The CATESOL Journal

Sue-Jin Lee
California State Polytechnic University, Pomona

Analyzing the Use of AI Writing Assistants in Generating Texts with Standard American English Conventions: A Case Study of ChatGPT and Bard

The emergence of AI writing assistants has raised concerns about their potential impact on language diversity, preservation, and education. This paper examines the capabilities and limitations of AI writing assistants in generating dialectic text in response to academic and professional writing prompts. The study uses a concordance tool to conduct N-gram and keyword analyses on texts generated by AI writing assistants to examine the collocational patterns and linguistic conventions in AI-generated text productions. The results suggest that AI writing assistants rely heavily on Standard American English (SAE) conventions. Pedagogical implications include integrating language technology to promote language diversity and preservation and utilizing register-diversified corpora to enhance students' understanding of language beyond SAE. This study emphasizes the importance of critically evaluating and revising AI-generated content and contributes to a better comprehension of the potential role of AI writing assistants in academic and professional writing.

Keywords: Artificial intelligence (AI) writing assistants, Standard American English (SAE), standard language ideology (SLI), language standardization, dialects

Introduction

English is commonly known as a "lingua franca," emphasizing its global influence and dominance as a universal mode of communication. Over the past two centuries, there has been a noticeable shift toward the "Americanization" of English, increasing the standardization of English writing conventions to reflect the norms of Standard American English, or SAE (Gonçalves, Loureiro-Porto, Ramasco, & Sánchez, 2018). As artificial intelligence (AI) and writing assistants become increasingly prevalent, they often default to Standard American English (SAE) conventions and rely on databases that primarily reflect SAE norms, catering to speakers of dominant language varieties (Nee, Smith, Sheares, & Rustagi, 2022). As a result, the suggested corrections and edits offered by writing assistants like ChatGPT, Grammarly, and spellcheckers reinforce SAE ideologies by setting it as the "default" and influencing the preference for SAE writing conventions (Shuttleworth, 2011).

Although new AI tools are praised for their convenience and accuracy, the disproportionate amount of SAE reflected in these systems may result in language inequality by prioritizing certain languages and dialects and upholding linguistic biases and prejudices (Nee et al., 2022). The prevalence of SAE in AI, especially those that use natural language processing (NLP) models to generate sophisticated, human-like texts, raises concerns about language standardization and the reinforcement of linguistic biases and SAE ideologies.

This study aims to address the following research questions:

1. Do AI writing assistants adhere to Standard American English (SAE) conventions when generating dialectic texts in response to academic and professional writing prompts?
2. What are the most common collocations employed by AI writing assistants in academic and professional writing texts?

Literature Review

English as a Lingua Franca and the Standard Language Ideology (SLI)

The idea of a standard language is often associated with notions of correctness, properness, and lack of accent, as exemplified by SAE. SAE is often described as the language heard on the news and spoken by politicians, implying acceptability, education, and recognition (Lippi-Green, 2012). However, the concept of a standard language is primarily an ideology that relies on the social construction of an imagined standard. The driving force behind the standard language ideology (SLI) is the belief that language is linked to skill, education, power, and prestige. This ideology empowers certain individuals and institutions to make decisions, impose them on others, and promote specific structures and rules that make language a commodity accessible only to specific educated, powerful, and privileged communities (Lippi-Green, 2012).

SAE is arguably considered the most prominent among other standard language varieties. The widespread influence of American culture has contributed to the spread of SAE, resulting in a shift in vocabulary and spelling conventions, even within the United Kingdom (Gonçalves et al., 2018). Factors such as media, economic dominance, and cultural influence play a significant role in the spread of SAE (Gonçalves et al., 2018). Given the global digital trends favoring Americanized vocabulary and spelling, new technologies like AI writing assistants could potentially shape the trajectory of the SLI and the growing prominence of SAE.

Exploring the Benefits and Limitations of AI Writing Assistants

As AI-powered writing assistants become more prevalent in academia, recent studies have highlighted their potential benefits for students. Tools like Grammarly have been found to promote self-directed learning and self-efficacy (Cavaleri & Dianati, 2016), provide grammar checks for spelling, sentence structures, and standard grammar (Fitria, 2021), and enhance meta-linguistic knowledge (Godwin-Jones, 2022). However, it is necessary to critically examine the notion of "standard grammar" as it may reinforce the SLI and linguistic hierarchies perpetuated by digital writing assistants (Schneider, 2022).

Research on AI-powered writing assistants acknowledges their limitations and promotes them as scaffolding support and supplemental assistance while being cautious of their often-flawed recommendations (Cavaleri & Dianati, 2016; Godwin-Jones, 2022). Concerns have been raised about the accuracy of Grammarly's feedback, including invalid corrections and errors due to preset American English settings, leading to erroneous markings of other English varieties as "correctness" errors (O'Neill & Russell, 2019; Calma, Cotronei-Baird, & Chia, 2022). Thus, writing assistants' inclination towards SAE and their limited consideration for other English varieties pose a significant limitation to their usage.

Concerns with AI System Designs and the Standard Language Ideology

Although a corpus-based approach helps eliminate biases and individual writing styles, it still relies on a model of "'standard" English derived from a corpus of L1 writing. (Napolitano & Stent, 2009). Although not explicitly stated, current language ideologies, such as the SLI, position SAE as the most valuable and "normal" variety of English. This becomes problematic in the context of what Kornai (2013) refers to as "digital language death." Digital media trends exacerbate the challenges faced by minority languages as dominant languages receive preferential treatment due to their larger speaker populations and prestige, further marginalizing minority languages that are already at risk of digital language death (Kornai, 2013).

According to Kornai (2013), it is predicted that less than five percent of all languages will be able to avoid digital language death, resulting in the survival of only 16 languages. This projection suggests a significant loss of language diversity and knowledge, with dominant languages continuing to spread and grow whereas minority languages face the risk of extinction.

The Gap in Current Research

Further research is necessary to understand the impact of AI tools on promoting the SLI, particularly SAE. This research gap limits our understanding of the consequences and potential perpetuation of dominant language ideologies and the devaluation of linguistic diversity. Closing this gap can provide insights and recommendations for AI systems and educational curricula, promoting a more inclusive and equitable approach to language education.

