ReCALL (2022), 34: 2, 185–200
doi:10.1017/S0958344021000252

RESEARCH ARTICLE

Developing an intelligent tool for computer-assisted formulaic language learning from YouTube videos

Phoebe Lin
The Hong Kong Polytechnic University, Hong Kong SAR, China (phoebe.lin@polyu.edu.hk)

Abstract
This paper presents IdiomsTube (https://www.idiomstube.com), the first-ever computer-assisted language learning tool for facilitating the learning of English idiomatic expressions from YouTube videos. The web-based tool is designed to automatically (1) assess the speech rate and lexical difficulty level of any learner-chosen English-captioned YouTube video; (2) generate a range of vocabulary-building and revision tasks (e.g. fill-in-the-blanks, the classic spelling game hangman, pronunciation, flashcards); and (3) recommend YouTube videos based on learners' individual learning progress, class teachers' instructions, and current trends. This paper describes the app's current features in conjunction with the challenges endemic in the development of applications that automatically generate English language learning content out of internet video content. At the time of writing, the IdiomsTube app had over 8,000 registered users worldwide.

Keywords: formulaic language; vocabulary learning; mobile learning app design; YouTube; videos

1. Introduction

For many years, the viewing of television and videos has remained one of the most popular strategies for self-directed English as a foreign language (EFL) learning for the younger generation. According to our recent survey concerning the learning strategies of Hong Kong university students (Lin, 2019), 75% of the 300 respondents reported that they habitually used English videos, television, and movies to improve their English proficiency. Ten percent (i.e. 30 respondents) spent at least five hours a week doing so. Today, access to multimedia content in English is increasingly ubiquitous. Millions of authentic English videos are available for keen learners to watch at a click. These qualities represent the beginning of a new chapter for independent English language learning in which everyone can learn English independently, anywhere, and at any time.

1.1 The challenges of independent English learning from watching internet videos

Although learners of English are now often spoiled for choice in terms of the variety of online videos available for independent English learning, two problems have arisen: choice overload and lack of guidance. It has been suggested that every minute 400 hours of new content is uploaded to YouTube alone (Wojcicki, 2015). These videos include a wide variety of genres ranging from the most formal (such as National Geographic documentaries, TED Talks, speeches, lectures, and political commentaries) to the most informal videos (such as how-to videos, video blogs, pranks, comedy, chat shows, music videos, and nursery rhymes). Everyone is guaranteed to

Cite this article: Lin, P. (2022). Developing an intelligent tool for computer-assisted formulaic language learning from YouTube videos. *ReCALL* 34(2): 185–200. https://doi.org/10.1017/S0958344021000252

© The Author(s), 2021. Published by Cambridge University Press on behalf of European Association for Computer Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

find some English video input that interests them, but the sheer volume of choice available can have a negative impact on learning (see Iyengar & Lepper, 2000). As Vohs *et al.* (2008) suggest, choosing from among so many alternatives is effortful and may deplete cognitive resources. This explains why their subjects showed a deterioration of self-control, stamina, persistence in the face of failure (i.e. they gave up more easily), as well as actual performance in tasks when offered several choices. In education, teachers are also keenly aware of the need to avoid choice overload in order to ensure the effectiveness of learning. It has been advised that

> Managing choice overload is critical for both student satisfaction and performance. When students choose from options tailored to their abilities and prior knowledge levels, they are both more satisfied and learn more effectively than when presented with too many or too few choices (Volz, Higdon & Lidwell, 2019: 11).

Unfortunately, this fundamental need to manage choice overload is often overlooked in discussions surrounding how best to support independent EFL learning through the use of online videos (see also Lin, 2014; Lin & Siyanova-Chanturia, 2014).

The second, and perhaps more pressing, concern is the lack of guidance. Given the rate at which new video content is uploaded online, it is impossible for teachers to design customised activities to facilitate learning from every online video, or to provide at least a list of noteworthy English usage features in it. Learners are often left to engage in freestyle self-reflection on what they have learned. For those who have weak independent language learning skills, this uncertainty surrounding the key knowledge to be gained from each video is unhelpful.

1.2 The aims of this paper

To address these two problems, the IdiomsTube app was designed to facilitate the noticing and independent learning of formulaic expressions (FEs) in any user-chosen English-captioned YouTube video. Informed by the latest research on second language vocabulary acquisition, the app offers key features known to enhance the effectiveness of FE learning in context. These features include automatic identification of noteworthy FEs in the English subtitles of any videos posted on YouTube, automatic assessment of the difficulty of the vocabulary, the automatic generation of a multimodal concordance of system-identified vocabulary items, the automatic generation of contextual usage, form recall and verbal repetition tasks, gamification features, a learning progress tracking system for class teachers, and so on.

