EUROCALL 2023: CALL for all Languages
15-18 August 2023, University of Iceland, Reykjavik
DOI: https://doi.org/10.4995/EuroCALL2023.2023.16997

Writing with automatic speech recognition: Examining user's behaviours and text quality (lexical diversity)

Walcir Cardoso[a] and Danial Mehdipour-Kolour[b]
[a]Department of Education Concordia University, walcir.cardoso@concordia.ca and [b]Department of Education, Concordia University, danial.mehdipourkolour@mail.concordia.ca

How to cite: Cardoso, W.; Mehdipour-Kolour, D. (2023). Writing with automatic speech recognition: Examining user's behaviours and text quality (lexical diversity). In *CALL for all Languages - EUROCALL 2023 Short Papers.* 15-18 August 2023, University of Iceland, Reykjavik. https://doi.org/10.4995/EuroCALL2023.2023.16997

Abstract

This study explores the potential of Automatic Speech Recognition (ASR) as a writing tool by investigating user behaviours (strategies henceforth) and text quality (lexical diversity) when users engage with the technology. Thirty English second language writers dictated texts into an ASR system (Google Voice Typing) while also using optional additional input devices, such as keyboards and mice. Analysis of video recordings and field observations revealed four strategies employed by users to produce texts: use of ASR exclusively, ASR in tandem with keyboarding, ASR followed by keyboarding, and ASR followed by both keyboarding and ASR. These strategies reflected cognitive differences and text generation challenges. Text quality was operationalized through lexical diversity metrics. Results showed that ASR use in tandem with keyboarding and ASR followed by both keyboarding and ASR yielded greater lexical diversity, whereas the use of ASR exclusively or ASR followed by keyboarding had lower diversity. Findings suggest that the integrated use of ASR and keyboarding activates dual channels, thus dispersing cognitive load and possibly improving text quality (i.e. lexical diversity). This exploratory study demonstrates potential for ASR as a complementary writing tool and lays groundwork for further research on the strategic integration of ASR and keyboarding to improve the quality of written texts.

Keywords: *Automatic Speech Recognition (ASR), writing, user's behaviour, writing strategies, lexical diversity.*

1. Introduction

This study explores a breakthrough in the development of writing: the use of one's voice, made possible via Automatic Speech Recognition (ASR). ASR is an application found in most modern devices (e.g. computers, smartphones) as well as in word-processing and communication software, such as Google Docs. ASR converts speech into text in real-time: via a microphone, the software receives and identifies the words a person speaks, analyzes it using a set of algorithms, and finally produces an output in the form of a text. In a survey conducted by Enge (2020), he found that 45% of U.S. respondents use their voices to input text and for commands daily, indicating that people are becoming increasingly accustomed to speaking to their devices, particularly after the rise of voice-activated virtual assistants, such as Amazon Alexa and Apple Siri.

Despite the growth in ASR use for writing, research in this field has not received the attention it deserves. While studies indicate that ASR is a useful pedagogical tool for improving the pronunciation of second/foreign language

(L2) learners (e.g. Cardoso, 2022; McCrocklin, 2018; Mehdipour-Kolour & Cardoso, in press), very few studies have fully explored the use of ASR as a writing tool. Notable exceptions include Haug and Klein (2018), who compared handwritten and ASR-created argumentative texts among elementary school students, and Johnson and Cardoso (in press) and Mehdipour-Kolour and Cardoso (2023), who examined users' perceptions of ASR as a writing method. Indeed, most of the research available on ASR for writing involves users with learning difficulties (e.g. Ballard et al., 2019; Le et al., 2018; Quinlan, 2004).

As research on using ASR for composing texts is still emerging, this study aims to build foundational knowledge by examining the technology's feasibility as a writing tool (for the rationale, see Cardoso, 2022). Rather than focusing on user perceptions, which are explored in recent related work (Johnson & Cardoso, in press; Mehdipour-Kolour & Cardoso, 2023), this research investigates the core capabilities of an accessible ASR system (Google Voice Typing) by analyzing: (1) writers' behaviours (a component of *learning strategy*, defined as the behaviors or steps that learners use to improve their learning; Oxford, 1990), and (2) the correlations of behaviour with text quality. In this study, text quality is operationalized via lexical diversity (Yu, 2010; Kyle et al., 2021), a measure of how many different words are used and the relative frequency of each word. This measure is regarded as an important indicator of writing knowledge (Yu, 2010) and text quality (Laufer & Nation, 1995).

This study explores the potential of ASR as a writing tool by examining the behaviours of users while engaged in ASR-based writing and the quality of the texts produced (operationalized as *lexical diversity*). The study was guided by the following Research Questions (RQs):

1. How do writers interact with the ASR to compose texts? Specifically, what strategies do they use to compose texts (e.g. exclusive use of ASR, combining keyboarding with ASR)?
2. Is there a connection between the strategy selected and the quality of the texts produced (i.e. in terms of lexical diversity)?

