Factors affecting successful inferencing of unfamiliar idiomatic expressions in a listening task: A case of Japanese university EFL learners

Junko Noguchi

Teikyo Heisei University, Japan

Bio data

Junko Noguchi received her M.A. in TESOL from Soka University of America and her Ph.D. from Temple University, Japan. She began her career as an English teacher at a public high school in Japan but has spent the last 14 years teaching and researching at various Japanese universities. Her research interests include articulatory phonology, self-directed learning, learning strategies, metacognition, listening, and multi-unit words.
Email: junko.noguchi.baierschmidt@gmail.com, Address: 4 Chome-21-2 Nakano, Nakano City, Tokyo 164-0001

Declaration of interests

The author declares that she has no known competing financial interests or personal relationships that could have influenced the work reported on in this paper.

Abstract

Many studies have been conducted to identify factors that influence learners' success in guessing the meaning of unknown words. Although lexical inferencing is the most frequently employed listening strategy when L2 learners face unfamiliar words, most research on lexical inferencing has been conducted in the context of reading. Additionally, the few studies that have investigated lexical inferencing during listening have focused on single word items despite the ubiquitous presence and importance of multi-word units. In this study, I investigated factors affecting successful inferencing during listening when EFL learners encountered unfamiliar multi-word idiomatic expressions. The participants were 89 Japanese students majoring in English at a private university in Japan. I collected data using a Listening Vocabulary Level Test, a listening span test, an idiom task, and an idiom identification task and analyzed these data using mixed-effects logistic regression. The results indicated that familiarity, listening comprehension skills, working memory, and L1 similarity were significant factors that influenced the success of learners in inferring the meaning of

unfamiliar idiomatic expressions while listening. These results suggest that, rather than relying on learners to infer the meanings of idiomatic expressions from context, explicit instruction of English idiomatic expressions to EFL learners is likely beneficial in settings where students have low listening comprehension or in cases where the target idiomatic expressions do not have a similar L1 counterpart.

Keywords: idiomatic expressions, listening comprehension, listening strategies, lexical inferencing, mixed-effects logistic regression.

Introduction

The large body of research on the lexical inferencing of EFL learners has concentrated mainly on identifying the necessary factors to make successful inferences when reading in a second language (L2). Kaivanpanah and Alavi (2008) listed aspects affecting guessing success in reading, dividing them into learner-related factors and text-related factors. The six learner-related factors they discussed are vocabulary knowledge (e.g., Hatami & Tavakoli, 2012; Nassaji, 2006), knowledge of grammar (e.g., Paribakht, 2005), language proficiency (e.g., Kaivanpandah & Moghaddam, 2012; Yin, 2013), attention to details (e.g., Laufer, 1997), cognitive and mental effort (e.g., Paribakht & Wesche, 1999), and readers' individual characteristics such as background knowledge, interest, familiarity with topic, previous learning experiences, and learning styles (e.g., Paribakht, 2005). In addition to learner-related factors, four text-related factors have been found to influence success: word characteristics (e.g., Na & Nation, 1985), text characteristics (e.g., Chegeni & Tabatabaei, 2014), the presence of contextual clues (e.g., Bengeleil & Paribakht, 2004; Haynes, 1994; Wesche & Paribakht, 2010), and topics (e.g., Pulido, 2007). Wesch and Paribackht (2010) furthermore claim that learners' first language (L1) influences their success in lexical inferencing, particularly the distance between the L1 and L2. Compared to reading, far fewer studies on L2 lexical inferencing have been conducted in the field of listening. Therefore, it is unclear if the findings on lexical inferencing in reading can also be applied in listening contexts.

Several studies have attempted to fill this gap. Farahani and Foomani (2015) investigated the relationship between listening proficiency and lexical inferencing success and found that more proficient listeners were more successful in inferring the meaning of unknown words than less proficient listeners. In a study conducted by van Zeeland (2014), similar results were produced as the reading studies cited above; higher background knowledge and larger vocabulary size led to more successful inferences while listening. It was also found that the achieved success rate was higher when an inference was made based on local clues (i.e., clues in a sentence that include the target unfamiliar lexical item) rather than global clues (i.e., clues that exist in the information produced by interpreting the surrounding sentences or by synthesizing the whole text) (Çetinavcı, 2014). Foomani (2015) also examined the role of depth of vocabulary knowledge in lexical inferencing during

listening by using regression analysis and found that the depth of vocabulary knowledge explained 48% of the success in the inference task.

Additionally, multiword units have been receiving increased attention among language teachers and researchers as a linguistic aspect that should be taught due to their ubiquitous existence, prevailing influence, and the possible advantages that figurative competence can bring for enhanced comprehension and fluency (Boers & Lindstromberg, 2012). Among the different types of multi-word units, those that are conventionalized, relatively fixed, and have different figurative meanings from the meaning of each component word of the phrase are called idioms or idiomatic expressions. Idiomatic expressions are beneficial for language learners to have knowledge of because they are omnipresent in everyday speech (Cooper, 1998). However, idioms can be difficult for L2 learners to acquire (Irujo, 1986b; Wang, 2020). Because inferencing is often the first step to lexical acquisition (Sternberg, 1987) as well as one of the most frequently employed strategies when L2 learners are faced with unfamiliar idiomatic expressions (Cooper 1999; Park & Chon, 2019), investigation into factors that determine the success of idiom inferencing are worth exploring. Nonetheless, so far, lexical inferencing studies in adult foreign language acquisition have been limited to single words. Therefore, in this study I investigated what factors affect the inferencing success rate when EFL learners are guessing the meaning of unfamiliar idiomatic expressions.

Literature Review

Because few studies have focused on inferencing strategies for unknown or unfamiliar idiomatic expressions during listening, possible factors that influence EFL learners' idiom inferencing process need to be hypothesized based on the relevant literature in the fields of listening comprehension, lexical inferencing in reading, and L1 and L2 idiom processing. Based on a review of the research in these fields, I identified several variables potentially important in English idiom inferencing during listening, which I categorized into three levels: person-level factors, sentence level factors, and lexical level factors. Person-level factors include the listener's: (1) familiarity with the idiom, (2) English listening proficiency, (3) listening vocabulary knowledge, and (4) working memory. At the sentence level, (1) lexical density and (2) sentence length seem important, while at the lexical level, (1) L1 similarity and (2) semantic transparency seem likely to influence inferencing success. Therefore, I

included all these variables for investigation in this study. In the following sections, I will introduce the findings from existing studies for each of the above factors.

Person Level Factors

In this section, I will discuss what has been found in studies related to the four person-level factors, namely (1) familiarity with the idiom, (2) English listening proficiency, (3) listening vocabulary knowledge, and (4) working memory.

Familiarity

Familiarity is often operationalized as an individual's subjective judgment of how frequently one has been exposed to and how well one knows the target lexical items (see Karlsson, 2019; Nippold & Rudzinski, 1993). Familiarity is an important factor that affects not only the level of L1 idiom comprehension (Nippold & Rudzinski, 1993; Nippold & Taylor, 2002) but also L2 idiom comprehension (Abel, 2003; Carrol et al., 2016). My aim in this study was to investigate the influence of various factors on the success probability of inferring *unfamiliar* idiomatic expressions in a listening task. Therefore, familiarity was included as a covariate to be controlled for.

