Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*, *55*(3), 1011-1023. doi:https://doi.org/10.1002/tesq.3055

In search of the optimal mode of input for the acquisition of formulaic expressions

Phoebe Lin

The Hong Kong Polytechnic University

phoebe.lin@polyu.edu.hk

http://orcid.org/0000-0001-7175-2099

https://www.researchgate.net/profile/Phoebe-Lin-2

Formulaic expressions (FEs) subsume all types of lexicalised word combinations including idioms, conversational routines, sayings, proverbs, binomials and collocations. These expressions seem to be difficult for EFL learners to master that they have been called the last hurdle to attaining nativelike performance (Spöttl & McCarthy, 2004). Often learners know very few FEs (Nguyen & Webb, 2017), and struggle with noticing them in texts (Littlemore, Chen, Koester, & Barnden, 2011; Martinez & Murphy, 2011).

Mode of input has been identified as a factor affecting FE acquisition (see Boers, 2020, for a review). In fact, spoken input is thought to be more conducive to FE acquisition than written input (Lin, 2012). The proposed advantage of aural input can be explained by considering the impact of the prosodic features inherent in speech that assist listeners in noticing them in discourse. These prosodic features include tendencies to: align with intonation unit boundaries (Lin, 2013, 2018b) and/or pauses (Wray, 2004), have a broad focus (Ashby, 2006), have a lower likelihood of receiving sentence stress (Lin, 2010; 2018b) or containing internal pauses (Wray, 2004), and be articulated faster due to their discourse functions (Lin, 2010; 2018b) and phonological reduction (Bybee, 2002). Together, these salient prosodic features facilitate the noticing and holistic processing of FEs (Lin, 2010; 2012). It is worth investigating whether prosodic features represent the only explanation for the advantage of aural input over visual input in the acquisition of FEs.

While studies investigating the effect of input modality on word learning abound (e.g., Brown et al., 2008; Vidal, 2011), FE learning studies remain scarce. Webb and Chang (2020) compared L2 collocation learning from graded readers in a between-subjects design and found that reading-while-listening generated significantly greater gains than listening, followed by reading. Lin (under review) compared computer-assisted FE learning from YouTube videos versus reading within-subject and found no significant difference in gains between the two conditions. Apparently, research results seem to vary depending on the types of FEs investigated, the treatment, input texts, measurements and subject backgrounds. Given these inconclusive findings, further research is necessary to explore other methods and other types of FEs.

This comparison of how L1 and EFL speakers respond to aural versus visual input has arisen from research suggesting that while L1 acquisition is aural input-dominant, EFL acquisition tends to be visual input-dominant which fundamentally affects FE acquisition and processing (Carter, 1998; Lin, 2012).

Number of repetitions is another factor affecting FE gains. Webb et al. (2013) found that collocation knowledge gains increased with the number of repetitions in context, with the

highest productive and receptive test scores at repetitions=15. Pellicer-Sanchez (2017), however, found no difference between 4 and 8 repetitions in the incidental learning of collocations. This discrepancy, together with a noticeable gap in empirical evidence evaluating the effect of repetition on FE learning out of context, justifies the need for this study.

To investigate the effect of input mode on FE learning while controlling for confounding variables arising from learning in context, a controlled experiment was conducted involving the learning of novel phrases presented both aurally and visually out of context. The focus was binomials. Subjects' form recall was assessed immediately after incidental learning and compared within-subject. The study aimed to address whether:

1) the outcome of out-of-context FE acquisition differed by input mode;
2) EFL learners and L1 English speakers differed in how well they acquired FEs from aural and visual input; and
3) the number of repetitions affected acquisition.

METHOD

Participants

Participants were 182 EFL learners (66 male, 116 female, aged 18-30, mean age: 21.7) and 30 L1 English speakers (7 male, 23 female, aged 17-32, mean age: 21.3). All were recruited at a Hong Kong university. The EFL learners were L1 Cantonese speakers with intermediate to advanced levels of English proficiency and no study-abroad experience.[1] The L1 English speakers were exchange students from the UK, the US, Canada and Australia.