Methods

Corpus-Based Approach

A corpus-based approach was chosen for this study due to its capacity to analyze large amounts of text and identify language usage patterns in different contexts and genres (Biber, Conrad, & Reppen, 1994). The concordance tool AntConc (Anthony, 2022) was used for corpus linguistic analysis, allowing the creation of concordances for the AI-generated text. AntConc was employed to summarize the data, compare texts and reference corpora, identify linguistic patterns, and provide a foundation for further inferential analysis. The study utilized AntConc's N-gram and keyword analysis functions to evaluate the frequency of collocations and terms in the texts for a comprehensive overview of the text data.

ChatGPT and Bard were selected for this study due to their NLP-based capabilities in understanding and generating text. Unlike Grammarly, which employs rule-based algorithms for evaluating and correcting pre-written texts, NLP models have the capacity to comprehend context and generate novel, human-like texts. Thus, the study aimed to assess the writing style and conventions of AI-generated texts, solely produced based on prompts, without human intervention.

ChatGPT and Bard[1]

OpenAI's ChatGPT and Google's Bard are conversational AI models trained to generate detailed responses. ChatGPT's GPT-3.5, released in November 2022, utilizes Reinforcement Learning from Human Feedback (RLHF) and AI trainers' input to improve response accuracy whereas Bard, released in March 2023, is powered by Google's Language Model for Dialogue Applications ("Introducing ChatGPT"; Pichai, 2023). Both Bard and ChatGPT leverage diverse training data to generate human-like responses and acknowledging potential inaccuracies and limitations. ChatGPT's training dataset is limited to information up to 2021, and Bard's dataset is narrower in scope due to its more recent release.

However, both models possess unique features such as comprehensibility, open-ended conversations, and the ability to learn from past interactions. Additionally, Bard auto-generates three distinct drafts for each prompt, providing multiple versions. For consistency, each prompt was entered into ChatGPT three times, in separate chat sessions, to generate an equal number of drafts. This approach helped avoid potential influence from previous responses and resulted in a corpus of 60 drafts, six per prompt, for analysis.

Data Collection

Ten diverse writing prompts were used to collect data, covering a range of English language variations in academic and professional contexts with a minimum of 200 words. The study employed an implicit approach to assessing the AI models' ability to recognize and produce text aligned with specific dialects, contexts, and cultures. The prompts targeted various English dialects, writing contexts, and styles, providing a comprehensive evaluation.

Table 1
Ten Prompts for AI-Generated Texts

#	Prompt
1	Write a 500-word essay discussing the importance of education in local communities where Scottish English is commonly spoken, and how it could shape students' experiences and perspectives in academic writing contexts, tailored for readers familiar with the dialect spoken in Scotland.
2	Write a 300-word news article reporting on a current event or issue affecting a specific community or cultural group within the United Kingdom, using Cockney English in a way that resonates with readers who are familiar with the dialect.
3	Write a 400-word academic paper analyzing the themes of race and identity in the works of a non-American author, such as Chinua Achebe or Arundhati Roy, for readers who speak Hawai'i Pidgin English.
4	Write a 250-word marketing pitch for a new product aimed at a specific ethnic or cultural community, highlighting how the product meets the unique needs and preferences of readers familiar with New Zealand English.
5	Write a 350-word opinion piece on how language and cultural differences can impact mental health, drawing from the experiences of people from non-English speaking backgrounds, tailored for readers who are familiar with Indian English.
6	Write a 200-word business email to a Singaporean colleague in a non-English speaking country, discussing the upcoming project deadline and any cultural considerations that may affect the timeline or communication style, with a tone and phrasing that is suitable for readers who speak Singlish.
7	Write a 450-word book review of a novel written in a non-English language, such as Gabriel Garcia Marquez's *One Hundred Years of Solitude* or Haruki Murakami's *Kafka on the Shore*, for readers familiar with Welsh English.
8	Write a 300-word personal statement outlining academic and career goals, and how diverse backgrounds and language skills in African American English can prepare a scholar for success in graduate study in applied linguistics, tailored for readers who are familiar with African American English (AAE).
9	Write a 400-word scientific paper discussing the impact of climate change on a specific non-Anglophone region, such as the Amazon rainforest or the Arctic Circle, with a style and phrasing that is suitable for readers who speak Nigerian English.
10	Write a 250-word travel blog post about a recent trip to a country where English is not the primary language, with a focus on experiences with the local culture, language, and customs, tailored for readers who speak Jamaican English.

Data Analysis

The text data collected from Bard and ChatGPT was analyzed using the concordance tool AntConc. Two reference corpora, the AmE06 Corpus of American English and the BE06 Corpus of British English, were used for comparison. These corpora represent general written American English and British English and consist of one million words each. The study aimed to determine the frequency of SAE collocations and keywords generated by the AI writing assistants. AntConc's N-gram and Keyword functions were employed to identify the 15 most common collocations and terms in the AI-generated text and compare them with those in the AmE06 and BE06 corpora. The analysis aimed to assess the AI writing assistants' adherence to SAE conventions across different dialects and writing styles.

AntConc's N-gram and Keyword functions were employed to analyze word chunks and terms in the text data. N-gram sizes of three and four were utilized to identify common collocational patterns and comprehend prevalent phrases and structures. To account for the variability of words within collocations, one placeholder word was included in each N-gram size. The keyword tool compared the specialized "target" corpus of AI-generated texts with the general "reference" corpora of American English and British

English, enabling a comprehensive examination of language usage variations across dialects and writing styles.

Results

N-Grams

In the N-gram tables, "type" represents the total number of distinct words or collocations in the dataset, whereas "frequency" indicates the number of times a word or collocation appears in a text. "Rank" denotes the relative position of keywords or collocations based on frequency, providing insights into the most common words in the corpus.

N-Gram Model of 3

The initial N-gram model was a size of three, consisting of two words and one placeholder word to form a "word chunk" of three words.