This paper describes the rationale, the design features, and the research underlying the development of the IdiomsTube app. It also discusses the theoretical and technical challenges facing the development of tools for computer-assisted learning of English from video streams. Since launching the app for open access use by individuals and schools in November 2018, over 8,000 learners and teachers worldwide have started using it. For empirical evidence supporting the effectiveness of the IdiomsTube app for facilitating FE learning, see Lin (under review).

2. Formulaic expressions: The key to native-like proficiency in English as a foreign language

FE is an umbrella term subsuming all types of lexicalised word combinations including idioms (e.g. *raining cats and dogs*, *spill the beans*), speech formulae (e.g. *easy does it*, *let's get on top of it*), proverbs (e.g. *an apple a day keeps the doctor away*), sayings (e.g. *better safe than sorry*, *time and tide wait for no man*), similes (e.g. *as white as snow*, *sleep like a log*), binomials (e.g. *safe and sound*, *hustle and bustle*, *kith and kin*), collocations (e.g. *ask-question, staggering-figures, prime-minister*), and so on. Although FEs are very common in daily communication, they are challenging for even advanced-level English learners (Spöttl & McCarthy, 2004).

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

Two studies in particular that provide strong empirical evidence for the difficulty of learning FEs are Webb, Newton and Chang (2013) and Pellicer-Sánchez (2017). Both studies examined incidental learning of FEs. In these studies, EFL learners were exposed repeatedly to a small set of collocates embedded within short stories and then tested on their knowledge of the items. In Webb *et al.*'s (2013) study, learners recalled only 12% of the target 18 verb-noun collocates in the immediate post-test even though they saw the items five times. In Pellicer-Sánchez's study (2017), the learning outcome was not better, with learners recalling only 7% of the six novel adjective-noun collocations in the one-week-delayed post-test despite having seen the items four times. These poor learning outcomes from incidental FE learning have prompted English language teaching (ELT) experts and researchers to explore ways of speeding up and enhancing FE learning through more explicit FE instruction.

2.1 Computer-assisted learning of formulaic expressions

Across the world, researchers are competing to solve the problems confronting second language FE acquisition. The main solution that has emerged involves enhancing the input by highlighting or glossing FEs (Bishop, 2004; Peters, 2012; Szudarski & Conklin, 2014; Wible, 2008) because learners' failure to notice the presence of FEs seems to be the source of the problem (Lin, 2012; Littlemore, Chen, Koester & Barnden, 2011; Martinez & Murphy, 2011).

Bishop (2004) and Wible (2008), in particular, pioneered the use of computer technology to increase the salience of FEs in electronic *written* texts. Bishop (2004) created a program that underlined, highlighted, and glossed FEs, whereas Wible (2008) developed a web browser plug-in that detected and listed the collocations present in any webpage that the learner visited. The development of these computer tools was motivated by the growing body of evidence supporting the effectiveness of input enhancement for FE acquisition (Bishop, 2004; Peters, 2012; Szudarski & Conklin, 2014; Wible, 2008). In fact, Peters' (2012) experiment indicated that input enhancement could lead to statistically significant gains in FE knowledge. Direct instruction, on the other hand, showed no significant impact.

While Bishop (2004) and Wible (2008) facilitated new possibilities for using computer technology to promote L2 English FE learning, these first-generation tools remain limited. First, they cannot handle spoken language or videos, despite initial evidence showing that spoken input may be more conducive to FE learning than written input (see Lin, 2021, for experimental evidence and Lin, 2012, for a discussion of the theoretical basis for this argument). Second, the tools stop at the point of input enhancement and lack components for the consolidation of FE knowledge. As Szudarski and Conklin (2014) show, consolidation is vital for the exposure to the input to produce a lasting gain in FE knowledge. In their study, significant knowledge gain was recorded immediately after the exposure to highlighted FEs, but had disappeared in the six-week-delayed post-test. Third, the tools stop at the point of learning the form and semantics of the FEs but contribute little to the learning of deeper levels of FE knowledge, such as their pragmatics or phonology.

