


ORIGINAL ARTICLE

What contributes to fluent L2 speech? Examining cognitive and utterance fluency link with underlying L2 collocational processing speed and accuracy

Kotaro Takizawa 

School of Education, Waseda University, Tokyo, Japan
Email: kzmw0628@fuji.waseda.jp

(Received 2 July 2023; revised 2 April 2024; accepted 2 April 2024; first published online 13 May 2024)

Abstract

Second language (L2) fluency research has suggested that a range of linguistic knowledge and processing speed serve as cognitive fluency (CF) underlying L2 utterance fluency (UF). Building on prior CF-UF link studies, this study explored L2 collocational processing speed and accuracy as underlying CF measures. This study also explored the phrasal frequency effect on the relationship between collocational processing speed and UF to test whether the processing speed of higher-frequency collocations is more strongly related to UF. A total of 108 Japanese university students completed a phrasal acceptability judgment task with adjective–noun collocations pooled from the Academic Collocation List. Speech was elicited in an argumentative speech task, and UF measures were computed based on speed, breakdown, and repair fluency. The results revealed that the collocational processing speed was weakly tied to articulation rate and mid-clause and end-clause silent pause ratio ($r_{ho} = |.271-.322|$), while the collocational processing accuracy was weakly to moderately tied to most UF measures except filled pause ratio ($r_{ho} = |.281-.368|$). The collocation frequency moderated the relationship between the processing speed and UF, suggesting that high-frequency collocations served as a proxy for automaticity in speech production. Implications for CF-UF link research are discussed.

Keywords: Cognitive fluency; collocational processing; frequency effect; multiword sequence; usage-based; utterance fluency

Introduction

Achieving oral fluency is a major goal of second language (L2) learning. Research shows that oral fluency is a crucial determinant of speaking proficiency (Iwashita et al., 2008). Moreover, fluency is often perceived as an indicator of overall oral proficiency in an L2 (Tavakoli & Hunter, 2018) or equated with nativelikeness (Boers et al., 2006). According to Segalowitz (2010), oral fluency comprises three

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

subdimensions: utterance, cognitive, and perceived fluency, with the first two being relevant in the current study. Utterance fluency (UF) refers to the actual properties of speech samples including the “temporal, hesitation, pausing, and repair characteristics” (Segalowitz, 2010, p.48). Cognitive fluency (CF) refers to “underlying cognitive processes responsible for fluency-relevant features of utterances” (Segalowitz, 2010, p. 50). Recently, L2 fluency research has highlighted that L2-specific CF—L2 linguistic knowledge and processing speed—underlies L2 UF (e.g., De Jong et al., 2013). The studies defined L2-specific CF according to Levelt’s (1989) speech production model and measured and estimated grammatical, lexical, and phonological knowledge and processing speed through various linguistic tasks (e.g., picture naming task). The CF-UF link revealed that the role of lexical knowledge and lexical retrieval speed is relatively large in predicting L2 UF.

Among various factors contributing to L2 UF, it is widely known that multiword sequences (MWS), defined as recurrent strings of words, enhance fluent L2 speech (Pawley & Syder, 1983; Tavakoli & Uchihara, 2019). The rationale behind the relationship between MWS and L2 UF is that MWS tend to be stored and retrieved as a whole from long-term memory, reducing processing costs in linguistic encoding and freeing up attentional resources in speech production (Wray, 2002; Kormos, 2006). One limitation in previous studies examining the relationship between MWS and L2 UF is that MWS were only examined at the performance level—MWS employed spontaneously in actual language use. However, closely aligned with Segalowitz’s (2010, 2016) definition of CF and the way CF was operationalized in the existing literature on the CF-UF link, it seems that MWS are better represented as part of lexical resources, or more specifically, phraseological competence, which can be assessed and estimated through linguistic tasks. The current study operationalized MWS as adjective–noun collocations, defined as the combinations of two adjacent words which co-occur more frequently than would be expected by chance alone and demonstrate statistical strength of co-occurrence by mutual information (MI) or t-score in a given corpus (e.g., Gablasova et al., 2017).

This study examined (1) how processing speed and accuracy of adjective–noun collocations (as L2-specific CF) were related to L2 UF and (2) how collocation frequency modulated the relationship between collocational processing speed and L2 UF. Usage-based accounts of language acquisition assume that human cognition is sensitive to frequency information (Ellis, 2002). A growing number of studies have explored the effects of phrasal frequency on collocational processing, revealing that high-frequency collocations are processed faster than low-frequency ones (e.g., Siyanova-Chanturia et al., 2011). Given that higher-frequency collocations are learned better and accordingly processed faster, it is expected that higher-frequency collocations would be activated and retrieved faster and more efficiently in speech production, which would lead to fluent speech.

Literature review

Utterance fluency

There exist various ways to operationalize UF measures, depending on the research aims or questions at hand (e.g., Segalowitz et al., 2017). The most commonly utilized UF measures originate from the triadic framework of UF, which encompasses speed,

breakdown, and repair fluency (Skehan, 2003; Tavakoli & Skehan, 2005). Speed fluency refers to the overall speed of articulation, which is commonly measured by articulation rate (AR) or mean syllable duration, serving as an indicator of overall cognitive efficiency in speech production. Breakdown fluency refers to pausing behavior, including silent and filled pauses (e.g., *um* or *eh*). Recent research examining the multifaceted nature of pausing behavior suggests that pauses are measured according to their frequency, duration, and location (Bosker et al., 2013; Saito et al., 2018; S. Suzuki & Kormos, 2019). Notably, pauses occurring within a clause and those occurring between clauses reflect distinct cognitive processes (e.g., Saito et al., 2018). Finally, repair fluency refers to disfluency phenomena that include self-correction, self-repetition, and false starts.

L2 fluency research over the past decade has shown that overall, speed, breakdown, and repair fluency are characteristics that help us better understand how L2 fluency develops and is perceived by others. Saito et al. (2018) found that AR and mid-clause pauses clearly distinguished separate levels of perceived fluency (low, mid, high, and native). More recently, Tavakoli et al. (2020) showed that speed fluency measures consistently distinguished between lowest to upper-intermediate levels, whereas pausing measures, particularly the frequency of mid-clause pauses, distinguished low-to-mid levels from mid-to-high levels. In addition, Kahng (2014), Tavakoli (2011), and De Jong (2016) revealed that L2 speakers paused more in the middle of a clause than first language (L1) speakers did. These findings lend themselves to the understanding that speed and pause measures (especially mid-clause pauses) are relatively stable characteristics distinguishing the level of proficiency or fluency. Recent meta-analysis further supports the importance of mid-clause pauses for listener-based judgment of fluency (S. Suzuki et al., 2021). Despite the good explanatory power of speed and pause measures for L2 fluency levels, repair measures are known to be subject to individual variation, and some researchers even assume that repairs reflect language-general characteristics including personal speaking style (De Jong et al., 2015).

Cognitive processes underlying utterance fluency

UF reflects the underlying cognitive processes in speech production. Levelt's (1989) blueprint of the L1 speaker's speech production model assumes the three major stages in speech production: conceptualization, formulation, and articulation. In conceptualization, macroplanning and microplanning offer conceptual preparation (e.g., communicative intentions) referring to the speaker's knowledge of the external world, which forms a preverbal message. Next, during formulation, the preverbal message is shaped linguistically by lexical, grammatical, and phonological encodings. Lemmas and lexemes are activated in the mental lexicon and shaped with morphological and phonological information. Finally, in the articulation stage, articulatory gestures are encoded, and language is produced as overt speech. The produced speech is then self-perceived and monitored, the phase known as self-monitoring. L2 fluency research has shown that speed fluency is indicative of overall processing efficiency in speech production, while breakdown fluency—typically measured as mid- and end-clause pauses—reflects formulation and conceptualization stages, respectively (De Jong, 2016; Skehan et al., 2016; Lambert et al., 2017,

2021; Kahng, 2018; Saito et al., 2018). Repair fluency reflects the availability of attentional resources and may be related to self-monitoring (Kormos, 1999; Lambert et al., 2021). Levelt's speech production model assumes that L1 speech processing in each stage is almost parallel, but L2 learners, particularly lower-level learners, tend to slow down in the linguistic encoding stage due to non-automatized linguistic knowledge, which leads to nonparallel, serial processing (Kormos, 2006).