Table 2

Comparison of N-gram Size of 3 to Reference Corpora

Target (AI)			AmE06			BE06		
Type	Rank	Freq	Type	Rank	Freq	Type	Rank	Freq
the + of	1	239	the + of	1	8696	the + of	1	9027
a + of	2	99	the + and	2	2620	a + of	2	2781
the + and	3	73	a + of	3	2598	the + and	3	2598
to + the	4	70	to + the	4	2243	to + the	4	2234
of + and	5	66	the + s	5	1629	of + and	5	1424
the + is	6	57	of + and	6	1436	and + the	6	1326
is + to	7	50	and + the	7	1401	the + to	7	1168
can + to	8	49	the + in	8	1121	the + s	8	1144
to + a	9	46	the + to	9	1079	the + in	9	1126
a + and	10	43	the + that	10	1054	to + a	10	1084
and + in	11	42	to + a	11	1022	the + the	11	982
and + the	11	42	the + the	12	978	the + is	12	915
it + a	13	40	the + is	13	812	the + that	13	856
the + rainforest	13	40	and + of	14	748	the + was	14	787
non + speaking	15	38	the + was	15	738	and + of	15	753

The top six collocations in the AI corpus consistently appeared within the top 15 collocations in both the AME06 and BE06 corpora, indicating similar rankings and frequencies across all corpora. Despite the prompts requesting different dialects, these overlapping collocations align with SAE conventions, suggesting that the AI-generated text predominantly follows common usage patterns. The repeated occurrence, rankings, and frequencies of these collocations throughout all corpora underscore the AI-generated text's adherence to common usage patterns and SAE conventions. These collocations consistently rank highly in the target corpus and exhibit similar positions in the reference corpora, indicating a strong alignment with SAE conventions in terms of their frequency of usage.

N-Gram Size of 4

The subsequent N-gram model was a size of four, including three words and one placeholder word, to identify chunks of four words.

Table 3

Comparison of N-gram Size of 4 to Reference Corpora

Target (AI)			AmE06			BE06		
Type	Rank	Freq	Type	Rank	Freq	Type	Rank	Freq
the + of the	1	44	the + of the	1	2000	the + of the	1	2250
from + English speaking	2	34	in the + of	2	938	in the + of	2	792
non + speaking backgrounds	2	34	the + of a	3	538	the + of a	3	535
from non + speaking	2	34	to the + of	4	432	to the + of	4	486
Non-English + backgrounds	2	34	of the + and	5	403	of the + of	5	469
people + non-English	6	29	of the + of	6	385	at the + of	6	443
people from + English	6	29	the + and the	7	378	on the + of	7	427
it is + to	8	26	at the + of	8	366	the + and the	8	370
a + impact on	9	24	on the + of	9	361	and the + of	9	352
a + that is	10	23	and the + of	10	324	of the + and	10	349
to the + of	10	23	a + of the	11	282	for the + of	11	314
education + Scottish English	12	22	the + in the	12	278	the + in the	12	273
education in + English	12	22	of the + s	13	258	in the + and	13	264
da + in which	14	21	for the + of	14	257	with the + of	14	245
is a + of	14	21	with the + of	15	237	a + of the	15	242

The analysis of 15 collocations across the corpora revealed only one overlapping collocation, “the + of the,” which ranked first in frequency in all three corpora. The addition of a unigram significantly reduced the occurrence of overlapping collocations. Although this top-ranked collocation had the highest frequency, it was notably lower compared to the frequencies in N-gram size three collocations, indicating that as N-gram sizes increase, AI writing assistants generate more diverse texts. Collocations ranked two through six in the target corpus demonstrate a wider range of patterns, highlighting the AI’s comprehension of prompt-related text and adaptable language usage. The difference in N-gram results between sizes three and four suggests that as the text sizes increase and become more prompt-specific, there is a decrease in the reliance on specific patterns and an increase in language diversity.

Keywords

The Keyword function identified unusually frequent words or phrases in the target corpus compared to a reference corpus for the exploration of significant content, vocabulary, style, and dialects. It identifies important words in the target corpus, considering common usage patterns in SAE, deepening the understanding of unique keywords in AI-generated text.

Table 4
Keywords from the Target and Reference Corpora Based on Rankings 1- 10

Target (AI) & Reference (AmE06)				Target (AI) & Reference (BE06)			
Type	Rank	Freq_Tar	Freq_Ref	Type	Rank	Freq_Tar	Freq_Ref
English	1	205	226	English	1	205	165
mental	2	113	39	language	2	152	132
da	3	104	25	da	3	104	14
language	4	152	251	mental	4	113	150
Scottish	5	73	8	novel	5	91	68
novel	6	91	72	Scottish	6	73	47
rainforest	7	59	0	rainforest	7	59	10
help	8	125	340	help	8	125	418
is	9	529	8420	is	9	529	8995
can	10	227	1771	can	10	227	2089

The analysis of the top 10 ranked keywords reveals patterns and discrepancies between the AI-generated texts and reference corpora. Although keywords like “English” and “language” show similar frequencies in both, indicating consistent usage patterns, there are notable differences. The keyword “da” stands out in the target corpus due to its higher frequency, reflecting its recognition and representation of “the” in Hawai’i Pidgin English by ChatGPT. Other keywords such as “mental,” “Scottish,” “novel,” and “rainforest” have lower frequencies in the reference corpora, reflecting the AI’s selection of specific, low-frequency keywords based on open-ended prompts. Conversely, keywords like “is” and “can” have higher frequencies in the reference corpora compared to the AI-generated text, suggesting potential discrepancies in usage patterns. These frequency differences indicate variations in AI-generated texts’ adherence to SAE conventions and understanding of contexts and dialects, leading to disparities in keywords.