Instrument

Two comparable lists of 18 novel, 3-word phrases of the form 'A and B' were created based on the following criteria: 1) all binomials should consist of three easy words known by all participants; 2) each list should have an equal number of syllables; 3) each list should include an equal number of phrases from each part-of-speech; 4) there should equal numbers of alliterated and non-alliterated phrases in each list; 5) phrases should be rare with frequencies of below 7 in the 520 million-word Corpus of Contemporary American English (COCA); 6) there should be no statistically significant frequency differences between the lists according to COCA; 7) there should be no statistically significant bigram frequency difference between the lists according to COCA; 8) the A and B words should belong to the same semantic category according to the UCREL Semantic Analysis System (Archer, Wilson, & Rayson, 2002); and 9) the A and B words should not be word associates according to USF Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) and the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy, & Piper, 1973). Table 1 shows the two lists of novel, 3-word phrases.

	Input mode	Part-of-speech	Alliteration	Repetitions
distant and violent	visual	adjective	N	4

[1]The majority of the EFL learners (i.e., 161) sat the HKDSE English language paper. The number of learners who got 5**, 5*, 5, 4, 3, 2, 1 and pending are, 7, 4, 33, 62, 41, 6, 1 and 2 respectively. These scores were benchmarked against IELTS scores (see https://www.hkeaa.edu.hk/DocLibrary/MainNews/press_20130430_eng.pdf). A 4 in the HKDSE English paper for example, which the majority of the EFL learners in this study received, is equivalent to an IELTS score of 6.31-6.51. The remaining 21 learners sat the older HKALE or HKCEE English language papers. Only 13% (i.e. 24) also attempted IELTS, mostly receiving an overall score of 6.5 or 7.

silent and sleepy	visual	adjective	Y	4
active and noisy	visual	adjective	N	6
hostile and hopeless	visual	adjective	Y	6
awkward and foolish	visual	adjective	N	8
smelly and sensitive	visual	adjective	Y	8
lawyer and colleague	visual	noun	N	4
movement and mission	visual	noun	Y	4
captain and keeper	visual	noun	N	6
demand and duty	visual	noun	Y	6
structure and balance	visual	noun	N	8
radio and response	visual	noun	Y	8
connect and engage	visual	verb	N	4
pretend and proceed	visual	verb	Y	4
gather and expand	visual	verb	N	6
divide and disturb	visual	verb	Y	6
admit and confirm	visual	verb	N	8
promise and perform	visual	verb	Y	8
risky and useless	aural	adjective	N	4
rotten and remote	aural	adjective	Y	4
peaceful and hopeful	aural	adjective	N	6
decent and diverse	aural	adjective	Y	6
exact and formal	aural	adjective	N	8
actual and absolute	aural	adjective	Y	8
conflict and witness	aural	noun	N	4
friendship and favour	aural	noun	Y	4
platform and schedule	aural	noun	N	6
sample and signal	aural	noun	Y	6
approach and culture	aural	noun	N	8
assault and allowance	aural	noun	Y	8
propose and convince	aural	verb	N	4
advise and assist	aural	verb	Y	4
select and inform	aural	verb	N	6
define and declare	aural	verb	Y	6
ignore and destroy	aural	verb	N	8
react and repeat	aural	verb	Y	8

Table 1 The list of phrases presented to participants

Procedures

Participants were presented with both lists to compare FE acquisition from aural and visual input within-subject. The order of presentation for different types of input was counterbalanced.

In the incidental learning phase, the novel phrases were presented through a 54-item same-different task. Participants responded by pressing a button to decide whether two phrases, presented sequentially, were the same (see Figure 1).[2] To investigate how the number of exposures affected memory of a phrase, a third of the phrases in each list were repeated 4, 6 and 8 times respectively. Participants were unaware they would be assessed later.