In recent years, researchers have investigated underlying L2 linguistic resources and processing speed as cognitive fluency (CF) in relation to L2 UF (De Jong et al., 2013; Kahng, 2014, 2020; S. Suzuki & Kormos, 2022). Segalowitz (2016) conceptualizes L2 CF as "the fluid operation (speed, efficiency) of the cognitive processes responsible for performing L2 speech acts" (p. 82). This sense of L2 CF can be of two types: (1) general-purpose cognitive control processes regulating mental activities and behaviors including L1 and (2) L2-specific modes of cognition that underlie L2 UF (Segalowitz, 2016). The current study defined L2 CF in the latter sense, particularly targeting L2 collocational processing speed and accuracy as part of the underlying L2 cognitive efficiencies responsible for UF. Recently, several scholars have attempted to systematically examine the link between L2-specific CF and L2 UF (De Jong et al., 2013; Kahng, 2020; S. Suzuki & Kormos, 2022). Those studies operationalized L2-specific CF broadly as L2 linguistic resources and processing speed, including vocabulary knowledge, lexical access speed, syntactic building speed, syntactic processing speed, and articulatory duration and latency.

De Jong et al. (2013) used various linguistic tests to measure L2-specific CF reflecting linguistic encodings, including vocabulary (including collocation) and grammatical knowledge, lexical retrieval speed, pronunciation quality and duration, sentence building speed, and articulation latency. The correlation analyses indicated that vocabulary knowledge and sentence building speed were significantly correlated with mean syllable duration with large effect sizes ($r = -.58$ and $-.66$), and the regression model showed that all CF measures explained 50% variations in mean syllable duration. Overall, vocabulary knowledge and sentence building speed were relatively stronger indicators of L2-specific CF. More recently, Kahng (2020) demonstrated similar but more nuanced results for the CF-UF link, additionally considering the confounding effect of L1 UF on the CF-UF link. Kahng similarly used various linguistic tests to measure L2-specific CF as in De Jong et al. (2013), but Kahng included a phrasal vocabulary size test as a measure of MWS knowledge. L2 UF measures were computed based on speed, breakdown, and repair fluency, but what differed from those in De Jong et al. (2013) was the consideration of pause location (within or between clauses). The multiple regression analyses highlighted the relative importance of lexical retrieval speed, syntactic encoding speed, and phrasal vocabulary knowledge. Particularly, L2 lexical retrieval speed and L2 phrasal vocabulary explained a 40% variance in L2 mean syllable duration, indicating the importance of word-related CF in line with De Jong et al. (2013). Suzuki and Kormos (2022) employed structural equation modeling (SEM) to model CF-UF links across four task types. The CF measures included vocabulary knowledge, lexical retrieval speed, syntactic encoding speed, grammatical knowledge, and articulation speed. To measure grammatical knowledge and syntactic encoding speed, they administered a grammaticality judgment task and a maze task and computed both reaction time (RT) and accuracy scores. There

emerged two latent structures of CF: linguistic resource (vocabulary knowledge and grammatical accuracy) and processing speed (RT measures and articulation speed). The structural regression model showed that speed fluency was associated with processing speed ($\beta = .431-.609$) while breakdown fluency was associated with both linguistic resource ($\beta = .240-.345$) and processing speed ($\beta = .376-.501$). Repair fluency was only associated with a linguistic resource in the limited task types ($\beta = .330-.375$). The regression coefficient of CF observed variables indicated that vocabulary knowledge and syntactic encoding speed had very strong effect sizes across all the task types ($\beta = .85-.88$ and $.79-.82$).

An interim summary of studies investigating the link between CF and UF suggests that vocabulary knowledge and retrieval speed, including both single word and collocations, and syntactic processing speed play a more prominent role in predicting speed and breakdown fluency, while their impact on repair fluency is comparatively less pronounced. These findings confirm the notion that speech production is largely lexically driven (Levelt, 1989; Kormos, 2006). However, despite the acknowledged importance of lexical resources to UF, the role of collocations as an underlying CF remains relatively underexplored. While Kahng (2020) has shed some light on the contribution of phrasal vocabulary size to UF, De Jong et al. (2013) included productive collocation knowledge as part of vocabulary knowledge, thereby obscuring the unique contribution of collocation knowledge to UF. In this context, two studies (Koizumi & In'nami, 2013; Uchihara et al., 2021) merit closer review since they estimated learners' collocation knowledge in association with their L2 UF performance, albeit without explicitly framing their studies within the CF-UF link framework. Koizumi and In'nami (2013) tested the predictive strength of vocabulary size, depth, and retrieval speed to L2 UF using SEM. Collocation knowledge was measured in form-recall (study 1) and form-recognition (study 2) formats. UF measures included speed fluency (the number of pruned tokens per minute) and repair fluency (e.g., the number of disfluency markers). The SEM model demonstrated that the R^2 value was very small for speed fluency (.17) and repair fluency (.07) in study 1 and small to moderate for speed fluency (.46) and repair fluency (.11) in study 2. Uchihara et al. (2021) measured collocation knowledge within a word-association paradigm for 40 Japanese university students. Frequency and association strength (MI and t-score) between the target word and the first and all responses were computed. L2 UF measures included AR and silent/filled pause ratio (FPR). The results showed moderate correlations for AR ($\rho = -.312$) and silent pause ratio ($\rho = .329$) for the first response analysis and FPR ($\rho = -.325$) for all response analyses.

A synthesis of the previous literature reviewed above uncovers the crucial role of collocation knowledge in predicting L2 UF. However, a notable gap in these studies is the lack of examination of the processing speed of collocations, which aligns with Segalowitz's (2016) initial narrow conceptualization of L2-specific CF. To address this gap, the current study sought to examine collocational processing speed as a measure of CF through a phrasal acceptability judgment task (PJT). In addition, this study examined the accuracy of collocational processing as another measure of CF (see S. Suzuki & Kormos, 2022, for accuracy measures elicited from timed linguistic tasks).

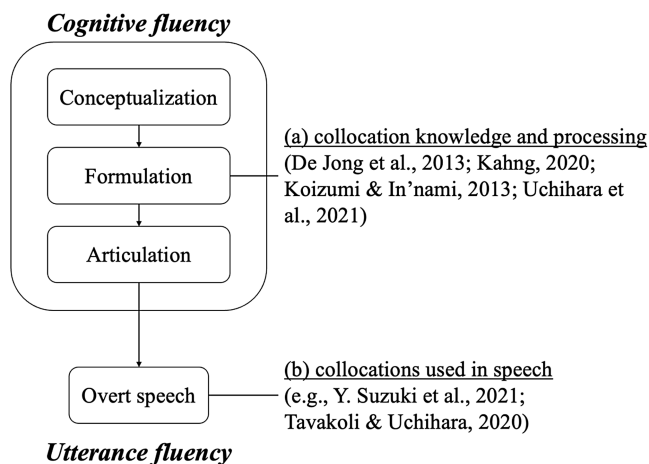


Figure 1. Schematic representation for the role of MWS relevant to UF.

Phraseological competence and oral fluency

Some scholars argue that collocations could be represented holistically in the mental lexicon (Wray, 2002; Kormos, 2006). This means that people do not necessarily construct phrases word by word but retrieve an entire phrase from long-term memory (for a discussion on the holistic processing of collocations, see Siyanova-Chanturia, 2015b). In this regard, collocation pertains to efficient speech production, automaticity, and fluency by attenuating cognitive processing load. Speech production is a complex process that requires significant attentional resources for conceptualization, linguistic encoding, articulation, and monitoring (Levelt, 1989; Kormos, 2006). In particular, L2 learners tend to face difficulties in linguistic encodings due to limited linguistic knowledge and processing speed, which takes much attentional resources otherwise spent on other tasks. Collocations potentially free up attentional resources by circumventing intricate linguistic encodings, which allows learners to achieve automaticity and fluency (Wray, 2002; Kormos, 2006).

Theoretical understandings of how collocations enhance UF are represented in Figure 1. One approach to examine the role of collocations in UF is represented as (a) in Figure 1, where collocation knowledge and processing speed as CF contribute to UF. The other approach is represented as (b), where the actual production of collocations in speech enhances UF. To date, the vast majority of empirical studies have analyzed the relationship between collocations and UF elicited during various speaking tasks (McGuire & Larson-Hall, 2017, 2021; Nergis, 2021; Y. Wood, 2006, 2009; Thomson, 2017; Tavakoli & Uchihara, 2019; Suzuki et al., 2022). These studies provide valuable insights into the ways in which instruction or training can develop phraseological competence, leading to improvements in UF over time. Furthermore, they demonstrate how the use of sophisticated collocations is linked to assessed levels of speaking proficiency (see Supplementary information [Appendix S4](#) for the top 20 most frequent bigrams in learner texts in the current study). However, a possible limitation of this approach (i.e., (b) in Figure 1) would be that produced

collocations can be affected by and restricted to a specific task or topic. This makes it hard to estimate the underlying phraseological competence contributing to UF. Employing linguistic tests external to speech tasks allows estimation of collocation knowledge and processing which contributes to UF, as empirically attested in De Jong et al. (2013) and Kahng (2020) (but see Yan, 2020 using predefined collocation as a prompt in an elicited imitation test).