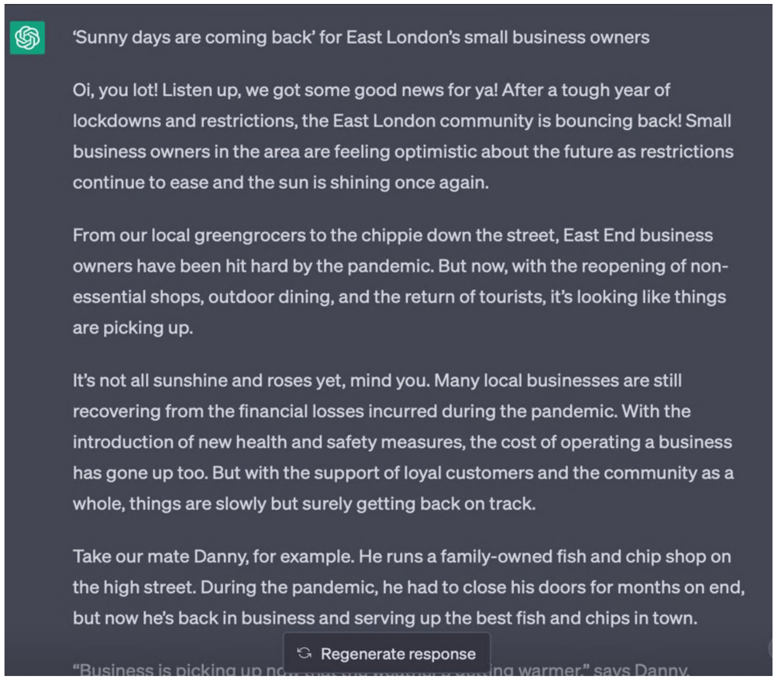
Implicit Understandings

“Implicit Understandings” encompass the underlying assumptions, biases, and nuances present in the generated texts. From the analysis of all 60 texts, three categories of implicit understandings were observed in the AI’s output: dialectical production, understanding context and tone, and biases. Dialectical production involved the AI’s attempts to generate dialect-specific speech patterns and incorporate dialect-specific keywords based on the given prompts. Understanding context and tone included the AI’s implicit comprehension of situational and emotional context and the ability to convey the intended tone. Biases were identified as unconscious biases present in the AI’s text, such as assumptions about gendered situations or stereotypical beliefs about speakers of a particular dialect.

Dialectical Production

Out of the 60 generated texts, six drafts (10%) were classified as dialectical productions. ChatGPT recognized and attempted to produce three dialects: Cockney English (see Figure 1), Hawai’i Pidgin English (see Figure 2), and Singaporean English (see Figure 3). On the other hand, Bard did not generate any of the prompted dialects in its texts.

Figure 1
ChatGPT's Production of Cockney English


'Sunny days are coming back' for East London's small business owners

Oi, you lot! Listen up, we got some good news for ya! After a tough year of lockdowns and restrictions, the East London community is bouncing back! Small business owners in the area are feeling optimistic about the future as restrictions continue to ease and the sun is shining once again.

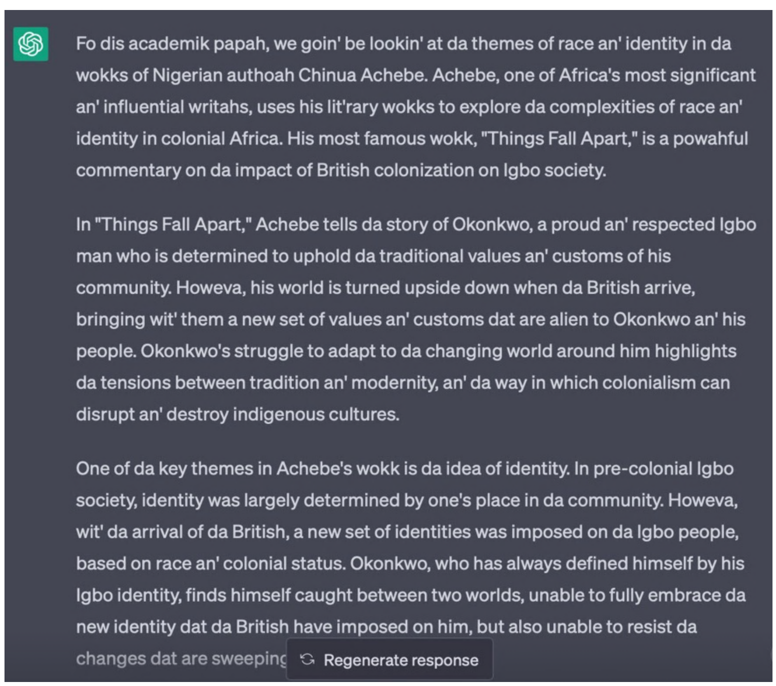
From our local greengrocers to the chippie down the street, East End business owners have been hit hard by the pandemic. But now, with the reopening of non-essential shops, outdoor dining, and the return of tourists, it's looking like things are picking up.

It's not all sunshine and roses yet, mind you. Many local businesses are still recovering from the financial losses incurred during the pandemic. With the introduction of new health and safety measures, the cost of operating a business has gone up too. But with the support of loyal customers and the community as a whole, things are slowly but surely getting back on track.

Take our mate Danny, for example. He runs a family-owned fish and chip shop on the high street. During the pandemic, he had to close his doors for months on end, but now he's back in business and serving up the best fish and chips in town.