English Listening Proficiency

Based on a synthesis of the extant research on lexical inferencing and idiom processing, it seems likely that English listening proficiency will be one determinant of idiom inference success. Studies on lexical inferencing in reading have found that the presence of contextual clues affect inference success (e.g., Bengeleil & Paribakht, 2004; Haynes, 1994; Wesche & Paribakht, 2010). Additionally, it has been pointed out that contextual clues need to be perceptually and conceptually recognizable to the reader for them to be used to disambiguate the meaning of unfamiliar words (Frantzen, 2003; Haynes, 1994; Laufer, 1997; Li, 1988). In other words, even if there is ample information to support readers in inferring the meaning of unfamiliar words in the surrounding context, the information will be of no use if readers do not possess the necessary linguistic knowledge to utilize such contextual clues in constructing an accurate semantic representation of the unfamiliar word or words. Thus, L2 proficiency is deemed to be a decisive factor in lexical inferencing procedures (e.g., Anvari & Farvardin, 2016; Bengeleil & Paribakht, 2004; Hu & Nassaji, 2014; Morrison, 1996; Qian,

2005). Some researchers assume that there is a threshold proficiency level that readers need to achieve to make accurate lexical inferences (e.g., Haynes, 1994; Kaivanpanah & Moghaddam, 2012). Therefore, it follows logically that lexical inferencing success in listening, which depends on understanding available information and utilizing contextual clues, is affected by the listener's comprehension ability. Farahani and Foomani (2015) investigated the relationship between listening proficiency and lexical inferencing success. The results indicated that more proficient listeners inferred the meaning of unknown words than less proficient listeners.

Vocabulary Knowledge

One of the most well-supported findings in all studies is that learners' lexical knowledge is a significant factor in L2 listening comprehension (Andringa et al., 2012; Mecartty, 2000; Stæhr, 2009; Wang & Treffers-Daller, 2017). For instance, Stæhr (2009) found that vocabulary size accounted for 49% of the variance in a measure of listening comprehension, and Mecartty (2000) reported that lexical knowledge accounted for 14% of the variance in listening comprehension. In addition, the results of Vandergrift and Baker's (2015) investigation of the path model indicated that L2 vocabulary accounted for 49% of the variance in listening (see Wang & Treffers-Daller, 2017, for a thorough review of the contribution of vocabulary to listening). Furthermore, Andringa et al. (2012) discovered that the Knowledge factor, which was extracted from measures of vocabulary, grammatical accuracy, and segmentation accuracy, had a strong correlation ($r(119) = .95$) with listening comprehension and, together with IQ, explained 96% of the variance in listening comprehension.

In a study conducted by van Zeeland (2014), greater background knowledge and a larger vocabulary led to more successful inferences in listening. Foomani (2015) also examined the role of depth of vocabulary knowledge in lexical inferencing during listening through a regression analysis and found that depth of vocabulary knowledge explained 48% of inference task success. Although Foomani (2015) investigated a different aspect of vocabulary than van Zeeland (2014), Foomani's (2015) study resonates with the finding of van Zeeland (2014) in that both researchers found vocabulary knowledge to be an important factor in lexical inferencing in listening.

Working Memory

Working memory, defined as the ability to mentally store and manipulate information relevant to a task (Baddeley, 2003), has been proposed as a vital factor that affects listening comprehension. However, relevant studies have produced mixed results. The results of some studies (e.g., Andringa et al., 2012; Vandergrift & Baker, 2015, 2018; Wallace & Lee, 2020) have indicated that working memory has no influence, a weak influence, or an indirect influence on L2 listening comprehension, while the results of other studies (Masrai, 2019; Satori, 2021; Vafaee & Suzuki, 2020) have suggested that working memory plays a significant role in listening comprehension. Vafaee and Suzuki (2020) argued that this inconsistency in the findings seems to lie in the differences in the working memory measures utilized.

Sentence Level Factors

In this section, I will cover what has been found in studies related to sentence level factors, namely (1) lexical density and (2) sentence length.

Lexical Density

Lexical density is measured as the number of content words (i.e., verbs, including infinitives and gerunds; nouns; adjectives; adverbs) per total words in the target text and seems to have an impact on listening comprehension (Revesz & Brunfaut, 2013). Investigating the influence of different lexical aspects on listening comprehension, Revesz and Brunfaut (2013) found that lexical corpus-based frequency and lexical density have a significant impact on listener performance, which warranted the inclusion of this variable in this study.

Sentence Length

One aspect of text-input factors that should be taken into account when considering listening difficulty is passage length. Intuitively, it is reasonable to think that the longer the sentences are, the harder it will be to comprehend them because of the increased working memory processing load. However, the findings in the literature so far have been mixed. Freedle and Kostin (1996), as well as Nissan et al. (1996), Kostin (2004), and Revesz and Brunfaut (2013) found that passage length had no effect on listening comprehension. On the

other hand, Moyer (2006), Rupp et al. (2001), as well as Henning (1991), reported opposite results, indicating that passage length does influence listening performance. The reason for the conflicting evidence might be because passage length is associated with other factors such as passage type, modality, the number of items per passage, the syntactic complexity of sentences, the number of words surrounding the key information, information redundancy, and information density. Therefore, when operationalized, the construct can easily be confounded with other variables. In addition, previous researchers have employed various ways to measure passage length, which is likely one reason for the mixed outcomes (see Bloomfield et al., 2010 for a detailed discussion). In this study, I explored the effects of sentence length in relation to other possible confounding variables such as working memory and lexical density.

Lexical Level Factors

In addition to person and sentence-level factors, lexical factors are also likely to influence inferencing success. In this section, I will discuss what has been found in studies related to the lexical level factors, namely (1) L1 similarity and (2) semantic transparence.

L1 Similarity

One characteristic inherent in the target idioms that have been investigated in L2 multi-word unit (MWU) comprehension/processing studies is the existence of the same or similar conceptual mappings in the L1 (i.e., the association between the literal expression and the metaphorical meaning). The level of similarity can be examined by two criteria: (a) whether a similar literal expression exists in the L1 and (b) whether the literal expression has a similar metaphorical meaning in the L1. For instance, an English idiomatic expression *open one's heart* has a similar expression in Japanese, *kokoro wo hiraku*, which has the same metaphorical meaning of “to share one's deepest or most intimate emotions, thoughts, or secrets.” On the other hand, *break one’s heart* does not have an exact literal counterpart although there is a similar metaphorical expression in Japanese: *kokoro wo kizutsukeru*, which literally means “damage/scar one’s heart” and has the metaphorical meaning of “causing one to feel great sadness.” However, unlike the previous two examples, the idiomatic expression *eat one’s heart out* does not have a literal or metaphorical counterpart in Japanese. Therefore, for Japanese learners of English, it is likely that the meaning of *open*

one's heart is easier to guess than *break one's heart*, and guessing the meaning of *eat one's heart out* would be most challenging.