Collocational processing as a window into usage-based language learning

A plethora of research has demonstrated that collocations are processed more quickly than their matched novel phrases. Moreover, higher-frequency collocations are processed at a faster rate than lower-frequency ones (for an overview, see Conklin & Schmitt, 2012; Siyanova-Chanturia & Martinez, 2015). These findings shed light on our remarkable sensitivity to recurrent lexical patterns, phrase frequency, and probabilistic nature of word co-occurrence, all of which align with the usage-based theory of language learning and processing (N. C. Ellis, 2002, 2012).

Factors that affect collocation processing include phrase frequency (e.g., Siyanova-Chanturia et al., 2011), constituent word frequency (e.g., Öksüz et al., 2020), congruency (e.g., Wolter & Yamashita, 2018), statistical associations (e.g., MI) (e.g., N. C. Ellis et al., 2008), learner's proficiency and learning experiences (Siyanova-Chanturia et al., 2011), and cognitive aptitude (Yi, 2018). Although the types of MWS (e.g., collocations, lexical bundles), as well as methodologies (e.g., self-paced reading task, eye-tracking technique), considerably differ across studies, common findings are that L1 speakers and L2 learners likewise demonstrate sensitivity to phrasal frequency of MWS (but see N. C. Ellis et al., 2008). For example, Siyanova-Chanturia et al. (2011) examined the processing of high-frequency binomials (e.g., bride and groom) against the low-frequency reversed form (e.g., groom and bride) and control phrases in a sample of native and nonnative speakers of English. Results showed that phrase type and phrasal frequency matter for both groups of participants.

A body of evidence suggests that collocations are entrenched in our memory, and the more frequently we encounter collocations, the more quickly we access and process them. As suggested by Wray (2002), this implies that collocation is likely stored and retrieved holistically from long-term memory. To measure the efficiency with which stored lexical knowledge is accessed, a common approach is to use a lexical decision task, which can be further extended to PJT, as demonstrated by Yi (2018), for evaluating the speed and accuracy of collocation processing. However, the extent to which this measure of CF is related to UF remains unexplored. Thus, the present study aimed to address this gap through the lens of the CF-UF link.

The current study

The main aim of this study is to investigate the association between L2 collocational processing speed and accuracy, as elicited in a PJT, and L2 UF. Building on the frequency effect observed in collocational processing research, the present study also tests whether the processing speed of higher-frequency collocations exhibits a

stronger association with UF compared with lower-frequency collocations. This study is guided by the following research questions and hypotheses:

- **RQ1.** How does collocational processing speed relate to L2 UF?
- **Hypothesis 1:** Collocational processing speed is expected to exhibit a stronger relationship with speed and breakdown fluency than repair fluency. Specifically, the mid-clause pause measure is expected to show the largest effect size, given that mid-clause pauses reflect one's linguistic knowledge and processing speed.
- **RQ2.** To what extent does the relationship between collocational processing speed and L2 UF differ across collocation frequency level?
- **Hypothesis 2:** In line with usage-based approaches, it is hypothesized that the processing speed of high-frequency collocation would be more strongly associated with speed and breakdown fluency, especially mid-clause pauses.
- **RQ3.** How does collocational processing accuracy relate to L2 UF?
- **Hypothesis 3:** Considering Kahng's (2020) findings that L2 phrasal vocabulary predicted various UF measures, it is expected that collocational processing accuracy would be related not only to speed and breakdown fluency but also to repair fluency. This expectation is also supported by Koizumi and In'nami (2013), which found a weak but significant relationship between collocation knowledge and repair fluency.

Method

Participants

Japanese undergraduate students ($N = 110$), ranging from 18 to 23 years old ($M = 20.49$, $SD = 1.29$), voluntarily joined the experiment online. About one-third of the participants have more than one month stay overseas in various countries ($M_{month} = 22.98$, $SD = 35.65$, range = 1–210). Their self-reported proficiency score on standardized tests indicated that they were mostly B1 to B2 levels of the Common European Framework of Reference for Languages (Council of Europe, 2001): A2 ($n = 2$), B1 ($n = 21$), B2 ($n = 46$), and C1 ($n = 12$). Table 1 is a breakdown of participants with their self-reported test scores. The mean age of acquisition was 9.20 ($SD = 3.88$).

Materials

Forty-two adjective–noun collocations were extracted from Ackermann and Chen's (2013) Academic Collocation List (ACL). ACL is a comprehensive compilation of academic collocations, combining computational and expert-review approaches. It is based on the written curricular component of the Pearson International Corpus of Academic English (PICAЕ) comprising over 25 million words. Following Wolter and Gyllstad (2013), adjective–noun collocations were selected as a type of collocation to control item variabilities otherwise present in verb–noun collocations.

Table 1. Breakdown of self-reported standardized test score

	<i>M</i>	<i>SD</i>	Range
TOEFL iBT (<i>n</i> = 23)	82.91	13.87	54–106
IELTS (<i>n</i> = 4)	6.75	0.87	6–8
TOEIC R&L (<i>n</i> = 40)	823.91	114.49	510–985
TOEFL ITP (<i>n</i> = 6)	543	65.24	450–640
EIKEN (<i>n</i> = 9)	Grade 2 (<i>n</i> = 2), Grade Pre-1 (<i>n</i> = 7)		

Note: Parentheses represent the number of participants reporting the score; *n* = 28 did not report any score.

The following outlines the construction procedure for the task items. As an initial list of the target collocations, 1,700 adjective–noun collocations were selected from ACL. To enhance theoretical and pedagogical soundness, 86 out of 1,700 collocations were selected as task items. These collocations were chosen based on Ackermann and Chen’s expert agreement rate concerning the pedagogical usefulness (Table 2 on p. 242 in Ackermann & Chen, 2013). Second, the 86 collocations were checked against the Japan Association of College English Teachers (JACET) basic word list, which covers 8,000 vocabulary items with 8 frequency bands (Aizawa et al., 2005). Of these lists, 1,000–4,000-word levels are designated as essential for Japanese university students. There were 11 items beyond the 4,000-word levels found on the list which were subsequently removed, resulting in 75 items. Third, the 75 items were rank-ordered based on the normed frequency available in PICA. These frequencies were then log-transformed using the natural log and divided into three frequency bins guided by Siyanova-Chanturia (2015a) and Yi (2018): high-frequency (*n* = 25), mid-frequency (*n* = 25), and low-frequency (*n* = 25). Fourth, to minimize the effect of any processing benefit from seeing the same word twice, overlapping items, either for adjectives or nouns, were excluded. Fifth, MI values (i.e., association strength of collocations), adjective frequency, noun frequency, and word length were closely matched across the three frequency bins, leading to 14 items each and 42 items in total.

The final factor to control for across the frequency bins was *congruency*, which is a similarity between the two languages and known as having an effect on collocation processing (Yamashita & Jiang, 2010; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2015, 2018). Seven graduate students in Applied Linguistics, native speakers of Japanese, were involved in translating adjectives and nouns for all stimuli (*n* = 42) into Japanese. Following the selection of the most plausible translations for each collocation component, the translated Japanese adjectives and nouns were combined to create the Japanese counterpart collocations. A separate group of three graduate students in Applied Linguistics, also native speakers of Japanese, assessed the naturalness of the Japanese collocations on a 4-point scale (1 = *very natural*, 4 = *very unnatural*). A Japanese collocation judged to be very natural indicates high congruency with its English counterpart. Spearman correlation analysis revealed that the first judge had a high-level agreement with the second and third judges ($\rho = .755, .629$) but a moderate-level agreement between the second and third judges

($\rho = .447$). Mean scores, based on the assessments of the three judges, were calculated for each item and used as congruency scores (see Supplementary information for the translation of adjectives and nouns and congruency judgment).