Regenerate response

Figure 2
ChatGPT's Production of Hawai'i Pidgin English

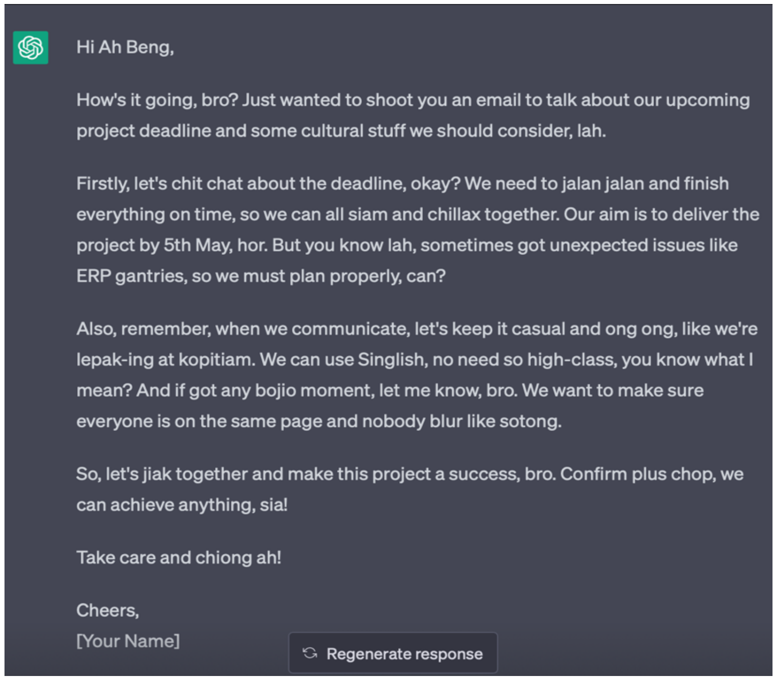

Fo dis academik papah, we goin' be lookin' at da themes of race an' identity in da wokks of Nigerian authoah Chinua Achebe. Achebe, one of Africa's most significant an' influential writahs, uses his lit'rary wokks to explore da complexities of race an' identity in colonial Africa. His most famous wokk, "Things Fall Apart," is a powahful commentary on da impact of British colonization on Igbo society.

In "Things Fall Apart," Achebe tells da story of Okonkwo, a proud an' respected Igbo man who is determined to uphold da traditional values an' customs of his community. Howeva, his world is turned upside down when da British arrive, bringing wit' them a new set of values an' customs dat are alien to Okonkwo an' his people. Okonkwo's struggle to adapt to da changing world around him highlights da tensions between tradition an' modernity, an' da way in which colonialism can disrupt an' destroy indigenous cultures.

One of da key themes in Achebe's wokk is da idea of identity. In pre-colonial Igbo society, identity was largely determined by one's place in da community. Howeva, wit' da arrival of da British, a new set of identities was imposed on da Igbo people, based on race an' colonial status. Okonkwo, who has always defined himself by his Igbo identity, finds himself caught between two worlds, unable to fully embrace da new identity dat da British have imposed on him, but also unable to resist da changes dat are sweeping

Regenerate response

Figure 3
ChatGPT's Production of Singaporean English



In the comparison of ChatGPT's text productions in the target corpus, its attempts at Cockney English, Hawai'i Pidgin English, and Singaporean English were among the closest representations of the implicitly prompted dialects. However, it is important to consider the style and register of these texts. All exhibit characteristics and terminology that are commonly associated with colloquial language, as evident in phrases like "Oi, you lot!," "Howeva,"and "no need so high-class" (OpenAI, 2023). The main focus in these texts appears to be the superficial inclusion of linguistic features specific to the dialects and contexts, or "eye dialect," which is "the phenomenon of unconventional spelling used to reproduce colloquial usage" (Brett, 2009, p. 49).

The generated texts, particularly in Hawai'i Pidgin English and Singaporean English, reflect features of eye dialect often associated with a level of denigration towards language varieties. The texts contained misspellings, such as "academik" instead of "academic," which do not reflect pronunciation differences, and included terms like "bro" and "la," which are highly colloquial and unsuitable for the genre of a business email, regardless of the dialect. The significant representations of eye dialect in nonstandard forms subtly reinforce the influence of prescriptive norms associated with the myth of a standard language, inherently undermining the voices conveyed through nonstandard orthography and the spellings themselves (Jaffe, 2000). This, in turn, overwhelmingly associates the historical context of linguistic representational practices with socially stigmatized identities (Jaffe, 2000).

Based on these text generations, it is apparent that the AI models do not discern variations in register when it comes to English dialects other than SAE or Standard British English. The attempted dialectical productions have a limited scope and should be viewed as superficial demonstrations of stylistic variations, or eye dialect, rather than precise representations of the linguistic characteristics associated with each dialect.

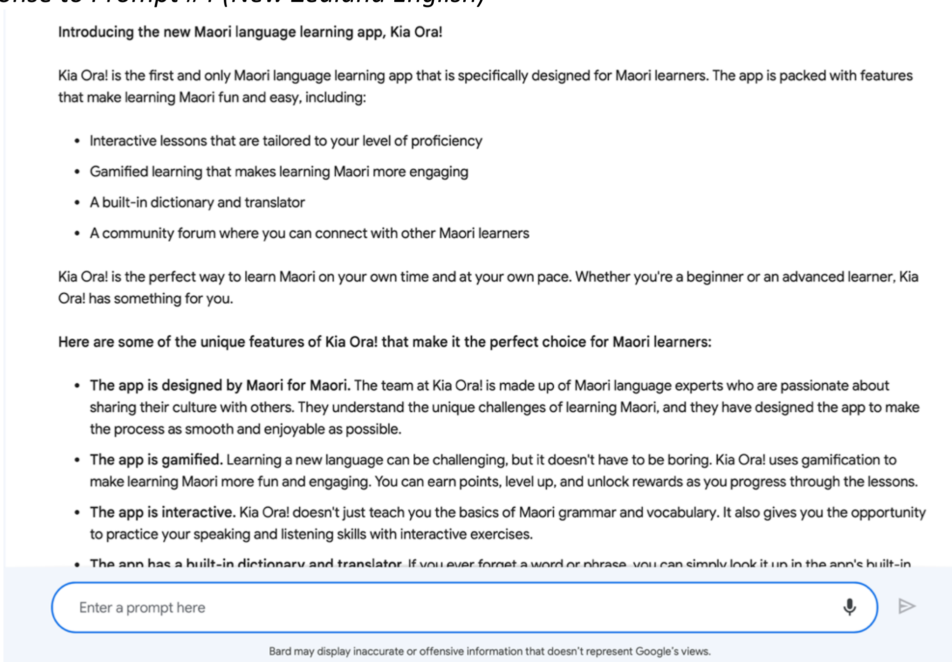
Understanding of Context and Tone

In Prompts #4 and #10, both Bard and ChatGPT demonstrated an understanding of context and conveyed appropriate tones, despite not generating the desired dialect. The marketing pitches for Prompt #4 had an informative, persuasive, and promotional tone, whereas the travel blog posts for Prompt #10 were enthusiastic, positive, and personable. Both models showed an understanding of cultural nuances and tailored their responses to target audiences.