According to Ibarretxe-Antuñano (2013), to understand an idiom means understanding its metaphorical meaning, which is deeply rooted in culture. Therefore, if the conceptual metaphor that underpins an English idiomatic expression has some overlap with the conceptual metaphor in one's L1 culture, it should be easier for learners to understand the English idiom and, in fact, some studies support this claim (e.g., Carrol et al., 2016; Charteris-Black, 2002; Laufer, 2000; Türker, 2016, 2019). In addition, L1 similarity when comprehending unfamiliar idiomatic expressions has been investigated by Irujo (1986a), who reported that L1 and L2 semantic similarity is important, especially when inferring idiomatic meanings of English expressions. She found that the meaning of English idioms that have metaphorically equivalent idioms in one's native language can be inferred more easily than those without such equivalents.

Semantic Transparency

Researchers in the field of L2 idioms have examined the effect of the characteristics inherent in the target idioms themselves. One such characteristic is semantic transparency. Semantic transparency is defined as the degree to which the literal and metaphorical meanings of the expression are related and associated (Wulff, 2008). For instance, comparing the two idiomatic expressions, *kick the bucket* and *twist one's arm*, the former has more distance between literal and figurative meanings than the latter. In other words, *kick the bucket* is less transparent and therefore more difficult to guess the metaphorical meaning from the literal meanings of the idiom components than *twist one's arm*. Semantic transparency is considered an influential factor for L1 idiom comprehension and development (Nippold & Rudzinski, 1993; Nippold & Taylor, 2002). Some studies show that the same applies to L2 idiom comprehension; the higher the level of semantic transparency of the target idioms, the better L2 learners process, understand, and acquire the idioms, at least in terms of their recognition ability (Liontas, 2003; Martinez & Murphy, 2011; Steinel et al., 2007).

Research questions

As I discussed in the section above, the studies in the relevant fields indicate that four person-level factors, (1) familiarity with the idiom, (2) English listening proficiency, (3)

listening vocabulary knowledge, and (4) working memory, are likely to affect the success in the lexical inferencing of English idiomatic expressions while listening. As for sentence-level factors, (1) lexical density and (2) sentence length seem to have an influential impact. At the lexical level, (1) L1 similarity and (2) semantic transparency are likely to influence inferencing outcomes. To date, there has not been a study that has investigated how these factors affect EFL learners' success in the lexical inferencing of English idiomatic expressions while listening. This study, which was conducted as part of my Ph.D. dissertation research project (Baierschmidt, 2022), aims to answer the following research questions:

1. How much does each of the person-level factors affect the probability of success of inferencing English idiomatic expressions by Japanese university EFL learners during listening?
2. How much does each of the sentence-level factors affect the probability of success of inferencing English idiomatic expressions by Japanese university EFL learners during listening?
3. How much does each of the lexical-level factors affect the probability of success of inferencing English idiomatic expressions by Japanese university EFL learners during listening?

Methods

Participants

The participants were 89 Japanese students (19 male and 70 female) whose ages ranged from 18 to 23 years old and who were majoring in English at a Japanese university. The male/female ratio of the participants was approximately 1 to 3, which reflects the female dominant student population of the department. The participants were selected using convenience sampling from four courses I taught (N = 60) and from four other courses taught by my colleagues (N = 19). These courses were composed of four compulsory courses for first-year students and four elective courses for third- and fourth-year students. In addition, I recruited 10 volunteers with a paper-based TOEFL (i.e., TOEFL ITP) score higher than 520 through a department-wide advertisement via e-mail to ensure the participant pool included a wide range of listening comprehension skill levels. The participants' TOEFL ITP scores ranged from 360 to 583 (M: 478, SD: 40).

Instruments

I used four instruments in this study: (a) an idiom task, (b) an idiom identification task, (c) a listening span test, and (d) a listening vocabulary level test.

Idiom Task

I created, validated, and administered the idiom task to assess the ability of students to infer idioms. The task was a picture identification task in which participants heard a narrative that included many idioms and chose pictures that they felt represented the story they had heard. There were two narratives, each of which was segmented into 20 sentences. In each sentence, participants heard sentences that included a single idiomatic expression that participants were assumed not to know the meaning of. While listening to the narrative, the participants were also presented with a story board that contained drawings that illustrated the gist of the scenes described in the story. For each sentence, four pictures were shown, only one of which visually represented the appropriate meaning of the target idiomatic expression. Upon hearing the story sentence, the participants were asked to choose the picture that they thought most accurately represented the part of the story they had just heard. The participants were allowed to listen to the sentence only once, although they chose for themselves when to start playing the recording of each sentence.

I selected the idioms from the Farlex Idioms and Slang dictionary (Farlex International, 2017), Street Talk 1 (Burke, 1995), McGraw-Hill's Dictionary of American Idioms and Phrasal Verbs (Spears, 2005), and Scholastic Dictionary of Idioms (Terban, 1996), as well as some online dictionaries, such as The Free Dictionary, and Macmillan dictionary. I designed this task to investigate if, based on available information, the listeners can form a mental image similar enough for them to be able to identify the most accurate representation of the meaning of the target idiomatic expression among the four candidates. The level of understanding targeted by this task is intended to reflect the ability to conduct a daily conversation while comprehending the general message sufficiently so that one can respond to the speaker if needed.

Idiom Identification Task

I included an idiom identification task to ensure that the idiom task described above included unknown or unfamiliar idiomatic expressions. In this way, I could ensure that the students would be prompted to infer the meaning of the expression, and their answers would indicate the level of their inferencing skills, not their previous knowledge level of the idiomatic expressions. The average rating score for each person was used as a familiarity score. In the idiom identification task, I asked participants to rate the 40 idiomatic expressions that appear in the idiom task in terms of their level of familiarity with the idiom on a scale from 1 to 5, with 1 indicating that the participant had never heard/seen the idiomatic expression before and 5 indicating that the participant knew the meaning of this idiomatic expression very well and could use it in actual communication. If a participant indicated that they were familiar with the expression, they were also asked to write the meaning of the idiomatic expressions in Japanese to confirm to the researcher that they knew the meaning of the idiom.

The Listening Vocabulary Level Test

I used the Listening Vocabulary Level Test (LVLT) created by McLean et al. (2015) in this study to measure participants' aural vocabulary knowledge. The test includes six parts corresponding to each of the first five 1000-word frequency levels and the Academic Word List. In this study, I administered Part 6 (Academic Word List), which includes 30 single words from the Academic Word List (Coxhead, 2000). The reason why I used only this part of the test is that it would take too long to administer the entire test within the limited time available for data collection. In addition, a pilot study I conducted a year before with students from the same population indicated that Part 6 would be an appropriate level at which to test participants' vocabulary levels.