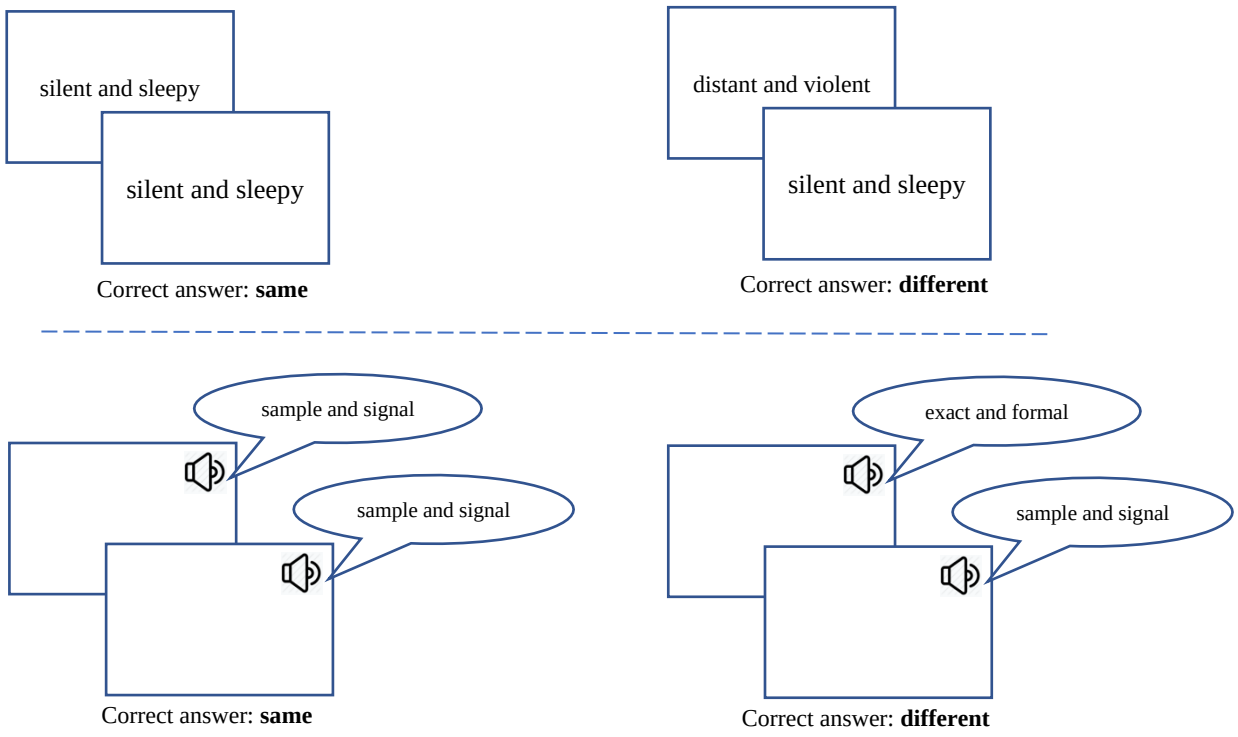

Figure 1 Schematic representation of the visual (top) and aural (bottom) stimuli in the same-different task

In the testing phase, participants completed an 18-item form recall task. They had to complete the phrases within 5 seconds either after seeing the first two words plus the first two letters of the third word (e.g., distant and vi___) or after hearing the first two words plus the first two phonemes (e.g., /ˈrɪs.ki ən juː/). The mode of presentation for the questions and the responses was consistent with the corresponding modes in the same-different task. Thus, aural responses were required following aural input whereas typed responses were required following visual input.

Both the same-different task and the form recall task were administered using the stimulus presentation software *Paradigm* (Perception Research Systems. 2007). Responses were scored

[2] The learning phase was considered incidental because learners were not warned about the test.

manually with 1 point for each correct and 0 for each incorrect answer. Typos in written responses (e.g., *vioelent*) were tolerated.

RESULTS

EFL learners

The accuracy of the responses in the same-different task was very high (visual input condition: 98.32%, aural input condition: 97.40%), which indicates participants' engagement in the exposure phase.

In the form recall test, EFL learners scored higher in the aural input condition (M=9.90, SD=2.96) than the visual input condition (M=8.30, SD=3.63). The Shapiro-Wilk normality test indicated that paired differences were normally distributed (W=0.99, p=0.19). A paired samples t-test confirmed that form recall accuracy was significantly higher for phrases presented aurally (t=4.70, df=181, p=5.20e-06, see Figure 2). The effect was between small and moderate (Cohen's d=0.35) according to Cohen (1992).

L1 speakers

The accuracy of the responses in the same-different task was also very high (visual input condition: 97.41%, aural input condition: 99.14%)

In the form recall test, L1 speakers also scored higher in the aural input condition (M=12.53, SD=2.36) than the visual input condition (M=7.53, SD=2.85). The Shapiro-Wilk normality test indicated that paired differences were normally distributed (W=0.94, p=0.11). A paired samples t-test confirmed that form recall accuracy was significantly higher for phrases presented aurally (t=8.35, df=29, p=3.34e-09, see Figure 2). The effect was large (Cohen's d=1.52).