Filler items ($n = 42$) were constructed by combining the adjectives and nouns systematically to achieve a collocation frequency close to zero. If the collocation frequency was not zero, the MI value was checked because a negative MI value does not signify any meaningful relationship between a pair of words (Gablasova et al., 2017). To accomplish this, the Corpus of Contemporary American English (COCA) was used as the reference corpus due to the limited accessibility to PICAE (see Supplementary information Appendix S1 for the full information about the characteristics of all the items). Table 2 shows the descriptive statistics for length, log-transformed adjective frequency, log-transformed noun frequency, MI, log-transformed collocation frequency, and congruency for all items ($N = 82$) by the frequency level. Two items were excluded from the analysis (see Results section for details). A series of One-way analysis of variance (ANOVA) were performed to examine statistically significant differences in all item characteristics across the frequency levels (see the Data analysis file in OSF). The results indicated that length, adjective frequency, and noun frequency were comparable across the three frequency levels and fillers. MI and congruency were also comparable across the three frequency levels. The only statistically significant difference observed was in collocation frequency across the three frequency levels ($H = 35.567, p < .001$).

Phrasal acceptability judgment task (PJT)

PJT assesses the online processing of MWS in L2 learners by asking them to judge whether the phrase presented on a screen is acceptable or not. In contrast to grammaticality or acceptability judgment tasks, which are often used to examine syntactic processing for grammatical correctness, PJT instructs learners to judge whether phrases are natural or not, rather than their grammaticality. This is because most phrases used in this kind of task are grammatically correct (e.g., Öksüz et al., 2020). Gorilla Experiment Builder (<http://www.gorilla.sc/>) was used to construct PJT (Anwyl-Irvine et al., 2020). Collocation items were presented one at a time in a random order. Participants were instructed to press a button on the keyboard (Y for *yes response* and N for *no response*) as quickly and accurately as possible. The task commenced with a fixation cross presented for 500 milliseconds. Next, a blank screen was presented for 65 milliseconds followed by a screen displaying a collocation for 4,000 milliseconds. Responses taking more than 4,000 milliseconds were considered incorrect, and the screen automatically moved on to the next. To familiarize participants with the task, a practice trial was conducted with six items (i.e., one item for each frequency level and three fillers). Due to online data collection, participants were carefully instructed to ensure the absence of distractors, allowing them to focus on the task.

Speech samples

Speech samples were obtained through an argumentative speech task, where participants expressed their opinions on the advantages and disadvantages of

Table 2. Test item characteristics

Item characteristics	Collocation frequency								<i>F</i> ^a	<i>p</i>
	High		Mid		Low		Filler			
	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI	<i>M</i> (<i>SD</i>)	95% CI		
Item length	14.57 (1.70)	[13.68, 15.46]	13.21 (4.37)	[10.92, 15.50]	15.46 (2.88)	[13.90, 17.03]	14.29 (2.96)	[13.39, 15.20]	0.923	0.442
Adjective frequency	11.60 (0.97)	[11.09, 12.11]	11.36 (0.92)	[10.88, 11.84]	10.82 (1.02)	[10.27, 11.38]	—	—	2.239	0.120
Noun frequency	11.72 (0.80)	[11.31, 12.14]	11.75 (0.52)	[11.47, 12.02]	11.14 (1.21)	[10.48, 11.79]	—	—	2.052	0.143
Congruency	2.07 (0.78)	[1.67, 2.48]	1.95 (1.10)	[1.38, 2.53]	1.90 (0.84)	[1.44, 2.36]	—	—	1.043	0.594
Collocational frequency	2.61 (0.54)	[2.33, 2.89]	1.85 (0.16)	[1.77, 1.93]	1.16 (0.29)	[1.01, 1.32]	—	—	35.567	< .001
MI scores	7.56 (1.38)	[6.84, 8.28]	7.54 (1.50)	[6.75, 8.32]	7.50 (1.19)	[6.86, 8.15]	—	—	0.178	0.915

^aMI, congruency, and collocational frequency were analyzed by the Kruskal-Wallis test due to non-normal distribution.

smartphone use in our everyday lives. Participants were given two minutes for preparation, and no time limit was set for delivering the speech. While preparing, participants were permitted to take notes, but consulting a dictionary or other resources (e.g., internet browsing) was not allowed. The author monitored participants throughout the task. Before the main speech, participants practiced with a self-introduction topic to ensure they were familiarized with the procedure. Speeches were recorded using Praat (Boersma & Weenink, 2020). In Praat, participants initiated and concluded the recording, saved the data on their computers, and sent it via the chat function on Zoom. The mean duration of speeches was 103.77 seconds ($SD = 55.13$, range = 24.88–346.30), and the mean number of total words was 109.78 ($SD = 55.36$, range = 26–306). One participant was unable to download Praat and used Audacity (Audacity Team, 2020). Another participant was also unable to download Praat and recorded his speech with iPhone 7 (iOS 14.3). Two participants were excluded from the analysis because they failed to save the speech data.

Analysis

L2 UF measures were computed based on Tavakoli and Skehan's (2005) framework of speed, breakdown, and repair fluency. For speed fluency, articulation rate (AR) was calculated by dividing the mean number of syllables per second by total spoken time (i.e., excluding pauses). For breakdown fluency, given the multidimensionality of pausing behaviors, the number and duration of pauses in the mid- and end-clauses were computed. Following the previous studies (e.g., De Jong et al., 2013), pauses 250 milliseconds or longer were defined as silent pauses. The definition of a clause followed Foster et al. (2000). Mid-clause silent pause ratio (MSPR) was calculated by the total number of silent pauses within a clause divided by the total number of syllables. End-clause silent pause ratio (ESPR) was calculated by the total number of silent pauses between clauses divided by the total number of syllables. Mid-clause silent pause duration (MSPD) was calculated by the mean duration of silent pauses within clauses expressed in seconds. End-clause silent pause duration (ESPD) was calculated by the mean duration of silent pauses between clauses expressed in seconds. Filled pauses were defined as non-lexicalized (e.g., *eh*, *um*) and lexicalized utterances (e.g., *well*, *like*). Filled pause ratio (FPR) was calculated by the total number of filled pauses divided by the total number of syllables. The repair fluency included self-repetitions, self-corrections, and false starts. Repair ratio (RR) was calculated by the total number of disfluency words divided by the total number of syllables. Syllables were counted on the publicly available website (*SyllableCounter.net*: <https://syllablecounter.net/>), which simply detects the total number of syllables in a text file. After detecting silent pauses (> 250 ms) by a built-in automatic annotation in Praat, silent pauses in mid-clauses or end-clauses, filled pauses, and repair behaviors were manually annotated by the author on TextGrid in Praat. A postdoctoral researcher in L2 fluency was asked to code silent pauses (mid or end), filled pauses, and repairs manually in Praat for 10% of the data, resulting in very high inter-coder reliability (Cronbach's $\alpha = .995$ for mid-clause silent pauses, .976 for end-clause silent pauses, .988 for filled pauses, .987 for disfluency words, .978 for self-repetitions, .924 for self-corrections, and .663 for false starts). All UF

Table 3. UF measures and calculations

Measures	Calculation
Speed fluency	
Articulation rate	The mean number of syllables per second/total phonation time
Breakdown fluency	
Mid-clause silent pause ratio	The total number of silent pauses within clauses/the total number of syllables
End-clause silent pause ratio	The total number of silent pauses between clauses/the total number of syllables
Mid-clause silent pause duration	The mean duration of silent pauses within clauses expressed in seconds
End-clause silent pause duration	The mean duration of silent pauses between clauses expressed in seconds
Filled pause ratio	The total number of filled pauses/the total number of syllables
Repair fluency	
Repair ratio	The total number of disfluency words/the total number of syllables

measures were automatically calculated in R (R Core Team, 2023) with the help of a publicly available R script (S. Suzuki & Révész, 2023). Table 3 summarizes the UF measures and the calculations.