Listening Span Test

I used a listening span test adapted from Ushiro and Sakuma (2000) listening span test in order to measure the listeners' working memory, defined as the ability to mentally store and manipulate information relevant to a task (Baddeley, 2003), specifically focusing on auditory information. To measure both the storage capability and processing skill that are at work simultaneously, the test takers are required to do two tasks at the same time. One task is

to judge whether what was heard was true or not. The other task is to remember the last word of the sentence they heard. The test is scored based on the number of target words that are recalled accurately. The test is composed of four parts under different conditions, and each part has three sets of sentences which are 10 to 12 words long. The condition that is different in each of the four parts is the number of sentences in each part, which varies from two to five.

Instrument Validation

I validated the idiom task, the Listening Vocabulary Level Test, and listening span test, using Rasch analysis by examining data produced by the instruments with WINSTEPS (Version 3.81.0). The reliability statistics of the instruments can be found in Table 1 and the validity and reliability analysis criteria, adopted from Bond et al. (2020), Boone et al. (2014), and Linacre (2007), in Table 2. Based on the criteria, all items seemed to function at an acceptable level. However, student reliability statistics imply that there may not have been enough variance among the participant group, which is understandable given their shared cultural and academic background. The lack of variability in the participants indicated by these instruments should be taken into consideration when interpreting the findings.

Table 1

Instrument Validation Results Based on Rasch Analysis

Instruments (Number of items)	Rasch person reliability	Rasch person separation	Rasch item reliability	Rasch item separation	Cronbach alpha (KR-20)
Listening Vocabulary Levels Test (30)	.69	1.49	.93	3.78	.74
Listening span test (44)	.58	1.18	.94	3.89	.72
Idiom task (40)	.75	1.72	.92	3.36	.76

Table 2

Criteria for Validity and Reliability Analysis

Aspects to be inspected	Criterion	Critical value

Reliability	Person reliability	> .80
	Item reliability	> .80
Separation	Person separation	> 2.00
	Item separation	> 3.00

Procedure

I used a cross-sectional design for this study to understand the association between possible factors that potentially influence inferencing success during listening. Due to the Coronavirus outbreak, I conducted data collection via Zoom, administering the instruments that I created using Google Forms. I collected data during regular class time in two sessions: I administered the idiom task and the idiom identification task in the first session, and the Listening Vocabulary Level Test and the listening span test in the second session. For each task, I provided participants with a link to the online task and assigned each participant to an individual breakout room where they worked on the task themselves. I required participants to share their screen with me while doing the task so that both their face and the task page on which they were working were visible to me, and I could confirm that participants were doing the tasks correctly. In addition, I asked the participants to use headphones to reduce background noise. I allowed them to do the tasks at their own pace and instructed them to ask questions or ask for help when needed. I visited each breakout room throughout the session to monitor the progress.

Analyses

After validating the data using a Rasch model as explained in the Instrument Validation section, I examined the data for descriptive characteristics and checked for linearity, multicollinearity, and outliers using IBM SPSS Statistics (Version 28). To test the linearity assumption, I employed the Box-Tidwell approach. All the variables met the criteria and indicated that they did not validate the linearity assumption. Tests to see if the data met the assumption of collinearity indicated that multicollinearity was not a concern; all the scale variables had a VIF value of less than 2, which is much smaller than the often-used cut-off point of 10. For outlier detection, a Mahalanobis distance was inspected and a cut-off value of 25 was used (Field, 2013, p. 307). Based on the chi-square test results, there were 30

outlier cases out of 3560 cases (89 participants x 40 items). After a close examination of each case, they were excluded from the analysis to avoid any undue influence on the results.

After the preliminary analyses and data preparation, I conducted a mixed-effects logistic regression analysis using R (Bates et al., 2015). I decided that a logistic regression analysis would be suitable for this study because the independent variables include a mix of continuous and categorical ones. Additionally, I incorporated mixed-effects modeling in the logistic regression analysis to account for the factors characterizing individual participants and items (Quené & van den Bergh, 2008). The dependent variable in the logistic regression analysis was the probability of correct inferences, given the modeled independent variables or covariates. Incorrect inferences were coded 0, and correct inferences were coded 1. The main eight independent variables in this study are (1) familiarity, (2) listening comprehension skills, (3) listening vocabulary size, (4) listening working memory, (5) lexical density, (6) sentence length, (7) L1 similarity, and (8) semantic transparency. The details of each variable can be found in Appendix A.

Results

I conducted a mixed-effects logistic regression analysis to investigate the relationship between eight possible target factors and the probability of successful idiom inferencing during listening by Japanese EFL learners. I entered the variables into the model simultaneously to examine the effect of each factor. Table 3 presents the descriptive statistics for all the numerical covariates investigated.

Table 3

Descriptive Statistics for All Covariates

	N	*M*	*SE*	*SD*	Minimum	Maximum
Person-level covariates						
LC	89	48.60	0.40	3.80	40.00	63.00
VK	89	1.79	0.13	1.22	-0.72	5.50
WM	89	1.40	0.09	0.87	-0.12	4.42
FamPA	89	0.46	0.04	0.35	0.00	1.40
Sentence-level covariates						

LD	40	0.40	0.01	0.08	0.27	0.60
SL	40	20.20	1.06	6.70	9.00	33.00

Note. LC = TOEFL ITP listening comprehension section score, VK = Listening Vocabulary Level Test Rasch person ability estimates in logits, WM = listening span test Rasch person ability estimates in logits, FamPA = average familiarity rating per person, LD = lexical density, SL = sentence length

The results (see Table 4) indicate that familiarity (Fam) (β = 0.13, *OR* = 1.14, *Wald* = 2.19, p = .03), listening comprehension (LC) (β = 0.10, *OR* = 1.11, *Wald* = 5.57, p < .0001), sentence length (SL) (β = 0.04, *OR* = 1.04, *Wald* = 2.18, p < .03), and listening working memory (WM) (β = 0.17, *OR* = 1.19, *Wald* = 2.18, p = .03) were significant variables. On the other hand, listening vocabulary size (VK) (β = -0.01, *OR* = 0.99, *Wald* = -0.15, p = .88), lexical density (LD) (β = 0.01, *OR* = 1.01, *Wald* = 0.90, p = .37), and semantic transparency (Trans) (β = -0.26, *OR* = 0.77, *Wald* = -1.46, p = .14) did not influence the probability of inference success in a systematic way.

Table 4

Results of Logistic Regression of the Model with All Covariates

Parameter	Fixed effects								Random effects	
							95% C.I.for EXP(B)		By subject	By item
	B	*SE*	*Wald*	*df*	*Sig.*	*Exp(B)*			*SD*	*SD*
Constant	-0.45	0.27	-1.69	1	.09	0.63	-0.98	0.07	0.44	0.69
Fam	0.13	0.06	2.19	1	.03*	1.14	0.01	0.25	—	—
LC	0.10	0.02	5.57	1	.00***	1.11	0.07	0.14	—	—
VK	-0.01	0.06	-0.15	1	.88	0.99	-0.13	0.11	—	—
WM	0.17	0.08	2.18	1	.03*	1.19	0.02	0.33	—	—
LD	0.01	0.01	0.90	1	.37	1.01	-0.01	0.04	—	—
SL	0.04	0.02	2.18	1	.03*	1.04	0.00	0.08	—	—
SIM	0.47	0.17	2.82	1	.00**	1.61	0.14	0.80	—	—
Trans	-0.26	0.17	-1.46	1	.14	0.77	-0.60	0.09	—	—

Note. Fam = familiarity rating; LC = TOEFL ITP listening comprehension section score; VK = Listening Vocabulary Level Test Rasch person ability estimates in logits; WM = listening span test Rasch person ability estimates in logits; LD = lexical density; SL = sentence length; SIM = L1 similarity rating; Trans = semantic transparency rating.

