.3858>
- Utami, E. W., Rokhman, F., & Mardikantoro, H. B. (2021). Language Choice in social media instagram found in @Jokowi account. *Seloka: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, 10(3).
- Vazquez-Calvo, B., & Thorne, S. L. (2022). Catalan identity and language attitudes through fan translation of video games in the digital wilds. *Journal of Language, Identity and Education*.  
<https://doi.org/10.1080/15348458.2022.2137168>
- Wahyuningsih, S. (2022). Istilah baru era pandemi COVID-19 di bahasa Indonesia dan bahasa Korea: Kajian neologisme pendekatan linguistik korpus. *JLA (Jurnal Lingua Applicata)*, 5(2), 82. <https://doi.org/10.22146/jla.70955>
- Whitley, E. A., & Schoemaker, E. (2022). On the sociopolitical configurations of digital identity principles. *Data and Policy*, 4(3). <https://doi.org/10.1017/dap.2022.30>