Before proceeding with the main statistical analyses to investigate the relationship between collocational processing speed and L2 UF, it is important to examine whether log-transformed RTs for collocational processing speed differ across frequency levels and fillers. This aims to demonstrate (1) faster processing speed for higher-frequency collocations (high < mid < low) and (2) slower processing speed for fillers. After confirming that the statistical assumptions of normal distribution and homogeneity of variance were met, a one-way repeated measures ANOVA was performed for collocational processing speed at each frequency level. The result indicated significant differences between all frequency levels and fillers, except for high-frequency and mid-frequency collocations (see Supplementary information Appendix S2). To address RQ1, log RTs of the three frequency levels were averaged as a single measure. Given the non-normal distributions of all UF measures except AR, Spearman's rank-order correlation analysis was performed to examine the relationship between log RTs of the collocational processing speed and UF measures. For RQ2, Spearman's rank-order correlation analyses were performed between log RTs of the three frequency levels (high, mid, and low) and UF measures. Subsequently, three frequency variables were incorporated as predictors in multiple regression analyses. To address RQ3, accuracy was calculated as the total number of correct marks obtained across all items (all frequency and fillers). Spearman's rank-order correlation analysis was performed to explore the relationship between accuracy and UF measures. The

Table 4. Descriptive statistics for collocational processing speed and UF

Measures	Min	Max	<i>M</i>	<i>SD</i>	95% CI
					Lower, upper
Collocational processing speed					
Overall (log)	6.78	7.80	7.24	0.19	7.20, 7.28
High-frequency (log)	6.62	7.90	7.21	0.22	7.17, 7.25
Mid-frequency (log)	6.74	7.74	7.19	0.19	7.15, 7.22
Low-frequency (log)	6.88	7.78	7.34	0.20	7.30, 7.37
Collocational processing accuracy					
Accuracy	40.00	72.00	56.29	7.59	54.83, 57.74
Speed fluency					
Articulation rate	1.20	4.22	2.80	0.70	1.43, 4.17
Breakdown fluency					
Mid-clause silent pause ratio	0.03	0.67	0.22	0.13	-0.04, 0.48
End-clause silent pause ratio	0.02	0.24	0.08	0.03	0.02, 0.13
Mid-clause silent pause duration (sec)	0.45	1.73	0.85	0.30	0.26, 1.44
End-clause silent pause duration (sec)	0.39	2.25	0.89	0.35	0.21, 1.57
Filled pause ratio	0.00	0.47	0.10	0.10	-0.09, 0.29
Repair fluency					
Repair ratio	0.00	0.50	0.11	0.09	-0.07, 0.29

significant alpha level was all set at $p < .007$ (calculated by $0.05/7$). Effect size interpretation in the correlation coefficient was guided by Plonsky and Oswald's (2014) recommendations: $r = .25$ for small, $r = .40$ for medium, and $r = .60$ for large effect sizes. All the statistical analyses were performed using Jeffrey's Amazing Statistics Program (JASP) (JASP Team, 2023).

Results

Prior to the main analysis, the PJT performance data underwent trimming and cleaning. This involved excluding one participant due to a technical issue and another participant responding too quickly to over half of the items (59.5%) ($M_{ms} = 429$, $SD_{ms} = 254$). An additional participant was excluded for having an accuracy level of more than -3 SD in all frequency levels. Additionally, one collocation item in the low-frequency list (item 34) was mistakenly copied from the original list and subsequently excluded from the analysis. A filler item that included a word from Item 34 was also excluded (Item 62). Responses exceeding 4,000 milliseconds (0.7% of all data) and responses faster than 450 milliseconds (0.09% of all data) were excluded. Incorrect responses, constituting 17.9% of all data (excluding fillers), were excluded from the analyses for RQ1 and RQ2 (see

Table 5. Spearman correlations between collocational processing speed and UF

Fluency measure	Collocation RT (overall)		
	<i>rho</i>	<i>p</i>	95% CI Lower, upper
Speed fluency			
Articulation rate	-0.271 ^a	0.005	-0.440, -0.084
Breakdown fluency			
Mid-clause silent pause ratio	0.322 ^a	<.001	0.139, 0.484
End-clause silent pause ratio	0.277 ^a	0.004	0.091, 0.445
Mid-clause silent pause duration (sec)	0.205	0.036	0.014, 0.382
End-clause silent pause duration (sec)	0.169	0.085	-0.024, 0.349
Filled pause ratio	-0.029	0.768	-0.220, 0.163
Repair fluency			
Repair ratio	0.135	0.171	-0.058, 0.318

^a*p* < .007.

Supplementary information [Appendix S3](#)). Descriptive statistics for collocational processing speed, accuracy, and UF measures are presented in Table 4.

RQ1 aimed to examine the relationship between collocational processing speed (CF) and L2 UF measures. Log RTs were averaged across the three frequency levels to examine the overall collocational processing speed. Table 5 shows the results of correlation analyses, revealing that collocational processing speed was significantly correlated with AR ($rho = -.271$), MSPR ($rho = .322$), and ESPR ($rho = .277$) with small to medium effect sizes. Conversely, MSPD, ESPD, FPR, and RR did not show any significant correlations with L2 UF with marginal effect sizes.

While RQ1 addressed the relationship between overall collocational processing speed and L2 UF, RQ2 examined whether collocation frequency modulated the relationship between processing speed and L2 UF. This aimed to test the assumption that the processing speed of higher-frequency collocations is more strongly associated with L2 UF in line with the usage-based theory of language acquisition. Correlation analyses results shown in Table 6 suggested that the processing speed of high-frequency collocations seemed to exhibit a stronger relationship with L2 UF measures than that of mid- and low-frequency collocations, particularly in AR ($rho = -.294$) and two pause frequency measures MSPR ($rho = .383$) and ESPR ($rho = .325$) (see scatterplots in Figure 2). Subsequently, three frequency variables were entered into the regression model using a stepwise approach, with assumptions of residual distributions being met (see Data analysis.pdf in OSF). The results in Table 7 showed that, overall, the processing speed of high-frequency collocation explained larger variances in L2 UF measures (3.2%–12.6%). It was noteworthy that the processing speed of low-frequency collocation explained small but additional variances in breakdown fluency measures (3%–5.1%).

Table 6. Spearman correlations between collocational processing speed for each frequency level and UF

	High-frequency collocation			Mid-frequency collocation			Low-frequency collocation		
	<i>rho</i>	<i>p</i>	95% CI	<i>rho</i>	<i>p</i>	95% CI	<i>rho</i>	<i>p</i>	95% CI
			Lower, upper			Lower, upper			Lower, upper
Speed fluency									
Articulation rate	-0.294 ^a	0.002	-0.46, -0.108	-0.231	0.018	-0.405, -0.042	-0.237	0.015	-0.41, -0.047
Breakdown fluency									
Mid-clause silent pause ratio	0.383 ^a	<.001	0.207, 0.535	0.277 ^a	0.004	0.09, 0.445	0.257	0.008	0.069, 0.428
End-clause silent pause ratio	0.325 ^a	<.001	0.143, 0.487	0.226	0.02	0.036, 0.401	0.242	0.013	0.053, 0.414
Mid-clause silent pause duration (sec)	0.245	0.012	0.056, 0.417	0.179	0.067	-0.013, 0.359	0.2	0.041	0.008, 0.377
End-clause silent pause duration (sec)	0.209	0.033	0.018, 0.385	0.154	0.117	-0.039, 0.336	0.141	0.152	-0.052, 0.324
Filled pause ratio	0.007	0.945	-0.185, 0.198	-0.099	0.313	-0.286, 0.094	-0.018	0.855	-0.209, 0.174
Repair fluency									
Repair ratio	0.166	0.09	-0.026, 0.347	0.078	0.427	-0.115, 0.266	0.126	0.2	-0.067, 0.31

^a*p* <.007.

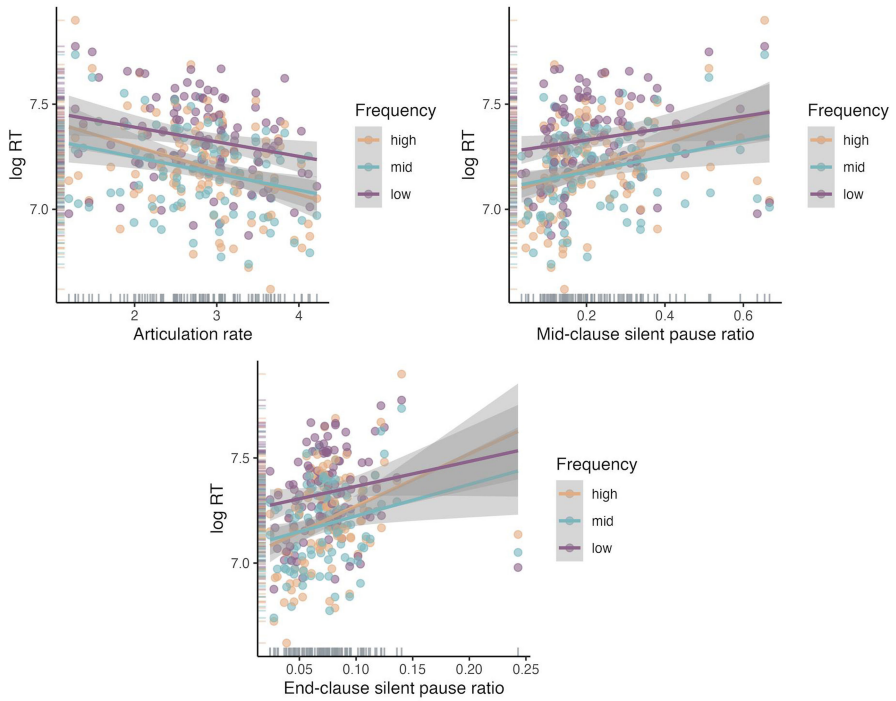


Figure 2. Scatterplots between collocational processing speed (log RT) and UF.