Bard's response to Prompt #4 showcased the fictional language learning app "Kia Ora" for Māori language learners (see Figure 4). The marketing tone was evident with statements like "Whether you're a beginner or advanced learner, Kia Ora has something for you" (Google, 2023). Bard used formatting techniques like bullet points and bolding for clarity and conciseness, making their text more reader-friendly. Although not directly related to New Zealand English, Bard's response effectively catered to the target audience of Māori language learners and aligned with the marketing context of the prompt.

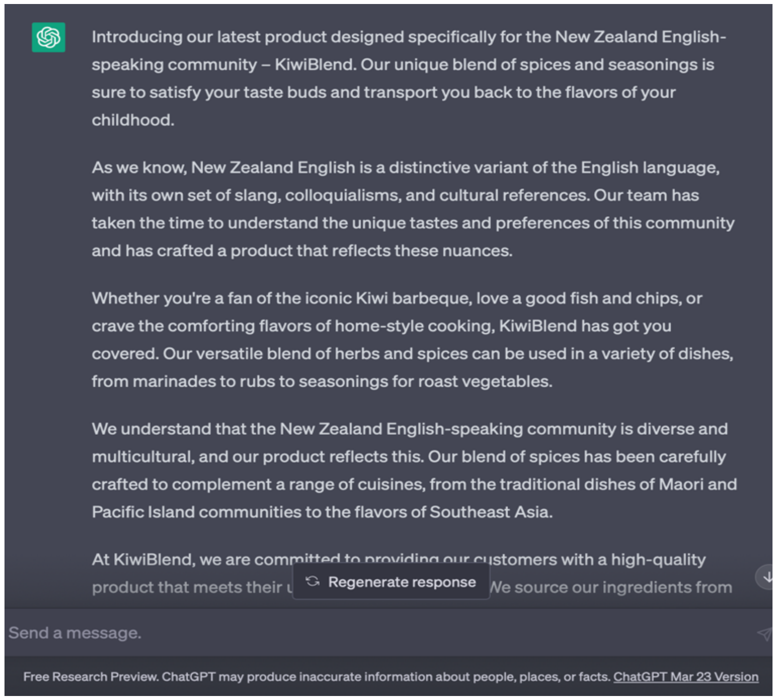
Figure 4

Bard's Response to Prompt #4 (New Zealand English)



ChatGPT's response to Prompt #4 introduced the fictional seasoning product "KiwiBlend" for a New Zealand English-speaking audience (see Figure 5)[2]. The tone was informative and promotional, highlighting features like the "unique blend of spices" and the nostalgic flavors it offers (OpenAI, 2023). ChatGPT demonstrated an understanding of the cultural nuances and preferences of the target audience, prioritizing the marketing of the product and its tailored benefits over readability.

Figure 5
ChatGPT's Response to Prompt #4 (New Zealand English)



Bard's response to Prompt #10 followed a similar approach to their response to Prompt #4, focusing on providing practical tips and utilizing formatting techniques to enhance readability and conciseness. Although the response included a relevant example related to food and scenery in Japan, it did not accurately capture the desired Jamaican English dialect. It appears that Bard may have misunderstood the prompt, writing the text from the perspective of a Jamaican English speaker instead of catering to a Jamaican English-speaking audience.

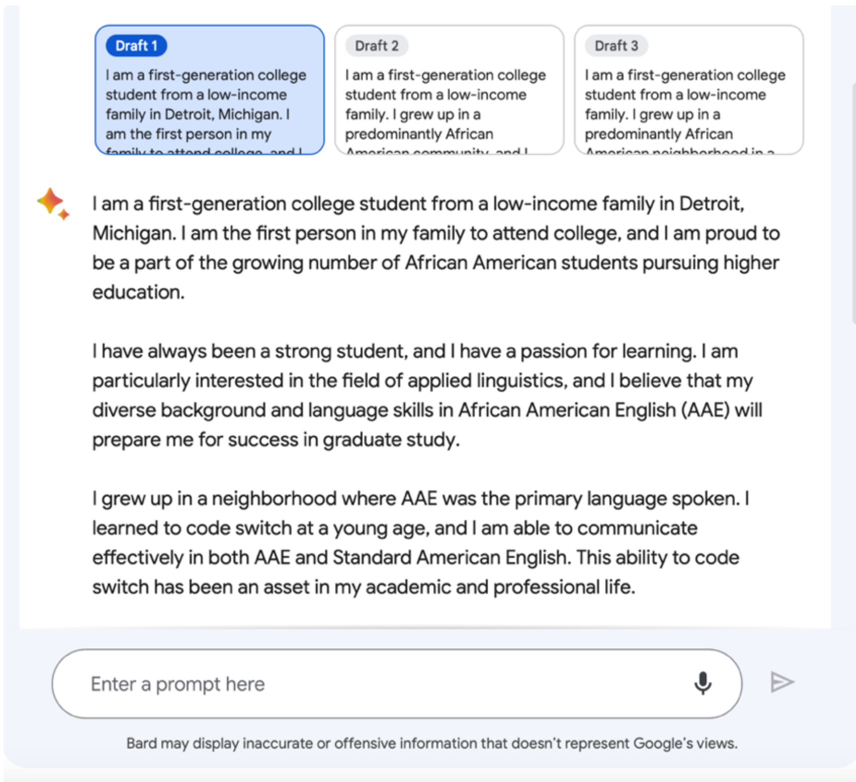
In contrast, ChatGPT's response to Prompt #10 adopted a more generalized approach, omitting specific tips or detailed information. Although it briefly mentioned the significance of learning key phrases in the destination, it did not provide further elaboration on customs or practices. This corresponds with the observations from Prompt #4, indicated that ChatGPT's responses may not be as audience oriented as Bard's. ChatGPT tends to generate broader and more comprehensive texts, whereas Bard places a higher emphasis on considering the text from the target audience's perspective and overall readability.