***$p < .001$ **$p < .01$ *$p < .05$

Next, I compared the model with all the variables (Model 1) and the model with non-significant variables removed (Model 2). The model comparison results indicated that the model excluding non-significant variables (Model 2) was a better model than the one including the non-significant variables (Model 1). However, sentence length (SL) was found to be non-significant in Model 2. Therefore, I compared Model 3, which included familiarity (Fam), listening comprehension (LC), listening working memory (WM), and L1 similarity (SIM), with Model 2, which included all of the variables in Model 3 as well as sentence length (SL). The result showed that including sentence length (SL) did not significantly improve the model. Therefore, I excluded sentence length (SL) from the final model. In other words, I concluded that Model 3 with familiarity (Fam), listening comprehension (LC), listening working memory (WM), and L1 similarity (SIM) should be kept as a final model. The details of the final model can be found in Table 5. Results of the model comparisons can be found in Table 6.

Table 5

Results of Logistic Regression of the Final Model (Model 3)

Parameter	Fixed effects								Random effects	
							95% C.I.for EXP(B)		By subject	By item
	B	*SE*	*Wald*	*df*	*Sig.*	*Exp(B)*			*SD*	*SD*
Constant	-0.64	0.22	-2.98	1	0.00**	0.53	-1.06	-0.22	0.44	0.74
Fam	0.13	0.06	2.20	1	0.03*	1.14	0.01	0.25	—	—
LC	0.10	0.02	6.00	1	0.00***	1.10	0.07	0.13	—	—
WM	0.17	0.07	2.41	1	0.02*	1.18	0.03	0.30	—	—

SIM	0.37	0.17	2.16	1	0.03*	1.45	0.03	0.71	—	—

Note. Fam = familiarity rating; LC = TOEFL ITP listening comprehension section score; WM = listening span test Rasch person ability estimates in logits; SIM = L1 similarity rating.

****p* < .001 ***p* < .01 **p* < .05

Table 6

Model Comparison Results

Model	*df*	AIC	Deviance (-2LL)	Chi-Square	Chi df	*Pr(>Chisq)*
Model 3	7	4431.5	4474.7			
Model 2	8	4431.5	4474.7	3.609	1	.06
Model 1	11	4431.5	4474.7	3.609	3	.50

Note. Model 1: intercept + Fam + LC +VK + WM + SL + LD + SIM + Trans + (1 | SubID) + (1 | Item); Model 2: intercept + Fam + LC + WM + SL + SIM + (1 | SubID) + (1 | Item); Model 3: intercept + Fam + LC + WM + SIM + (1 | SubID) + (1 | Item)

Discussion

In this study, I examined variables at the person-level, the sentence-level, and the lexical-level in order to investigate their influence on the probability of success of Japanese EFL university students inferring unfamiliar English idiomatic expressions during a listening task. I will discuss the results for the variables at each level in the following sections.

Person Level Factors

Among the four person-level variables (i.e., familiarity, listening comprehension, vocabulary knowledge, and working memory), familiarity, listening comprehension and working memory were found to be significant in predicting success. I included Familiarity as a controlling variable in this study since the aim of my research was to investigate the influence of various factors on the success probability of inferring *unfamiliar* idiomatic expressions in a listening task. As for listening comprehension, it was found to be an influential factor in auditory inferencing success. This result seems reasonable, as listening idiom inferencing requires listeners to understand the surrounding texts so that the information from the text can be used as clues to make sense of what the target idiomatic

expression may mean (Frantzen, 2003; Haynes, 1983; Laufer, 1997; Li, 1988). This also suggests that successful inferencing can be achieved only when successful comprehension precedes it.

The non-significant result for vocabulary knowledge contradicts the findings of some previous studies that have found it to be a significant factor for listening comprehension (Andringa et al., 2012; Mecartty, 2000; Stæhr, 2009; Vandergrift & Baker, 2015; Wang & Treffers-Daller, 2017) and listening inferencing (Foomani, 2015; van Zeeland, 2014). In van Zeeland (2014), for instance, vocabulary knowledge was found to be one of the chief factors in lexical inferencing in listening. The differing results could come from the fact that listening comprehension skills and L1 background are controlled for in this study, whereas they were not controlled for in the study by van Zeeland (2014). In the limitations and suggestions for future research section of van Zeeland's paper, these two factors were indeed mentioned as those that can produce different results (van Zeeland, 2014). Another possible explanation is that the aspect of vocabulary knowledge measured in van Zeeland's study (i.e., knowledge of literal meanings) was different from the aspect of vocabulary knowledge required for lexical inferencing during the idiom task (i.e., knowledge of figurative meanings). Additionally, judging from the relatively low Rasch person reliability estimate of the listening vocabulary size test, it can be speculated that there was not enough variance among the participants in this study, which possibly contributed to the result of vocabulary knowledge not being statistically significant.

Regarding working memory, its significant role in lexical inferencing during listening in this study contradicts with the results of some prior studies that have found that working memory has no influence, a weak influence, or an indirect influence on L2 listening comprehension (see Andringa et al., 2012; Vandergrift & Baker, 2015; Wallace & Lee, 2020). This contradiction may imply that lexical inferencing is an additional sub-process of listening comprehension that requires extra working memory capacity on top of what is used for comprehension. However, like the Listening Vocabulary Size Test, the listening span test's relatively low Rasch person reliability estimate requires that the results for working memory be interpreted cautiously.

Sentence Level Factors

The results indicate that both sentence level factors (i.e., lexical density and sentence length) were not significant factors in inferencing in this study. However, one needs to take into consideration that I limited the ranges of both lexical density and sentence length in this study to keep the difficulty level of the task to a suitable level for the participants. In addition, I wrote the passages in a conversational style, and I allowed the listeners to control the timing of when the recording of each sentence was played, giving them enough time to process each sentence before the next one was played. If the listening passages were longer with more formal academic words, and/or if the timing of the recording was not controlled by the listener, the results might differ from what I found in this study regarding the effect of sentence level factors.

Lexical Level Factors

While none of the sentence-level factors were significant, one of the lexical factors (i.e., L1 similarity) was found to be significant. The importance of lexical items, especially content words, in the decoding process for comprehension aligns with what researchers have found so far (Ross, 1997; VanPatten, 2015). The dominance of L1 similarity over lexical density and sentence length may imply that the useful information that the idiomatic component carries as clues for inferencing matters more than the percentage of content words in a sentence or the number of letters in the sentence.