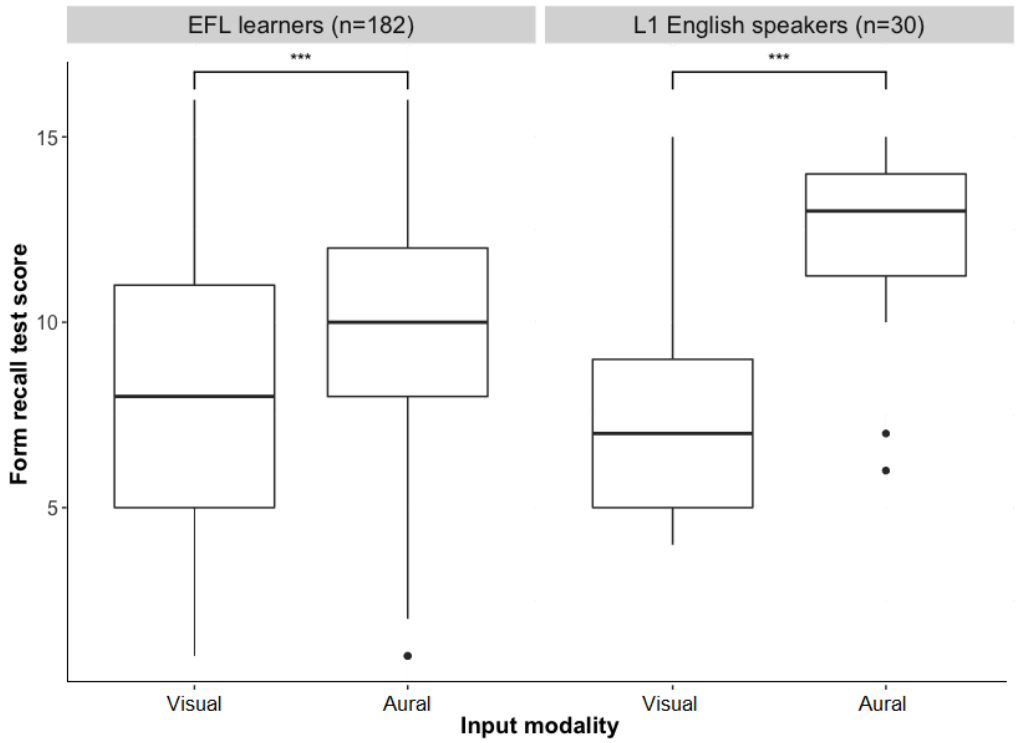

Figure 2 EFL learners' and L1 speakers' form recall test scores following visual and aural input

EFL learners versus L1 English speakers

A comparison of the EFL learners' and L1 English speakers' performances in the test revealed an interesting pattern. The mean test score was highest for the L1 speaker aural input condition (M=12.53) and lowest for the L1 speaker visual input condition (M=7.53). Thus, when acquiring phrases in their L1 from visual input, L1 speakers performed worse than the EFL learners when acquiring phrases from visual or aural input in their foreign language (see Figure 2).

Further analyses were conducted to establish whether the advantage of aural input was statistically significant and whether it differed between the two groups.[3] The size of the advantage (i.e., aural - visual) was 5 points in the L1 speaker group and 1.60 in the EFL learner group. A two-way independent samples t-test with Welch correction indicated that the aural input advantage was significantly larger (t=4.93, df=50.09, p=9.58e-06) for the L1 speaker group. The effect was large (Cohen's d=0.95).

The effect of the number of repetitions on FE acquisition

Table 2 shows the breakdown of the form recall test scores by the number of times phrases were presented. In the visual condition, the mean test score increased with the number of repetitions. However, in the aural condition, the mean form recall test score was highest at repetitions=6. At repetitions=6, the difference in test scores between the aural and visual conditions also reached the maximum.