Bard and ChatGPT effectively conveyed the intended message and tone, showcasing the ability of NLP models to understand and adapt to various styles and contexts. Despite not achieving the desired dialect, both models demonstrated their capacity to recognize contextual nuances and accurately capture the intended tone, highlighting their strong language command and semantic understanding. This underscores the significant capability of NLP models to comprehend contextual cues and generate coherent and meaningful texts.

Biases

An unexpected finding in the study was the presence of explicit biases in the generated texts, even though the prompts only specified a particular dialect of English. Both ChatGPT and Bard exhibited biases, suggesting that even subtle prompts related to nonstandard dialects could elicit prejudiced and biased responses. This bias was particularly distinct in Bard's response to Prompt #8, which requested a "personal statement outlining academic and career goals, and how diverse backgrounds and language skills in African American English can prepare a scholar for success in graduate study in applied linguistics, tailored for readers who are familiar with African American English (AAE)." In all three versions of the response, Bard included unsolicited information about the fictional student being first-generation, low-income, and from a predominantly African American location. This production raises significant concerns as the inclusion of such information likely stems from biases and stereotypes associated with the implicitly suggested dialect, potentially originating from the AI writing assistant's text database.

Figure 4
Bard's Response to Prompt #8 (African American English)



In contrast, ChatGPT's response to Prompt #8 avoided incorporating any stereotypical details about the fictional student. ChatGPT focused solely on academic and career goals without including unsolicited information about the student's background.

Both texts followed the conventions of a professional personal statement style and expressed the desire to pursue graduate studies in the specified field. However, it is important to highlight the difference between Bard's and ChatGPT's responses. Bard included unsolicited stereotypical information about speakers of African American English, and ChatGPT's response did not. This finding suggests that Bard's biases emerged in response to the prompt involving a nonstandard English variety as the target audience. This distinction highlights the difference in how the two models handle text generations and potential biases that may arise.

This is not to say that ChatGPT's responses were more accurate or reliable than Bard's. In response to Prompt #6, which asked for a business email to a Singaporean colleague in a non-English speaking country, ChatGPT's text included a reference to "Ah Beng" (OpenAI, 2023; see Figure 3). *Ah Beng* is a colloquial term in Singaporean English that is:

> used to describe ethnic Chinese youths in Southeast Asia, particularly in Singapore and Malaysia, who have a rather loud, and/or […] terrible sense of fashion. […] Although a stereotype, *Ah Beng* refers to someone who is not highly educated, who is loud and unsophisticated, and associates with street gangs; the term also indicates a strong sense of nativeness. (Lin, 2023, p. 905)

The use of this term should be approached cautiously as its potential offensiveness depends on the context and tone of the discourse. In a business email to a colleague, ChatGPT's generated text failed to consider the professional context and used a term that could be seen as inappropriate. Moreover, ChatGPT consistently addressed male colleagues and employed gender-specific terms, giving rise to concerns regarding bias and stereotypes in AI-generated responses. All three drafts from ChatGPT in response to Prompt #6 were directed at a male colleague, utilizing "Hey bro!" and *Ah Beng* in the salutation, both of which are references to men. This suggests that ChatGPT may have also perceived the prompt as gendered, further raising concerns about bias and stereotypes in AI-generated texts.

The disparities between ChatGPT and Bard in terms of accuracy and appropriateness do not establish the superiority of one model over the other. Rather, they underscore the limitations of AI writing assistants in capturing context and cultural nuances, occasionally resulting in the generation of inappropriate and unprompted information. Both models possess the capacity to generate inaccurate and offensive content, highlighting the significance of user awareness regarding these issues. It is crucial to recognize that the generated text may present such information as factual and objective by seamlessly integrating it with a given prompt. These findings highlight the presence of inaccuracies and offensive information in the output of AI writing assistants. Each model exhibits its own predispositions towards linguistic biases, necessitating caution regarding the potential inaccuracies and offensiveness in the generated text.

Discussion and Implications

The emergence of AI writing assistants like ChatGPT and Bard raises important pedagogical considerations, given their increasing prevalence and human-like productions. Kornai (2013) emphasizes the need for collaborative language preservation efforts involving language education, policy, and technology development to support minority languages and prevent the loss of heritage, knowledge, and identity in the face of the growing digital media. In agreement with Kornai (2013), integrating language technology into existing pedagogy appears to be a practical approach to leverage its benefits for students and educators to promote language preservation and foster linguistic diversity. It is unlikely that education can avoid or prohibit the use of AI tools. These tools are not only easily accessible and user-friendly but also continuously enhance their performance through advanced training and feedback. Therefore, leveraging AI tools to help students develop critical thinking and evaluative skills is most beneficial. Meeting the expectations and requirements of academic success often necessitates adherence to SAE conventions; restricting students' exposure to and learning of SAE would be disadvantageous in academic and professional contexts. Teaching students to utilize AI texts for a better understanding of linguistic diversity and effective communication in professional and academic writing would be valuable in writing programs.

I echo Biber's (1993) recommendation of using register-diversified corpora in teaching, which acknowledges systematic linguistic variations among registers and contributes to a comprehensive understanding of language. Incorporating linguistic corpora that represent a broad range of registers is essential in writing focused classrooms, enabling a deeper understanding of language complexity across registers, genres, contexts, and dialects. Authentic and diverse texts, including nonstandard language varieties, challenge the notion that SAE is the only acceptable form of English, and exposing students to these representations could foster an inclusive perspective of language. The goal is not to avoid SAE but to broaden the understanding and comprehension of language variations by using register diversified corpora to facilitate analysis and discussion.