Comparing the two lexical level factors, it was found that L1 similarity has an influential impact on idiom inference success probability in a positive way, while semantic transparency, operationalized through native speakers' intuitive semantic transparency judgment, did not. The results regarding L1 similarity resonate to some degree with the findings of previous studies that have suggested that linguistic and conceptual similarities assist the learner in understanding figurative expressions (e.g., Carrol et al., 2016; Charteris-Black, 2002; Irujo, 1986a; Laufer, 2000; Türker, 2016, 2019).

On the other hand, the statistically non-significant effect of semantic transparency somewhat contradicts with what has been found in previous research studies on the effects of semantic transparency on idiom and collocation processing. A possible reason for this is that most existing studies have focused on the processing of *known/familiar* idiomatic expressions or collocations, while examining processing speed as the dependent variable (e.g., Cieślicka,

2006; Cieślicka & Heredia, 2019; Gyllstad & Wolter, 2016). In other words, these studies did not focus on the inferencing of *unfamiliar* idiomatic expressions. Additionally, the participants in such studies were usually advanced learners or bilingual speakers of English instead of EFL learners with a low- or intermediate- level of English proficiency. Semantic transparency of known idiomatic expressions, which is formed after one learns the meaning of the expression, would be a different construct from the perceived semantic transparency of unfamiliar idiomatic expressions before one learns the actual figurative meaning of the expression. Therefore, it is likely that the difference in the findings in terms of the effect of semantic transparency of idioms lies in whether the learners have already developed direct links between L2 idiomatic expressions and their metaphorical meanings. If learners have already acquired such links, the semantic transparency of idioms can influence and assist the way they process the learned idiomatic expressions (Wang et al., 2020). On the other hand, if the target idiomatic expressions are unfamiliar to the learners, their meaning may not be transparent due to many plausible metaphorical interpretations of the component words or their combinations.

Limitations and Suggestions for Future Research

The participants in this study were all intermediate EFL learners majoring in English at a Japanese university. Therefore, more research is needed to generalize the findings to other learning contexts. Similarly, since I controlled many aspects of the idiom task used in this study, the ecological validity was compromised. In other words, the findings of this study may not exactly reflect the inferencing process in more natural settings. In addition, one of the important listening factors that I did not control for in this study was the phonological aspects of task input (i.e., the speaker's sex, accent, tone, speech speed, amount of reduction, length of pauses, etc.). Further research investigating the influence of phonological elements in inferencing success during listening might be beneficial to capture a more holistic view of the L2 idiom inferencing mechanism. It is also important to mention that one type of idiomatic expressions based on L1 similarity was excluded from this study, namely, the type of expressions that have a similar linguistic form but differ in meaning. The reason why I did not include this type of expression in this study is that they are known to be quite challenging for L2 learners (Charteris-Black, 2002) and I avoided such elements to design an idiom task where available information is comprehensible enough to be able to use as clues.

Additionally, the range and variety of syntactic complexity and lexical density were limited, which could be a reason why sentence level variables were not found to be statistically significant. Lastly, I did not consider contextual information used for idiom inferencing in this paper. Even though qualitative retrospective interviews were conducted as part of the larger study (Baierschmidt, 2022), due to space limitations, I will not include the results of the qualitative investigation in this paper. Information regarding how clues from the co-text and context, depth of vocabulary knowledge, as well as individuals' background knowledge affect inferencing could shed light on a more comprehensive view of how the L2 idiom inferencing process works.

Conclusion

In this study, I investigated the factors underlying successful lexical inferencing when Japanese EFL listeners encounter unfamiliar idiomatic expressions during listening. The results indicated that familiarity, listening comprehension skills, working memory, and L1 similarity were significant factors influencing participants' inferencing success. Although many researchers have investigated how quickly advanced learners process known idiomatic expressions, few researchers have looked at how low- and intermediate-proficiency learners successfully infer the meaning of unfamiliar idiomatic expressions.

The results of this study seem to suggest that whether inferencing can be an effective strategy when encountering unfamiliar idiomatic expressions depends on the listener's comprehension skills. Therefore, new idiomatic expressions should be taught explicitly to EFL learners with low-listening comprehension skills, rather than implicitly having them infer the meaning of unfamiliar idiomatic expressions. For those EFL teachers who would prefer to teach new idiomatic expressions implicitly, it is recommended that they ensure that students' listening comprehension skills are sufficient enough to understand the texts surrounding the target idiomatic expressions. In addition, teachers should be aware that their intuitive judgment about how easily the meaning of the target idiomatic expressions can be inferred based on the semantic transparency (i.e., how closely the literal meaning and the figurative meaning of the target idiom are associated) can be misleading. Instead, they should be advised to estimate the likelihood of students being able to infer the meaning of such expressions based on the existence of corresponding L1 lexical and/or conceptual counterparts. Furthermore, if target idiomatic expressions have no corresponding L1 lexical

and/or conceptual counterparts, it would be challenging for EFL learners to infer the meaning of the target expression. Thus, such idiomatic expressions should be explicitly taught by introducing the underpinning conceptual metaphors.

Though I believe this study has helped clarify some of the factors involved in idiom inferencing in listening in an EFL setting, much remains to be explored, and even more to be gained by further research in this field. Future relevant research should be conducted by overcoming the limitations of my study to gain further insight into the elaborate relationships among the factors that influence the nature and mechanisms of idiom inferencing, comprehension, and retention. I hope that the results of this study have filled in some of the missing pieces from the picture of idiomatic inferencing in an L2 and help illustrate how several factors affect the processing of unfamiliar idioms by EFL learners. Furthermore, I hope that this study will make a useful contribution to the development of effective instructional materials and teaching practices, which, in turn, will help learners gain much needed idiomatic competence during authentic and meaningful communication.

References

Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second Language Research*, *19*(4), 329–358. https://doi.org/10.1191/0267658303sr226oa

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, *62*, 49–78. https://doi.org/10.1111/j.1467-9922.2012.00706.x.

Anvari, S., & Farvardin, M. T. (2016). Revisiting lexical inferencing strategies in L2 reading: A comparison of successful and less successful EFL inferencers. *The Reading Matrix*, *16*(1), 63–77.

Baierschmidt, J. (2022). *Factors influencing Japanese university learners' inferences of unfamiliar idiomatic expressions in listening*. [Doctoral dissertation, Temple University]. http://dx.doi.org/10.34944/dspace/8290

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8. http://CRAN.R-project.org/package=lme4

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. https://doi.org/10.1038/nrn1201

Bengeleil, N. F., & Paribakht, T. S. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *Canadian Modern Language Review*, *61*(2), 225–249. http://dx.doi.org/10.3138/cmlr.61.2.225

Bloomfield, A., Wayland, S., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult?: Factors affecting second language listening comprehension*. University of Maryland. https://apps.dtic.mil/dtic/tr/fulltext/u2/a550176.pdf

Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32,* 83-110. https://doi.org/10.1017/S0267190512000050

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Burke, D. (1995). *Street talk 1*. Optima books.