To test whether the observed effect of the number of repetitions on phrase learning was statistically significant, a Friedman test was conducted, with difference in scores between aural and visual input as the dependent variable and number of repetitions as a categorical, individual variable.[4] The Friedman test results showed that the form recall test scores differed significantly with the number of repetitions ($\chi^2(2)$=32.2, p=9.95e-8), with a small effect size (Kendall W=0.08). A pairwise Wilcoxon signed-rank test (with continuity correction and Bonferroni adjustment of the p-value) between groups revealed statistically significant differences in the form recall test scores between repetitions=4 and repetitions=6 (Z=4737, p=1.87e-05, p_{adj}=5.61e-05), and between repetitions=6 and repetitions=8 (Z=12570, p=1.30e-08, p_{adj}=3.90e-08). The difference in scores between repetitions=4 and repetitions=8 was non-significant (Z=7470, p=0.052, p_{adj}=0.158, see Figure 3).

		Mean form recall test scores					
		Aural		Visual		Difference (aural-visual)	
repetitions	N	Mean	SD	Mean	SD	Mean	SD
4 (6 items)	212	3.14	1.33	2.56	1.3	0.58	1.85
6 (6 items)	212	3.96	1.27	2.77	1.44	1.18	1.94
8 (6 items)	212	3.18	1.3	2.86	1.54	0.321	2

[3] Aural input advantage (or spoken input advantage) here refers to the difference between the test score in the aural input condition and the test score in the visual input condition.

[4] The Friedman test was conducted because the normality assumption of the one-way repeated measures ANOVA was not met, as indicated by the Shapiro-Wilk normality test (W=0.98, p=8.30e-09).

Table 2 Breakdown of form recall test scores by number of repetitions

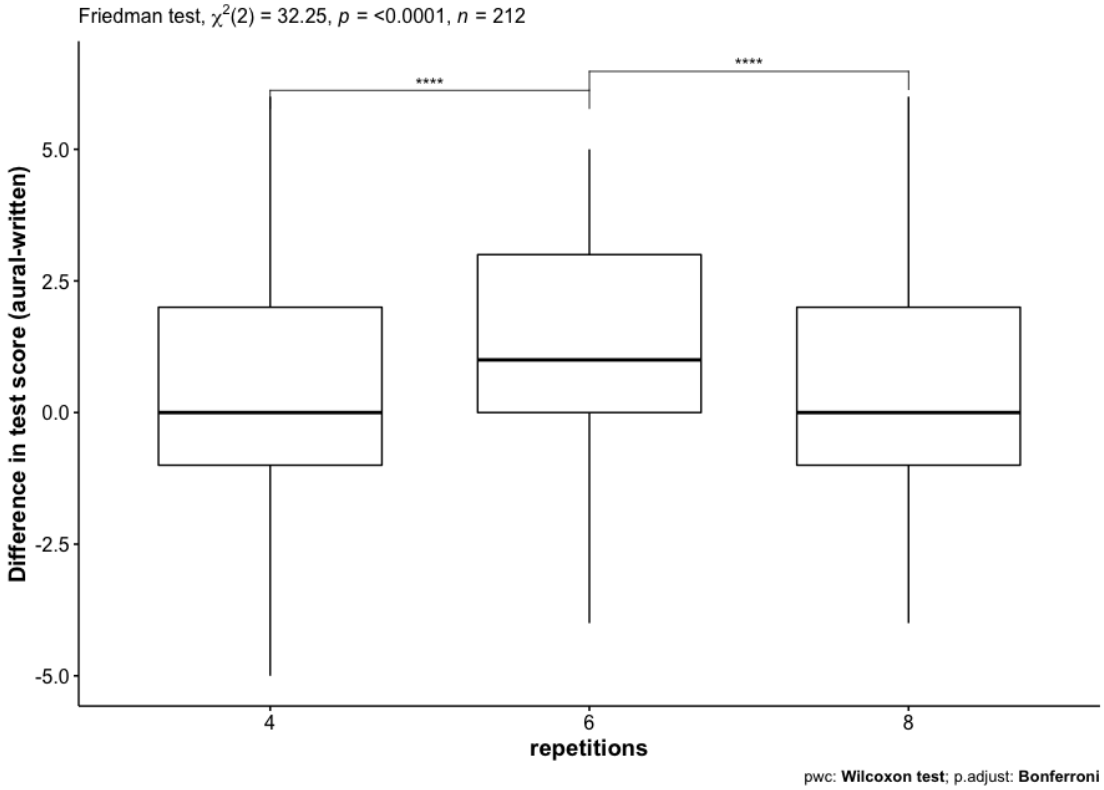

Figure 3 The difference in test score following aural versus visual input

DISCUSSION AND CONCLUSION

The results confirm that input modality affects English phrase learning. Spoken input appears more conducive to FE learning than written input, for both L1 and EFL learner groups, although the magnitude and impact of the aural input advantage were significantly larger for the L1 speaker group. The number of repetitions had a statistically significant impact on the magnitude of the aural input advantage, but the relationship between the two variables was not directly proportional.