Conclusion

The present study sheds light on the significance of style, context, and dialectical differences in AI writing assistants, ChatGPT and Bard. These models predominantly rely on SAE conventions, as evidenced by their use of common collocations and keywords associated with SAE. N-gram analyses further confirm their adherence to specific collocational patterns and linguistic conventions of SAE. However, as the length of the generated text increases, there is a reduced reliance on specific collocational patterns, allowing for greater language usage diversity. Despite defaulting to SAE, both ChatGPT and Bard demonstrate adaptability to diverse dialects and contexts. Nevertheless, biases and inaccuracies may arise, highlighting the need for ongoing evaluation and improvement. It is essential to recognize the potential of AI writing assistants in revolutionizing academic and professional writing, offering contextually adaptable and culturally conscious texts. However, the reliance on SAE conventions may promote language standardization and marginalize speakers of nonstandard dialects and languages. AI's attempts to generate dialects often resulted in superficial representations that prioritized style and keywords over authentic dialect-specific features. Although the presence of diversity in keywords, style, and tone indicates the potential for linguistic diversity, the majority of text generations in this study did not accurately reflect the desired dialect.

The study's limitations include the use of AntConc for data analysis, which provided a general overview but limited the depth of analysis. Future research employing statistical methods could yield more precise results and identify significant differences and patterns in AI-generated text data. Additionally, writing assistants utilizing NLP models perform better with explicit prompts, potentially influencing their adherence to SAE conventions. Lastly, the focus on SAE collocations and keywords overshadowed a thorough examination of biases, stereotypes, and cultural sensitivities in AI-generated texts. A qualitative approach could provide a comprehensive understanding of biases and language ideologies in AI writing assistants, shedding light on language subordination and standardization implications.

To harness AI writing assistants' potential in academia, engaging in academic discussions and considering pedagogical implications is crucial. Although these tools could contribute to language standardization, they should be complemented with approaches that prioritize linguistic diversity and inclusivity to enable a nuanced understanding of the impact of AI tools on language use and foster an inclusive and diverse academic environment without completely avoiding the use of AI tools in academia.

Author

Sue-Jin Lee holds a master's degree in English, specializing in Teaching English to Speakers of Other Languages, from California State Polytechnic University, Pomona, and a bachelor's degree in Education, Society, and Human Development, focusing on Learning and Behavioral Studies, from the University of California, Riverside. Her research interests include corpus linguistics, AI and language, multilingual education, and language revitalization.

Note(s)

[1]*The AI tools employed in this study, ChatGPT's GPT-3.5 and Google's Bard, have undergone significant improvements since the time of writing in late 2022 to early 2023. As of early 2024, GPT-3.5 has since been upgraded to the more advanced GPT-4 and Google's Bard has been improved and rebranded as "Gemini." Both new releases boast enhanced capabilities, such as contextual understanding, increased accuracy, and*

practical real-world applications. Therefore, readers should interpret the methodology and findings of this study in the context of the capabilities of the AI tools available at the time the research was conducted.

[2] *ChatGPT's response in "New Zealand English" consistently uses the spelling "flavor," following SAE conventions. It is important to note that Standard New Zealand English prefers the spelling "flavour," indicating a different spelling preference for SAE.*

References

Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Biber, D. (1993). Using register-diversified corpora for general language studies. *Computational Linguistics, 19*(2), 219–241.

Biber, D., Conrad, S., & Reppen, R. (1994). Corpus-based approaches to issues in applied linguistics. *Applied Linguistics*, *15*(2), 169-189.

Brett, D. F. (2009). Eye dialect: Translating the untranslatable. *Annali della Facoltà di Lingue e Letterature Straniere di Sassari*, *6*, 49-62.

Calma, A., Cotronei-Baird, V., & Chia, A. (2022). Grammarly: An instructional intervention for writing enhancement in management education. *The International Journal of Management Education*, *20*(3), 100704. https://doi.org/10.1016/j.ijme.2022.100704

Cavaleri, M. R., & Dianati, S. (2016). You want me to check your grammar again? The usefulness of an online grammar checker as perceived by students. *Journal of Academic Language and Learning*, *10*(1), 223-236.

Fitria, T. N. (2021). Grammarly as AI-powered English writing assistant: Students' alternative for writing English. *Metathesis: Journal of English Language, Literature, and Teaching*, *5*(1), 65-78.

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning & Technology*, *26*(2), 5-24.

Gonçalves, B., Loureiro-Porto, L., Ramasco, J. J., & Sánchez, D. (2018). Mapping the Americanization of English in space and time. *PLoS One*, *13*(5), e0197741. https://doi.org/10.1371/journal.pone.0197741

Google. (2023). *Bard* (Beta testing) [Large language model]. https://bard.ai/

Introducing ChatGPT. (2022, November 20). OpenAI. https://openai.com/blog/chatgpt#OpenAI

Jaffe, A. (2000). Introduction: Non-standard orthography and non-standard speech. *Journal of Sociolinguistics*, *4*(4), 497-513.

Kornai, A. (2013). Digital language death. *Plos One, 8*(10), e77056. https://doi.org/10.1371/journal.pone.0077056

Lin, F. C. H. (2023). Ang Mo, Ah Beng, and Rojak: Singapore's architectural orientalism. *Journal of Asian Architecture and Building Engineering*, *22*(2), 896-913. https://doi.org/10.1080/13467581.2022.2064474

Lippi- Green, R. (2012). *English with an accent: Language, ideology, and discrimination in the United States*. 2nd Ed. Taylor & Francis.

Napolitano, D. M., & Stent, A. (2009). TechWriter: An evolving system for writing assistance for advanced learners of English. *Calico Journal*, *26*(3), 611-625.

Nee, J., Smith, G. M., Sheares, A., & Rustagi, I. (2022). Linguistic justice as a framework for designing, developing, and managing natural language processing tools. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517221090930

O'Neill, R. & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*(1), 42-56.

OpenAI. (2023). *ChatGPT* (Mar 23 ver.) [Large language model]. https://chat.openai.com/chat

Pichai, S. (2023, February 6). *An important next step on our AI journey*. Google. https://blog.google/technology/ai/bard-google-ai-search-updates/

Schneider, B. (2022). Multilingualism and AI: The regimentation of language in the age of digital capitalism. *Signs and Society*, *10*(3), 362-387.

Shuttleworth, R. (2011). Computer language settings and Canadian spellings. *TESL Canada Journal, 9*(1), 121-130.