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition, 38*(3), 403–443. https://doi.org/10.1017/S0272263115000492

Charteris-Black, J. (2002). Second language figurative proficiency: A comparative study of Malay and English. *Applied Linguistics, 23*(1). 104-133. https://doi.org/10.1093/applin/23.1.104

Chegeni, N., & Tabatabaeiei, O. (2014). Lexical inferencing: The relationship between number and density of lexical items and L2 learners' reading comprehension achievement. *Journal of Language Teaching and Research, 5*(2), 306–312. https://doi.org/10.4304/jltr.5.2.306-312

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research, 22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

Cieślicka, B. A., & Heredia, R. R. (2019). Figurative language in foreign/second language learning and teaching. In Abrar-ul-Hassan (Eds.), *TESOL Encyclopedia of English Language Teaching: Teaching vocabulary*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118784235.eelt0959

Cobb, T. (n.d.). Web VP Classic [an adaptation of Heatley, Nation & Coxhead's (2002) Range]. https://www.lextutor.ca/vp/eng/

Cooper, T. C. (1998). Teaching idioms. *Journal of Foreign Language Annals, 31*(2), 255–266. https://doi.org/10.1111/j.1944-9720.1998.tb00572.x

Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly, 33*(2), 233–262. https://doi.org/10.2307/3587719

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Farahani, A. A. K., & Foomani, E. M. (2015). Lexical inferencing in listening: Patterns of knowledge source use across L2 listening proficiency levels. *Journal of Applied Linguistics and Language Research, 2*(6), 39–56.

https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.836.844&rep=rep1&type=pdf

Farlex International (2017). *The Farlex idioms and slang dictionary.* Farlex International.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE Publications Ltd.

Foomani, E. M. (2015). Lexical inferencing in listening: Depth of vocabulary knowledge and listening proficiency. *International Journal of English Language Education*, *3*(2), 105–117. https://doi.org/10.5296/ijele.v3i2.8081

Frantzen, D. (2003), Factors affecting how second language Spanish students derive meaning from context. *The Modern Language Journal*, *87,* 168–199. https://doi.org/10.1111/1540-4781.00185

Freedle, R., & Kostin, I. (1996). The prediction of TOEFL listening comprehension item difficulty for minitalk passages: Implications for construct validity. *TOEFL Research Report, 56*. Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.1996.tb01707.x

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, *66*, 296-323. https://doi.org/10.1111/lang.12143

Hatami, S., & Tavakoli, M. (2012) The role of depth versus breadth of vocabulary knowledge in success and ease in L2 lexical inferencing. *TESL Canada Journal*, *30*(1), 1–21. https://doi.org/10.18806/tesl.v30i1.1123

Haynes, M. (1983). Patterns and perils of guessing in second language reading. In J. Handscombe, R. A. Orem, & B. P. Taylor (Eds), *On TESOL '83: The question of control* (pp. 163–176). https://files.eric.ed.gov/fulltext/ED275137.pdf#page=164

Henning, G. (1991). A study of the effects of variation of short-term memory load, reading response length, and processing hierarchy on TOEFL listening comprehension item performance. *TOEFL Research Reports Report 33* February 1991. Educational Testing Service. https://doi.org/10.1002/j.2333-8504.1990.tb01354.x

Hu, H. M., & Nassaji, H. (2014). Lexical inferencing strategies: The case of successful versus less successful inferencers. *System*, *45*, 27–38. https://doi.org/10.1016/j.system.2014.04.004

Ibarretxe-Antuñano, I. (2013). The relationship between conceptual metaphor and culture. *Intercultural Pragmatics*, *10*(2), 315–339. https://doi.org/10.1515/ip-2013-0014

Irujo, S. (1986a). Don't put your leg in your mouth: Transfer in the acquisition of idioms in a second language. *TESOL Quarterly*, *20*, 287–304. https://doi.org/10.2307/3586545

Irujo, S. (1986b). A piece of cake: Learning and teaching idioms. *ELT Journal*, *40*(3), 236–242. https://doi.org/10.1093/elt/40.3.236

Jackson, D. O. (2020). Working memory and second language development: A complex, dynamic future? *Studies in Second Language Learning and Teaching*, *10*(1), 89–109. https://doi.org/10.14746/ssllt.2020.10.1.5

Kaivanpanah, S., & Alavi, M. (2008). Deriving unfamiliar word meaning from context: Is it reliable? *Regional Language Centre Journal*, *39*(1), 77–95. https://doi.org/10.1177/0033688208091141

Kaivanpanah, S., & Moghaddam, M. S. (2012). Knowledge sources in EFL leaners' lexical inferencing across reading proficiency levels. *RELC journal*, *43*(3), 373–391. https://doi.org/10.1177/0033688212469219

Karlsson, M. (2019). *Idiomatic mastery in a first and second language*. Multilingual Matters. https://doi.org/10.21832/9781788922371

Kostin, I. (2004). Exploring item characteristics that are related to the difficulty of TOEFL dialogue items. *TOEFL Research Report RR-79* July 2004. Educational Testing Service. https://www.ets.org/Media/Research/pdf/RR-04-11.pdf

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (Cambridge Applied Linguistics, pp. 20-34). Cambridge University Press. https://doi.org/10.1017/CBO9781139524643.004

Laufer, B. (2000). Avoidance of idioms in a second language: Effect of L1-L2 degree of similarity. *Studia Linguistica*, *54*(2), 186–196. https://doi.org/10.1111/1467-9582.00059

Li, X. (1998). Effects of contextual cues on inferring and remembering meanings of new words, *Applied Linguistics*, *9*(4), 402–413. https://doi.org/10.1093/applin/9.4.402

Linacre, J. M. (2007). Rasch measurement and the R statistics environment. *Rasch Measurement Transactions*, *21*(1), 1095. https://www.rasch.org/rmt/rmt211m.htm

Liontas, J. I. (2003). Killing two birds with one stone: Understanding Spanish VP idioms in and out of context. *Hispania*, *86*(2), 289–301. https://pdfs.semanticscholar.org/77d8/117d5b21feb0d6d6885cd51ceda838394c4f.pdf

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, *19*(6), 741–760. https://doi.org/10.1177/1362168814567889

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly*, *45*, 267–290. https://doi.org/10.5054/tq.2011.247708

Masrai, A. (2019). Exploring the impact of individual differences in aural vocabulary knowledge, written vocabulary knowledge and working memory capacity on explaining L2 learners' listening comprehension. *Applied Linguistics Review*, *11*(3), 423–447. https://doi.org/10.1515/applirev-2018-0106

Mecartty, F. (2000). Lexical and grammatical knowledge in reading and listening comprehension by foreign language learners of Spanish. *Applied Language Learning*, *11*(2), 323–348. https://www.dliflc.edu/wp-content/uploads/2014/04/all-v11-n2.pdf

Morrison, L. (1996). Talking about words: A study of French as a second language learners' lexical inferencing procedures. *The Canadian Modern Language Review*, *53*(1), 41–75. https://doi.org/10.3138/cmlr.53.1.41

Moyer, A. (2006). Language contact and confidence in second language listening comprehension: A pilot study of advanced learners of German. *Foreign Language Annals*, *39*(2), 255–275. https://doi.org/10.1111/j.1944-9720.2006.tb02265.x