Table 7. Summary of multiple regression analyses

Outcome variable	Predictors	Adj. R^2	β	F	p
Speed fluency					
Articulation rate	High	0.126	-0.366	15.93	<.001
Breakdown fluency					
Mid-clause silent pause ratio	High	0.115	0.773	14.509	<.001
	High + low	0.165	-0.485	11.241	0.009
End-clause silent pause ratio	High	0.108	0.769	13.634	<.001
	High + low	0.159	-0.491	10.851	0.008
Mid-clause silent pause duration (sec)	High	0.063	0.269	8.010	0.006
End-clause silent pause duration (sec)	High	0.032	0.552	4.460	0.037
	High + low	0.062	-0.400	4.306	0.040

Note: Predictors were high-frequency, mid-frequency, and low-frequency collocational processing speed (log RT); filled pause ratio and repair ratio were not significantly predicted; variance inflation factor of high and low-frequency RT was 4.117.

Table 8. Spearman correlations between collocational processing accuracy and UF

Fluency measure	Collocation accuracy		
	ρ	p	95% CI Lower, upper
Speed fluency			
Articulation rate	0.356 ^a	<.001	0.176, 0.513
Breakdown fluency			
Mid-clause silent pause ratio	-0.368 ^a	<.001	-0.523, -0.19
End-clause silent pause ratio	-0.342 ^a	<.001	-0.501, -0.161
Mid-clause silent pause duration (sec)	-0.321 ^a	<.001	-0.483, -0.138
End-clause silent pause duration (sec)	-0.29 ^a	0.003	-0.457, -0.105
Filled pause ratio	-0.129	0.19	-0.313, 0.064
Repair fluency			
Repair ratio	-0.281 ^a	0.004	-0.448, -0.094

^a $p < .007$.

Finally, RQ3 addressed the relationship between collocational processing accuracy and L2 UF measures. The results of correlation analyses in Table 8 revealed that collocational processing accuracy was significantly correlated with all L2 UF measures except FPR with small to medium effect sizes ($\rho = |.281-.368|$).

Discussion

Motivated by the recent L2 fluency research shedding light on underlying linguistic resources and processing speed that enable efficient speech production, the current study aimed to investigate the connection of collocational processing speed and accuracy to L2 UF. Moreover, the study sought to determine whether the relationship between collocational processing speed and L2 UF was modulated by collocation frequency.

Collocational processing speed and L2 UF

The first RQ addressed whether RT of collocational processing speed was associated with L2 UF. The results of the correlation analyses indicated that RT was significantly correlated with AR ($\rho = -.271$), MSPR ($\rho = .322$), and ESPR ($\rho = .277$), with small to medium effect sizes. This suggests that faster access to collocation knowledge leads to more efficient overall speech production and fewer pauses at both mid- and end-clauses, aligning with recent CF-UF link research (De Jong et al., 2013; Kahng, 2020; S. Suzuki & Kormos, 2022). Speed fluency, measured as mean syllable duration or AR, serves as an indicator of L2-specific CF, as evidenced in studies by De Jong et al. (2013), Kahng (2020), and Suzuki and Kormos (2022), where approximately 40%–50% of the variance was explained by CF

measures. Breakdown fluency, especially a measure of mid-clause pause frequency, also emerges as a robust indicator of L2-specific CF (Kahng, 2020; S. Suzuki & Kormos, 2022). Regarding lexical processing speed, previous studies employed a picture naming task and found that the RT of lexical retrieval was related to speed fluency (e.g., mean syllable duration) and the frequency of silent pauses. The current study extends prior findings by demonstrating that collocational processing speed also underlies fluent speech, highlighting the need for a more diverse measure of lexical knowledge, including collocation knowledge and processing speed, as part of CF measures. While the importance of phraseological proficiency underlying fluent speech production has been emphasized in previous studies (De Jong et al., 2013; Koizumi & In'nami, 2013; Kahng, 2020; Uchihara et al., 2021), most studies have focused on knowledge measures (i.e., linguistic resource) rather than speed measures (i.e., processing speed).

The current study revealed a relatively weak association between collocational processing speed and silent pause duration, filled pauses, and repairs. This finding is consistent with earlier research conducted by De Jong et al. (2013) and Kahng (2020), who reported a limited relationship between mean length of silent pauses and most measures of CF, suggesting that pause duration may reflect language-general characteristics such as personality. However, care needs to be taken as S. Suzuki and Kormos (2022) found significant correlations between MSPD and ESPD and various measures of CF. With respect to filled pauses and repairs, Kahng (2020) suggests that these two reflect language-general fluency, rather than L2-specific fluency, a view echoed by De Jong et al. (2015). In addition, S. Suzuki and Kormos (2022) did not find a meaningful relationship between the processing speed factor and repair fluency factor. Given that repair behaviors reflect monitoring—self-monitoring speech processing in terms of accuracy and appropriateness (Kormos, 1999)—it could be accurate collocation knowledge rather than speed of access to it that matters in reducing repairs. It is also possible that participants' proficiency levels confounded the results because oral fluency is influenced by language proficiency (Tavakoli et al., 2020). Specifically, Tavakoli et al. showed that AR and pause frequency had a relatively linear relationship with proficiency, but pause duration, filled pauses, and repairs exhibited less linearity with proficiency.

Frequency effects on the relationship between collocational processing speed and L2 UF

Usage-based theories of language acquisition suggest that the frequency and probability of linguistic units are crucial for learning and processing language. Ellis (2009) suggested that fluent speakers' "unconscious language representations are adaptively probability-tuned to predict the linguistic constructions that are most likely to be relevant in the ongoing discourse context, optimally preparing them for comprehension and production" (p. 152). Previous evidence indicates that L2 learners' collocation knowledge tends to favor high-frequency collocations (Ellis et al., 2008; González Fernández & Schmitt, 2015; Siyanova-Chanturia et al., 2011). This suggests that high-frequency collocations are more likely to be entrenched in an L2 learner's memory such that high-frequency collocations would be more robust in language comprehension and production.

RQ2 tested whether the processing speed of higher-frequency collocations was more closely associated with L2 UF than lower-frequency ones. Correlation analyses revealed that the processing speed of high-frequency collocations seemed to be more associated with AR ($\rho = -.294$), MSPR ($\rho = .383$), and ESPR ($\rho = .325$). Examining the predictive strength of processing speed at each frequency level through multiple regressions showed that the processing speed of high-frequency collocations explained a larger or unique variance in speed and breakdown fluency measures (adjusted $R^2 = 0.32-0.126$), while the processing speed of low-frequency collocations explained additional variance in several pause measures. These results largely confirmed previous findings that the role of higher-frequency collocations is more significant than lower-frequency ones (Ellis et al., 2008; Durrant & Schmitt, 2009; González Fernández & Schmitt, 2015; Siyanova-Chanturia et al., 2011; Yi, 2018). Ellis et al. (2008) suggested that L2 learners' processing of MWS (ngrams) was affected by phrase frequency rather than mutual exclusivity in lexical decision, priming, and production tasks. Similar evidence was found by Durrant and Schmitt (2009), who observed that L2 learners tended to overuse high-frequency collocations and underuse exclusive collocations (measured as MI). L2 learners' sensitivity to collocation frequency was also evident in collocation processing research (Siyanova-Chanturia et al., 2011; Yi, 2018). Despite the predominant role of high-frequency collocations in L2 learners' phrasal lexicon, it has been less clear how high-frequency collocation knowledge is employed in fluent speech compared to lower-frequency ones. This study contributes additional evidence to CF-UF link research, highlighting that the relationship between collocational processing speed and L2 UF was largely modulated by high-frequency collocations.