2. Method

A total of 30 adult participants were recruited to participate in this study. They were English L2 speakers and had a variety of first languages (e.g. Chinese, Farsi, French, Portuguese) and had no previous experience using ASR for composing narrative texts.

The experiment took place remotely using Zoom and lasted approximately two hours. Participants first received an introduction to Google Voice Typing (GVT)(accessed by navigating to *Tools* and then *Voice Typing* in Google Docs) and its writing capabilities, including verbal cues such as *comma* and *period* for punctuation, and *new paragraph* and *underline last word* for formatting text. They were then instructed to compose texts by dictating into GVT while using any additional resources available to them, such as a keyboard, mouse, or handwriting. There were no restrictions on strategies or tools that participants could leverage to plan, monitor recognition, correct errors, and revise content. This enabled the observation of natural behaviors when using ASR for composing texts. The collected texts had an average length of about 300 words, with the mean of approximately 105.47 unique words across the 30 texts. Following insights from Holdsworth (2021) for a holistic analysis of users' experience and behaviours and thick descriptions (Geertz, 2008), data collection was triangulated to include: observation, fieldnotes, conversations with the participants, audio-visual materials (video recordings and screen capture of interactions with ASR), and an analysis of the texts produced to analyze lexical density.

To address the goals of the study, participants were given three narrative writing prompts (e.g. *First Love: Write a story about how you met your first love ...* – simplified due to space limitations) and asked to choose two to compose their narratives. At the end of the experiment, all texts were compiled for lexical diversity analysis. For this analysis, texts were imported to the Tool for the Automatic Analysis of Lexical Diversity (TAALED) (Kyle et al., 2021), an analysis tool developed to compute a wide range of lexical diversity indices. Four indices were selected (for the rationale and details, see Kyle et al., 2021): Root Type-Token Ratio, Moving-Average TTR, Measure of Textual Lexical Diversity, and MTLD moving-average-wrap.

3. Results

Adopting an exploratory approach**,** this study analyzed video recordings and field notes to answer the question of how writers interact with ASR to produce their texts regarding strategies used (RQ1). As illustrated in Table 1, participants employed a variety of strategies, broadly categorized into four types: the use of (1) ASR exclusively (ASR-Only; *n*=2; or 7% of all participants), (2) ASR in tandem with keyboarding (ASR=KB; *n*=13; 40%), (3) ASR followed by keyboarding (ASR>KB; *n*=5; 17%), and (4) ASR followed by both keyboarding and ASR (ASR>ASR=KB; *n*=10; 33%).

To answer RQ2 on whether there is a correlation between the ASR writers' choice of strategies and the lexical diversity of their written text, Eta and Eta squared tests were conducted. These tests are suited for analyzing categorical data and measuring effect size, and effective for exploring the correlation between strategy use and lexical diversity. We calculated the lexical diversity of participants' texts using TAALED (Kyle et al., 2021). Five levels of lexical diversity were identified in the collected narratives. The levels 'below average,' 'slightly below average,' 'average,' 'slightly above average,' and 'above average' were assigned based on predefined scores ranging from 'below average' (<50) to 'above average' (>50), as shown in Table 1.

Table 1. Writing strategies and associations with lexical diversity

			Lexical Diversity in Texts				
			Below	Slightly below	Average	Slightly above	Above
Writing Strategies	ASR-Only	*n*=2	2	0	0	0	0
	ASR=KB	*n*=13	1	1	0	7	4
	ASR>KB	*n*=5	3	1	0	1	0
	ASR>ASR=KB	*n*=10	0	1	1	4	4
Total		*n*=30	6	3	1	12	8

The findings of the Eta and Eta-squared tests are illustrated in Table 2, indicating that using ASR-Only corresponds to a significant moderate negative effect on lexical diversity, while ASR=KB exhibits a moderate positive association, but it lacks statistical significance in its effect. The adoption of ASR>KB is significantly and strongly correlated with a low level of lexical diversity. Finally, the use of ASR>ASR=KB demonstrates a moderate positive relationship with high lexical diversity, albeit without a statistically significant effect.

Table 2. Summary of the measures of associations and their effect sizes

Writing Strategies	*F*	*p*	η	η^2
ASR-Only	6.80	.014	.44	.19
ASR=KB	1.77	.194	.24	.05
ASR>KB	8.99	.006	.49	.24
ASR>ASR=KB	3.68	.065	.34	.11

4. Discussion and conclusions

This study examined L2 writers' interactions with ASR during text composition, specifically focusing on the strategies employed (RQ1), and explored whether a connection exists between the chosen strategy and the resultant quality of the produced texts (RQ2).

The findings for RQ1 indicate that L2 writers use a variety of strategies when composing with ASR technology. Four distinct approaches were identified through an analysis of participant behaviours: (1) the use of ASR exclusively, (2) ASR in tandem with keyboarding, (3) ASR followed by keyboarding, and (4) ASR followed by both keyboarding and ASR. The diverse array of strategies observed implies that writers exhibit distinct writing and, potentially, learning preferences. These preferences can be attributed to Individual Differences (IDs), which encompass a spectrum of factors such as varying cognitive aptitudes and learning styles (e.g., visual, auditory). Dörnyei (2005) defines as the distinctive traits and attributes that set individuals apart from one another. For example, considering the participants who used ASR exclusively for composing their texts, it is possible to speculate that these participants have heightened auditory and oral abilities in comparison with their peers. This ability led them to rely on ASR throughout all phases of the writing process, including planning, drafting, revising, and editing (for a discussion of these types of learners in SLA, see Pawlak & Kruk, 2022).