Spoken input for English phrase learning

Webb and Chang (2020) found reading-while-listening to graded readers generated significantly greater gains in collocation knowledge than reading alone. This study also observed the facilitative effects of hearing FEs, using an alternative research design (i.e., novel phrases learned out of context and within-subject comparison).

These results can be explained from prosodic, evolutionary and psycholinguistic perspectives. As Lin (2012) suggests, aural input may have an advantage over visual input because certain prosodic features make FEs more noticeable. This prosodic effect is especially pronounced in discourse contexts, where intonation contours coincide with FE boundaries (see Lin, 2018a, 2018b). However, since the novel binomials were presented in isolation, the present results can only uncover the effect of the tonicity, tonal and rhythmic aspects of the prosodic salience of FEs on how learners notice, memorise and learn them.

Humans are better at learning through auditory input because speech is the most primitive form of communication. Humans are biologically predisposed to rely on prosodic cues to learn languages (see Lin, 2012, for a review). Even in the womb, foetuses listen to the prosody of their L1. As children grow, prosodic cues continue to guide learning of syntax and FEs. In contrast, the processing of written input is an acquired skill.

Psychologically, the processing advantage of aural over visual input could be explained by working memory. According to Baddeley (2006), verbal short-term memory is phonological in nature. Aural input can be processed within the phonological loop directly, but written input must first be converted into phonological representations through subvocal naming. Thus, the processing pathway for written input is not only more indirect, but the strength of the memory of written input (and performance in the subsequent recall test) also depends on how much the person covertly rehearses the written signals to convert them into phonological representations. Previous studies show that this subvocal naming process can be easily disrupted (see discussions of articulatory suppression, Baddeley, 2006).

The role of speech prosody in L1 and L2 FE acquisition

The discovery of a significantly larger aural input advantage in the L1 speaker group was surprising. The inferior performance of the L1 speaker group in the form recall test following visual input (compared with the EFL group in both conditions) indicates that aural input processing is deeply ingrained in the L1 system. Therefore, the L1 speakers took a bigger hit when given written input. This could be explained by the fact that L1 acquisition and L1 use is predominantly spoken. As Carter (2012, personal communication) estimates, the linguistic encounters of an ordinary L1 speaker are 90% spoken. The situation facing the learners in the present study, however, is completely different, given their text-dominant EFL curriculum (see Lin, 2012, 2018a).

The effect of repetitions on FE learning

This study contributes new empirical evidence to the literature on FE (phrase) learning and the effect of repetitions on L2 vocabulary learning. Webb et al. (2013) noted a 12% form recall rate at 5 repetitions when collocations were learned in context. Using different tasks, our participants recalled 52.33% (=3.14÷6×100%) and 42.67% (=2.56÷6×100%), respectively, of the items presented aurally and visually out of context at repetitions=4. At repetitions=6, the immediate recall rate increased to 66% (=3.96÷6×100%) for aural input and 46.17% (=2.77÷6×100%) for visual input. Importantly, this study confirms that number of repetitions affects FE learning.

Nevertheless, the number of repetitions needed for optimal FE learning warrants further investigation. According to the present learning and assessment task design, optimal FE learning was achieved at repetitions=6. This unexpected advantage for 6 repetitions could be evidence that FE learning depends not only on the number of repetitions, but also on the interaction between repetitions and other factors (Uchihara et al., 2019).

This study's findings have highlighted the value of spoken input in L2 FE learning. Pedagogically, it provides some empirical groundwork for reconsidering the balance between spoken and written input in EFL curricula. Although, written input allows learners more time to engage with new vocabulary, especially its form and composition, when the focus shifts to learning the syntagmatic relationships between known English words, spoken input seems to be superior both for L1 English speakers and EFL learners.