Contrary to the expectation that the strength of prediction would follow the order of high- > mid- > low-frequency, regression analysis revealed that the processing speed of mid-frequency collocations was not a significant predictor for any L2 UF measures. The preliminary analysis indicated that high-frequency and mid-frequency collocations did not significantly differ in processing speed (see Supplementary information Appendix S2), suggesting that the explained variance of L2 UF by mid-frequency collocation processing could be overshadowed by high-frequency collocation processing. One speculation is that, as suggested in Siyanova-Chanturia and Schmitt (2008), the participants might not have been able to differentiate between high-frequency and mid-frequency collocations, thereby reducing the predictive power of mid-frequency collocation processing for L2 UF. Another potential factor could stem from the selection of target collocations in ACL (Ackermann & Chen, 2013). The initial list of 86 adjective-noun collocations was chosen based on the highest expert agreement regarding pedagogical usefulness. Therefore, although collocation frequency in the target corpus (high, mid, and low) was statistically significantly different (see Table 2), high- and mid-frequency collocations might have been equally important and frequent for the participants in their course of learning history.

Collocational processing accuracy and L2 UF

As an extension of recent CF-UF link research, this study examined collocational processing accuracy as an underlying CF measure, operationalized as the number of

correct marks across the entire item set. S. Suzuki and Kormos (2022) took a similar approach to measure accuracy in timed grammatical processing tasks, a part of the CF measure (linguistic resource factor). Correlation analyses showed that collocational processing accuracy was associated with all L2 UF measures except FPR, with small to medium effect sizes ($\rho = .281-.368$). The findings largely confirmed previous studies that demonstrated collocation knowledge underpins L2 UF (De Jong et al., 2013; Koizumi & In'nami, 2013; Kahng, 2020; Uchihara et al., 2021), claiming that speech production could be phraseologically driven. The results demonstrated that having larger accurate collocation knowledge enhanced the overall speed of articulation and reduced disruptions and repairs during speech. It should be noted, however, that the results were possibly affected by the proficiency level of the participants.

It seems that collocational processing accuracy had more robust associations with L2 UF measures than collocational processing speed. It is notable that silent pause durations and repairs were significantly associated with accuracy. This discrepancy between collocational processing speed and accuracy can be interpreted in two ways. One possibility is that collocational processing speed and accuracy might tap into different constructs, thus having differential associations with L2 UF, as S. Suzuki and Kormos (2022) illustrated in their analysis. The other possibility could be due to the fact that collocational processing speed and accuracy were computed differently. That is, the processing speed analysis was based only on correct responses to the target collocations, while the accuracy analysis was based on all responses, including fillers (see Supplementary information Appendix S3).

Limitations and future directions

Several limitations are worth pointing out. First and foremost, this study did not incorporate an external measure of participants' proficiency levels into the relationship between collocational processing speed and accuracy and L2 UF measures. The participants in this study were at a relatively advanced level according to self-reported proficiency scores. Thus, it is possible that the significant relationship between collocational processing speed and accuracy with L2 UF would be weaker in lower-level L2 learners. Second, this study focused on collocation as a type of MWS, which have various types according to the definition and operationalization (Wood, 2020). Thus, in future studies, it is imperative to see the extent to which the current findings are generalizable to different types of MWS (e.g., idioms). Third, this study operationalized collocation as an adjective–noun combination, a pair often utilized in collocation processing research to control for the variabilities of nouns otherwise present in a verb–noun collocation (i.e., a/the or plural-s) (e.g., Wolter & Gyllstad, 2013). Future studies are expected to remedy this convention to explore different forms of collocations and test collocational processing as an underlying CF. Fourth, it should be noted that the corpus from which the target collocations in ACL (Ackermann & Chen, 2013) were derived consisted of written academic English. Given the population in this study were university students, it was deemed appropriate to test the processing of academic collocations. However, in future studies, it would be meaningful to investigate whether the processing of collocations from other genres (e.g., TV shows) attests to

underlying CF measures to predict L2 UF. Finally, the speech samples in this study were elicited based solely on an argumentative task. The task effect on CF-UF links was aptly illustrated by Suzuki and Kormos (2022), concluding that the availability of content or linguistic input in tasks influenced the complex interplay between CF and L2 UF measures. An argumentative task is relatively less controlled due to the absence of predefined content or linguistic items. Future investigations are worth exploring the relationship between the processing of MWS and L2 UF through tasks with different speech processing demands.

Conclusion

Motivated by the recent investigations that underlying cognitive efficiencies of oral fluency predict L2 UF, this study operationalized L2-specific CF as speed and accuracy of collocational processing elicited in a PJT. The results demonstrated that L2 learners' collocational processing speed underpinned AR, MSPR, and ESPR, while collocational processing accuracy underpinned most L2 UF measures. The analysis of frequency effects on the relationship between collocational processing speed and L2 UF measures indicated that the processing speed of high-frequency collocations was predominant in predicting speed and breakdown fluency measures, aligning with the usage-based model of language acquisition (Ellis, 2002). This study contributes to L2 fluency research by showing that L2 learners' (receptive) phraseological proficiency also underlies fluent speech production, a topic relatively under-researched in recent CF-UF link investigations (e.g., S. Suzuki & Kormos, 2022). Despite the fact that the role of MWS in fluent speech production has been a central topic for decades in Second Language Acquisition (Pawley & Syder, 1983; Wray, 2002; Kormos, 2006), very few attempts have been made explicitly so far to explore what and how phraseological competence underlies L2 UF and automaticity in speech production (Koizumi & In'nami, 2013; Kahng, 2020; Uchihara et al., 2021). To better understand the multidimensionality of oral fluency and how it develops in the course of L2 learning, more research is necessary to explore the link between CF and UF considering the wide variety of linguistic knowledge and processing, including different types of MWS.

Replication package. (1) The study materials, (2) analysis code, and (2) data used for the present study can be found at <https://osf.io/kyuaw/>.

Acknowledgments. This study was supported by the EIKEN Research Assistance Program in 2020 from the EIKEN Foundation of Japan. I would like to express my gratitude toward Tetsuo Harada for his great mentorship. I would also like to acknowledge Shungo Suzuki for his insightful comments, suggestions, and supports for the entire project. Finally, I extend my gratitude and appreciation to Masaki Eguchi, Takumi Uchihara, and three anonymous *Applied Psycholinguistics* reviewers as well as the associate editor, Amanda Huensch, for their constructive feedback on earlier versions of the manuscript.

Competing interests. The author declares none.

References

- Ackermann, K., & Chen, Y. H. (2013). Developing the academic collocation list (ACL): A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235–247. <https://doi.org/10.1016/j.jeap.2013.08.002>

- Aizawa, K., Ishikawa, S., Murata, M., Iso, T., Kamimura, T., Ogawa, T., Shimizu, S., Sugimori, N., Haisa, A., & Mochizuki, M. (2005). *JACET 8000 Eittango [JACET 8000 word list]*. Kirihara-shoten.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52(1), 388–407. <https://doi.org/10.3758/s13428-019-01237-x>
- Audacity Team. (2020). *Audacity(R): Free Audio Editor and Recorder* (Version 2.4.2) [computer application]. <https://www.audacityteam.org/> (accessed December 2020).
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a Lexical Approach to the test. *Language Teaching Research*, 10(3), 245–261. <https://doi.org/10.1191/1362168806lr195oa>
- Boersma, P., & Weenink, D. (2020). PRAAT: Doing phonetics by computer (version 6.1.44) [Computer program]. www.praat.org
- Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, 30(2), 159–175. <https://doi.org/10.1177/0265532212455394>
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45–61. <https://doi.org/10.1017/S0267190512000074>
- Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *IRAL - International Review of Applied Linguistics in Language Teaching*, 54(2), 113–132. <https://doi.org/10.1515/iral-2016-9993>
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, 36(2), 223–243. <https://doi.org/10.1017/S0142716413000210>
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(5), 893–916. <https://doi.org/10.1017/S0142716412000069>
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL - International Review of Applied Linguistics in Language Teaching*, 47(2), 157–177. <https://doi.org/10.1515/iral.2009.007>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143–188.
- Ellis, N. C. (2009). Optimizing the input: Frequency and sampling in usage-based and form-focused learning. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 139–158). Blackwell Publishing. <https://doi.org/10.1002/9781444315783.ch9>
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44. <https://doi.org/10.1017/S0267190512000025>
- Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 61–78. <https://doi.org/10.1515/CLLT.2009.003>
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354–375. <https://doi.org/10.1093/applin/21.3.354>
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155–179. <https://doi.org/10.1111/lang.12225>
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? *ITL - International Journal of Applied Linguistics*, 166(1), 94–126. <https://doi.org/10.1075/itl.166.1.03fer>
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24–49. <https://doi.org/10.1093/applin/amm017>
- JASP Team. (2023). *JASP* (Version 0.17) [computer software]. <https://jasp-stats.org/> (accessed February 2023).