Regarding RQ2, Eta (η) and Eta-squared (η2) tests were conducted to investigate the connection between the writing strategies employed and the lexical diversity of the collected texts. The findings indicate that the use of ASR *combined* with keyboarding, namely ASR=KB and ASR>ASR=KB, resulted in a higher level of lexical diversity, as compared to the two other strategies. The latter strategies, which did not combine ASR and keyboarding, were significantly correlated with lower levels of lexical diversity as a measure of text quality. This finding confirms the Modality Effect Theory proposed by Leahy and Sweller (2011), who hypothesize that using several input modalities (e.g. visual and auditory channels) might assist in evenly distributing the cognitive burden, consequently yielding enhanced writing performance. Incorporating these strategies, the participants could cultivate ideas and reduce the need for constant text monitoring and repetitive corrections compared to exclusively using ASR, resulting in texts with higher lexical diversity. Considering that the results favouring ASR=KB and ASR>ASR=KB as writing strategies were not statistically significant due to the small sample size, cautious generalizations can still be made. One such generalization is that our findings highlight the potential benefits of leveraging ASR as a *complementary* rather than a standalone tool in the writing process.

As an exploratory study that aimed to provide initial insights into L2 writer strategies and lexical diversity patterns in ASR-based writing, there are certain limitations that need be acknowledged. These include the small sample size and the short duration of the study. Future research should examine these effects over an extended period of time, with more participants, and using additional textual measures beyond lexical diversity. However, the findings lay important groundwork for understanding how students strategically use ASR when writing, and how to potentially incorporate ASR effectively alongside typing for enriching text production with lexical variety. This has relevant implications for writing instruction and the use of ASR in L2 education.

Acknowledgements

We sincerely thank all the participants who generously contributed their time and insights to make this research possible. We also acknowledge partial funding from the *Social Sciences and Humanities Research Council of Canada* (Grant #430-2022-00512).

References

Ballard, K., Etter, N. M., Shen, S., Monroe, P., & Tan, C. T. (2019). Feasibility of automatic speech recognition for providing feedback during tablet-based treatment for Apraxia of Speech Plus Aphasia. *American Journal of Speech-Language Pathology, 28*, 818–834. https://doi.org/10.1044/2018_AJSLP-MSC18-18-0109

Cardoso, W. (2022). Technology for speaking development. In T.M. Derwing., M.J. Murray., & R.I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 299–313). Routledge. https://doi.org/10.4324/9781003022497

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge.

Enge, E. (2020). Mobile Voice Usage Trends in 2020. *Perficient, Inc*. https://www.perficient.com/insights/research-hub/voice-usage-trends

Geertz, C. (2008). Thick description: Toward an interpretive theory of culture. In T. Oakes, & P. L. Price (Eds.), *The cultural geography reader* (pp. 41–51). Routledge.

Haug, K.N., & Klein, P.D. (2018). The effect of speech-to-text technology on learning a writing strategy. *Reading & Writing Quarterly, 34*(1), 47–62. https://doi.org/10.1080/10573569.2017.1326014

Holdsworth, J. (2021). *Doing ethnographic research: a practical guide*. Sage Publications.

Johnson, C. & Cardoso, W. (in press). Hey Google, let's write: Examining L2 learners' acceptance of automatic speech recognition as a writing tool. *CALICO Journal*.

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Le, D., Licata, K., & Provost, E.M. (2018). Automatic quantitative analysis of spontaneous aphasic speech. *Speech Communication, 100*, 1–12. https://doi.org/10.1016/j.specom.2018.04.001

Mehdipour-Kolour, D. & Cardoso, W. (in press). A systematic literature review of automatic speech recognition in L2 learning: A case for L2 writing. In M. Peterson & N. Jabbari (Eds.), *Routledge Frontiers in computer assisted language learning* (17 pages). London, UK: Routledge. https://doi.org/10.4324/9781003395218

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Newbury House Publishers. https://escholarship.org/uc/item/1446j36q

Pawlak, M., & Kruk, M. (2022). *Individual differences in computer-assisted language learning research*. https://doi.org/10.4324/9781003240051

Quinlan, T. (2004). Speech recognition technology and students with writing difficulties: improving fluency. *Journal of Educational Psychology, 96*(2), 337–346. https://doi.org/10.1037/0022-0663.96.2.337

McCrocklin, S. (2018). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal, 36*(2), 119–137. https://doi.org/10.1558/cj.34738

Yu, G. (2010). Lexical diversity in writing and speaking task performances. *Applied Linguistics, 31*(2), 236–259. https://doi.org/10.1093/applin/amp024