Na, L., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, *16*(1), 33–42. https://doi.org/10.1177/003368828501600103

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*(3), 387–401. https://doi.org/10.1111/j.1540-4781.2006.00431.x

Nation, P. & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. http://www.jalt-publications.org/archive/tlt/2007/07_2007TLT.pdf

Nippold, M. A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech and Hearing Research*, *36*, 728–737. https://doi.org/10.1044/jshr.3604.728

Nippold, M. A., & Taylor, C. L. (2002). Judgments of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech and Hearing Research, 45*. 384–391. https://doi.org/10.1044/1092-4388(2002/030)

Nissan, S., DeVincenzi, F., & Tang, K. L. (1996). An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension. *TOEFL Research Reports 51* February 1996*)*. Educational Testing Service. https://files.eric.ed.gov/fulltext/ED400328.pdf

Paribakht, T. S. (2005), The influence of first language lexicalization on second language lexical inferencing: A study of Farsi-speaking learners of English as a foreign language. *Language Learning*, *55,* 701-748. https://doi.org/10.1111/j.0023-8333.2005.00321.x

Paribakht, T., & Wesche, M. (1999). Reading and "incidental" L2 vocabulary acquisition: An introspective study of lexical inferencing. *Studies in Second Language Acquisition*, *21*(2), 195-224. https://doi.org/10.1017/S027226319900203X

Park, J., & Chon, Y. V. (2019). EFL learners' knowledge of high-frequent words in the comprehension of idioms: A boost or a burden?. *RELC Journal*, *50*(2), 219–234. https://doi.org/10.1177/0033688217748024

Pulido, D. (2007). The effects of topic familiarity and passage sight vocabulary on L2 lexical inferencing and retention through reading. *Applied Linguistics, 28*(1), 66–86. https://doi.org/10.1093/applin/aml049

Qian, D. D. (2005). Demystifying lexical inferencing: The role of aspects of vocabulary knowledge. *TESL Canada Journa*l, *22*(2), 34–54. https://doi.org/10.18806/tesl.v22i2.86

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. https://doi.org/10.1016/j.jml.2008.02.002

Revesz, A., & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition*, *35*(1), 31–65. https://doi.org/10.1017/S0272263112000678

Ross, S. (1997). An introspective analysis of listener inferencing on a second language listening test. In G. Kasper & E. Kellerman (Eds.), *Communication strategies: Psycholinguistic and sociolinguistic perspectives* (pp. 216–237). Longman.

Rupp, A. A., Garcia, P., & Jamieson, J. (2001). Combining multiple regression and CART to understand difficulty in second language reading and listening comprehension test items. *International Journal of Testing, 1*(3 & 4), 185–216. https://doi.org/10.1080/15305058.2001.9669470

Satori, M. (2021). Effects of working memory on L2 linguistic knowledge and L2 listening comprehension. *Applied Psycholinguistics*, *42*, 1313–1340. https://doi.org/10.1017/S0142716421000345

Spears, A. R. (2005). *McGraw-Hill's dictionary of American idioms and phrasal verbs*. McGraw-Hill.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, *31,* 577–607. https://doi.org/10.1017/S0272263109990039

Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. *Studies in Second Language Acquisition*, *29*(3), 449–484. https://doi.org/10.1017/S0272263107070271

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. C. McKeown & M. E. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 89–105). Lawrence Erlbaum Associates, Inc.

Terban, M. (1996). *Scholastic dictionary of idioms.* Scholastic.

Türker, E. (2016). The role of L1 conceptual and linguistic knowledge and frequency in the acquisition of L2 metaphorical expressions. *Second Language Research*, *32*(1), 25–48. https://doi.org/10.1177/0267658315593336

Türker, E. (2019). Idiom acquisition by second language learners: The influence of cross-linguistic similarity and context. *The Language Learning Journal*, *47*(2), 133–144. https://doi.org/10.1080/09571736.2016.1221441

Ushiro, Y., & Sakuma Y. (2000). Modifying reading and listening span tests for group testing. *JLTA Journal*, *3*, 67–82. https://doi.org/10.4992/jjpsy.82.150

Vafaee, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, *42*(2), 383–410. https://doi.org/10.1017/S0272263119000676

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*, 390–416. https://doi.org/10.1111/lang.12105

VanPatten, B. (2015). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition* (2nd ed.) (pp. 113–135). Routledge.

van Zeeland, H. (2014). Lexical inferencing in first and second language listening. *The Modern Language Journal*, *98*(4), 1006–1021. https://doi.org/10.1111/modl.12152

Wallace, M. P., & Lee, K. (2020). Examining second language listening, vocabulary, and executive functioning. *Frontiers in Psychology*, *June*(11), 1–14. https://doi.org/10.3389/fpsyg.2020.01122

Wang, X. (2020). *Applying cognitive linguistics to second language idiom learning.* [Doctoral dissertation, Victoria University of Wellington]. http://researcharchive.vuw.ac.nz/handle/10063/9153

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139–150. https://doi.org/10.1016/j.system.2016.12.013

Wang, X., Boers, F., & Warren, P. (2020). Using literal underpinnings to help learners remember figurative idioms: Does the connection need to be crystal clear?. In Ana P. María & A. Rafael (Eds.), *Metaphor in foreign language instruction* (pp. 221-240). De Gruyter Mouton. https://doi.org/10.1515/9783110630367-011

Wesche, M. B., & Paribakht, T. S. (2010). *Lexical inferencing in a first and second language: Cross-linguistic dimensions*. Multilingual Matters.

Wulff, S. (2008*). Rethinking idiomaticity: A usage-based approach.* Continuum.

Yin, Z. (2013). The success of Chinese EFL learners'lexical inferencing and the subsequent vocabulary knowledge acquisition. *TESOL Journal*, *8*(1).

Appendix A

Details of Variables and Instruments

Variable type	Variable	Instrument	Measurement
Dependent variable	Lexical inference skill for idiomatic expressions	Idiom task (40 items)	The outcome of participants' Idiom task answer (correct = 1 / incorrect = 0)
Covariates (Person level)	Familiarity	Idiom identification task	The familiarity rating (1-5) for each person and each item
	Listening comprehension	TOEFL ITP listening section	Participants' TOEFL ITP listening section score (mean centered)
	Listening vocabulary size	The Listening Vocabulary Levels Test (LVLT) created by McLean et al. (2015)	Participants' LVLT Rasch person ability estimates
	Listening working memory	Listening span test	Participants' listening span test Rasch person ability estimates
Covariates (Item level)	Lexical density	Counting (using Web VP Classic created by Cobb)	Content words (i.e., verbs, nouns, adjectives, adverbs) per total words in the target sentence. Number of verb phrases (i.e., infinitive forms, adjective or adverbial past participles, and gerunds) is included as verbs.
	Sentence length	Counting (using Web VP Classic created by Cobb)	Total number of words in the target sentence
	L1 similarity	Japanese NS intuitive judgment for the 40 target idiomatic expressions	Rating of 1-3

	Semantic transparency	English NS intuitive judgment for the 40 target idiomatic expressions	Rating of 1-3