- Kahng, J.** (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809–854. <https://doi.org/10.1111/lang.12084>
- Kahng, J.** (2018). The effect of pause location on perceived fluency. *Applied Psycholinguistics*, *39*(3), 569–591. <https://doi.org/10.1017/S0142716417000534>
- Kahng, J.** (2020). Explaining second language utterance fluency: Contribution of cognitive fluency and first language utterance fluency. *Applied Psycholinguistics*, *41*(2), 457–480. <https://doi.org/10.1017/S0142716420000065>
- Koizumi, R., & In'nami, Y.** (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*(5), 900–913. <https://doi.org/10.4304/jltr.4.5.900-913>
- Kormos, J.** (1999). Monitoring and self-repair in L2. *Language Learning*, *49*(2), 303–342. <https://doi.org/10.1111/0023-8333.00090>
- Kormos, J.** (2006). *Speech production and second language acquisition*. Lawrence Erlbaum.
- Lambert, C., Aubrey, S., & Leeming, P.** (2021). Task preparation and second language speech production. *TESOL Quarterly*, *55*(2), 331–365. <https://doi.org/10.1002/tesq.598>
- Lambert, C., Kormos, J., & Minn, D.** (2017). Task repetition and second language speech processing. *Studies in Second Language Acquisition*, *39*(1), 167–196. <https://doi.org/10.1017/S0272263116000085>
- Levelt, W. J. M.** (1989). *Speaking: From intention to articulation*. MIT Press.
- McGuire, M., & Larson-Hall, J.** (2017). Teaching formulaic sequences in the classroom: Effects on spoken fluency. *TESL Canada Journal*, *34*(3), 1–25. <https://doi.org/10.18806/tesl.v34i3.1271>
- McGuire, M., & Larson-Hall, J.** (2021). The contribution of high-frequency multi-word sequences to speech rate and listening perception among EFL learners. *Vocabulary Learning and Instruction*, *10*(2), 18–29. <https://doi.org/10.7820/vli.v10.2.mcguire.larson-hall>
- Nergis, A.** (2021). Can explicit instruction of formulaic sequences enhance L2 oral fluency? *Lingua*, *255*, 103072. <https://doi.org/10.1016/j.lingua.2021.103072>
- Öksüz, D., Brezina, V., & Rebuschat, P.** (2020). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, *71*(1), 55–98. <https://doi.org/10.1111/lang.12427>
- Pawley, A., & Syder, F. H.** (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 191–226). Longman. <https://doi.org/10.4324/9781315836027-12>
- Plonsky, L., & Oswald, F. L.** (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. <https://doi.org/10.1111/lang.12079>
- R Core Team.** (2023). *R: A language and environment for statistical computing* (Version 4.3.1). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S.** (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, *39*(3), 593–617. <https://doi.org/10.1017/S0142716417000571>
- Segalowitz, N.** (2010). *Cognitive bases of second language fluency*. Routledge.
- Segalowitz, N.** (2016). Second language fluency and its underlying cognitive and social determinants. *IRAL - International Review of Applied Linguistics in Language Teaching*, *54*(2), 79–95. <https://doi.org/10.1515/iral-2016-9991>
- Segalowitz, N., French, L., Guay, J., Segalowitz, N., & French, L.** (2017). What features best characterize adult second language utterance fluency and what do they reveal about fluency gains in short-term immersion? *Canadian Journal of Applied Linguistics*, *20*(2), 90–116. <https://doi.org/10.7202/1050813ar>
- Siyanova-Chanturia, A.** (2015a). Collocation in beginner learner writing: A longitudinal study. *System*, *53*, 148–160. <https://doi.org/10.1016/j.system.2015.07.003>
- Siyanova-Chanturia, A.** (2015b). On the “holistic” nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, *11*(2), 285–301. <https://doi.org/10.1515/clt-2014-0016>
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B.** (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 776–784. <https://doi.org/10.1037/a0022531>
- Siyanova-Chanturia, A., & Martinez, R.** (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–569. <https://doi.org/10.1093/applin/amt054>

- Siyanova-Chanturia, A., & Schmitt, N.** (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, *64*(3), 429–458. <https://doi.org/10.3138/cmlr.64.3.429>
- Skehan, P.** (2003). Task-based instruction. *Language Teaching*, *36*, 1–14.
- Skehan, P., Foster, P., & Shum, S.** (2016). Ladders and snakes in second language fluency. *IRAL - International Review of Applied Linguistics in Language Teaching*, *54*(2), 97–111. <https://doi.org/10.1515/iral-2016-9992>
- Suzuki, S., & Kormos, J.** (2019). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition*, *42*, 143–167. <https://doi.org/10.1017/S0272263119000421>
- Suzuki, S., & Kormos, J.** (2022). The multidimensionality of second language oral fluency: Interfacing cognitive fluency and utterance fluency. *Studies in Second Language Acquisition*, *45*(1), 38–64. <https://doi.org/10.1017/S0272263121000899>
- Suzuki, S., Kormos, J., & Uchihara, T.** (2021). The relationship between utterance and perceived fluency: A meta-analysis of correlational studies. *The Modern Language Journal*, *105*(2), 435–463. <https://doi.org/10.1111/modl.12706>
- Suzuki, S., & Révész, A.** (2023). Measuring speaking and writing fluency: A methodological synthesis focusing on automaticity. In S. Yuichi (Ed.), *Practice and automatization in second language research: Theory, methods, and pedagogical implications* (pp. 235–264). Routledge.
- Suzuki, Y., Eguchi, M., & De Jong, N.** (2022). Does the reuse of constructions promote fluency development in task repetition? A usage-based perspective. *TESOL Quarterly*, *56*(4), 1290–1319. <https://doi.org/10.1002/tesq.3103>
- Tavakoli, P.** (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71–79. <https://doi.org/10.1093/elt/ccq020>
- Tavakoli, P., & Hunter, A. M.** (2018). Is fluency being ‘neglected’ in the classroom? Teacher understanding of fluency and related classroom practices. *Language Teaching Research*, *22*(3), 330–349. <https://doi.org/10.1177/1362168817708462>
- Tavakoli, P., Nakatsuhara, F., & Hunter, A.** (2020). Aspects of fluency across assessed levels of speaking proficiency. *Modern Language Journal*, *104*(1), 169–191. <https://doi.org/10.1111/modl.12620>
- Tavakoli, P., & Skehan, P.** (2005). Strategic planning, task structure and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–277). John Benjamins.
- Tavakoli, P., & Uchihara, T.** (2019). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*(2), 506–547. <https://doi.org/10.1111/lang.12384>
- Thomson, H.** (2017). Building speaking fluency with multiword expressions. *TESL Canada Journal*, *34*(3), 26–53. <https://doi.org/10.18806/tesl.v34i3.1272>
- Uchihara, T., Eguchi, M., Clenton, J., & Saito, K.** (2021). To what extent is collocation knowledge associated with oral proficiency? A corpus-based approach to word association. *Language and Speech*, *65*(2), 311–336. <https://doi.org/10.1177/00238309211013865>
- Wolter, B., & Gyllstad, H.** (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, *35*(3), 451–482. <https://doi.org/10.1017/S0272263113000107>
- Wolter, B., & Yamashita, J.** (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193–1221. <https://doi.org/10.1017/S0142716414000113>
- Wolter, B., & Yamashita, J.** (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, *40*(2), 395–416. <https://doi.org/10.1017/S0272263117000237>
- Wood, D.** (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *The Canadian Modern Language Review*, *63*(1), 13–33.
- Wood, D.** (2009). Effects of focused instruction of formulaic sequences on fluent expression in second language narratives: A case study. *Canadian Journal of Applied Linguistics*, *12*(1), 39–58.
- Wood, D.** (2020). Classifying and identifying formulaic language. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 30–45). Routledge.
- Wray, A.** (2002). *Formulaic language and the lexicon*. Cambridge University Press.

- Yamashita, J., & Jiang, N.** (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, *44*(4), 647–668. <https://doi.org/10.5054/tq.2010.235998>
- Yan, X.** (2020). Unpacking the relationship between formulaic sequences and speech fluency on elicited imitation tasks: Proficiency level, sentence length, and fluency dimensions. *TESOL Quarterly*, *54*(2), 460–487. <https://doi.org/10.1002/tesq.556>
- Yi, W.** (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*(4), 831–856. <https://doi.org/10.1017/S0272263118000141>

Cite this article: Takizawa, K. (2024). What contributes to fluent L2 speech? Examining cognitive and utterance fluency link with underlying L2 collocational processing speed and accuracy. *Applied Psycholinguistics* *45*, 516–541. <https://doi.org/10.1017/S014271642400016X>