Lin, P. (2021). In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly, 55*(3), 1011-1023. doi:https://doi.org/10.1002/tesq.3055

Acknowledgement
This project was funded by the Hong Kong Research Grants Council (Ref: 256121/16H and 156001/21H). The author would like to thank the editor and the anonymous reviewers for their valuable comments.

About the author
Phoebe Lin is an assistant professor and a member of the Research Centre for Professional Communication in English, The Hong Kong Polytechnic University. Her research interests include vocabulary acquisition, speech prosody, experimental research, second language acquisition, corpus linguistics, natural language processing, data science, crowd-sourcing and computer-assisted language learning. https://orcid.org/0000-0001-7175-2099

REFERENCES
Archer, D., Wilson, A., & Rayson, P. (2002). Introduction to the USAS category system. Retrieved from http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf

Ashby, M. (2006). Prosody and idioms in English. *Journal of Pragmatics, 38*, 1580-1597. https://doi.org/10.1016/j.pragma.2005.03.018

Baddeley, A. D. (2006). Working memory: An overview. In S. J. Pickering (Ed.), *Working memory and education* (pp. 1-31). London: Academic Press. https://doi.org/10.1016/B978-012554465-8/50003-X

Boers, F. (2020). Factors affecting the learning of multiword items. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 143-157). London: Routledge. https://doi.org/10.4324/9780429291586-10

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20*(2), 136-163.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215-221. https://doi.org/10.1017/S0272263102002061

Carter, R. (1998). *Vocabulary: Applied linguistic perspectives* (2 ed.). London: Allen and Unwin. https://doi.org/10.4324/9780203270110

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159. https://doi.org/10.1037//0033-2909.112.1.155

Ellis, N. C. (1995). Psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning, 8*, 103-112. https://doi.org/10.1080/0958822940080202

Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. J. Aitken, R. W. Bailey, & N. Hamilton-Smith (Eds.), *The computer and literary studies*. Edinburgh: Edinburgh University Press.

Lin, P. (2010). The phonology of formulaic sequences: A review. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 174-193). London: Continuum.

Lin, P. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics, 33*, 342-347. https://doi.org/10.1093/applin/ams017

Lin, P. (2013). The prosody of idiomatic expressions in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics, 18*, 561-588. https://doi.org/10.1075/ijcl.18.4.05lin

Lin, P. (2018a). Formulaic language and speech prosody. In A. Siyanova-Chanturia & A.

Pellicer-Sanchez (Eds.), *Understanding formulaic language: A second language acquisition perspective* (pp. 78-94). London: Routledge. https://doi.org/10.4324/9781315206615-5

Lin, P. (2018b). *The prosody of formulaic sequences: A corpus and discourse approach*. London: Continuum.

Littlemore, J., Chen, P. T., Koester, A., & Barnden, J. (2011). Difficulties in metaphor comprehension faced by international students whose first language is not English. *Applied Linguistics, 32*(4), 408-429. https://doi.org/10.1093/applin/amr009

Martinez, R., & Murphy, V. A. (2011). Effect of frequency and idiomaticity on second language reading comprehension. *TESOL Quarterly, 45*(2), 267-290. https://doi.org/10.5054/tq.2011.247708

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. *http://www.usf.edu/FreeAssociation/*.

Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298-320. https://doi.org/10.1177/1362168816639619

Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research, 21*(3), 381-402. https://doi.org/10.1177/1362168815618428

Perception Research Systems. 2007. *Paradigm Stimulus Presentation*, Retrieved from http://www.paradigmexperiments.com

Siyanova-Chanturia, A., & Lin, P. (2018). Production of ambiguous idioms in English: A reading aloud study. *International Journal of Applied Linguistics, 28*, 58-70. https://doi.org/10.1111/ijal.12183

Spöttl, C., & McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 191-226). Amsterdam; Philadelphia: John Benjamins. https://doi.org/10.1075/lllt.9.11spo

Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning, 69*(3), 559-599. https://doi.org/10.1111/lang.12343

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*, 219-258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S., & Chang, A. C. S. (2020). How does mode of input affect the incidental learning of collocations? *Studies in Second Language Acquisition*, 1-22. https://doi.org/10.1017/S0272263120000297

Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning, 63*, 91-120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 249-268). Amsterdam; Philadelphia: John Benjamins. https://doi.org/10.1075/lllt.9.13wra