3. IdiomsTube: A mobile and personalised web-based app for FE learning through YouTube videos

IdiomsTube (https://www.idiomstube.com) is the first computer-assisted language learning tool for facilitating the independent learning of FEs from YouTube videos (see Figure 1). It is unique in the sense that it (1) enables FE learning from extensive exposure to online videos, and (2) seeks to offer a comprehensive and effective FE learning experience with a range of personalised and adaptive FE learning tasks powered by an individualised FE learning profile database that stores learners' video preferences and learning histories.

IdiomsTube consists of three key components: *automatic assessment of video difficulty*, *automatic tasks generation*, and *video recommendations*, in addition to a range of features

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

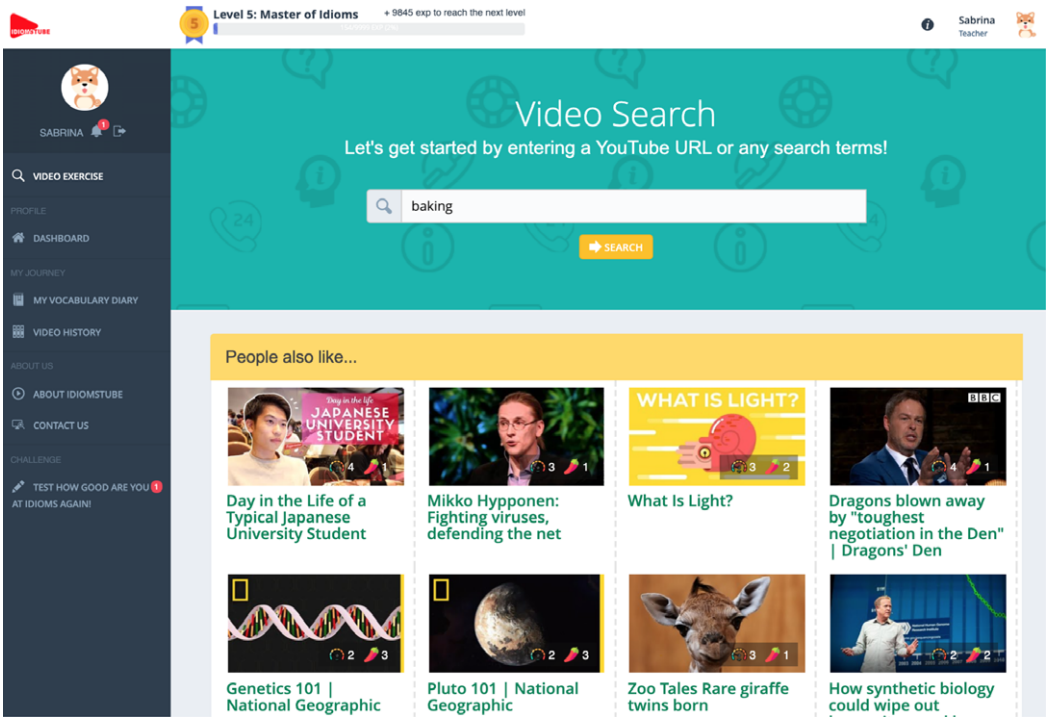


Figure 1. The homepage of IdiomsTube with the video search box and recommended videos

designed to provide a comprehensive and enjoyable FE learning experience such as gamification, bookmarks, revision flashcards, idiom-of-the-day posts, IdiomsTube Facebook and Instagram pages, and an IdiomsTube Teacher's interface.

This section gives an overview of these features. The rationale behind and the technical details concerning these features will be discussed in section 4.

3.1 Automatic tasks generation

This component automatically generates FE learning tasks for all English-captioned videos available for online streaming on YouTube.[1] Learners can find videos based on topics of personal interest (e.g. *documentary*, *TED Talks*, *baking*). The system then retrieves the 50 most recent related videos in real time from the YouTube website using the YouTube official API (see Figure 1). Once the learner has chosen a video to watch, the system automatically detects the presence of noteworthy FEs in the captions and generates three types of learning tasks based around the items, namely pre-learning, knowledge consolidation, and progress management tasks. These tasks direct learners' attention to the FEs in the videos and progressively consolidate their memory of their form and usage (see Figure 2).

3.1.1 Pre-learning tasks

The pre-learning tasks consist of two parts (i.e. the glossary and the dictionary lookup) that aim to draw learners' attention to the presence and meanings of noteworthy vocabulary items present in

[1]Due to uncertainties about the accuracy of YouTube auto-caption, the generation of learning tasks has been disabled for videos without user-uploaded English captions in the current version of IdiomsTube.

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

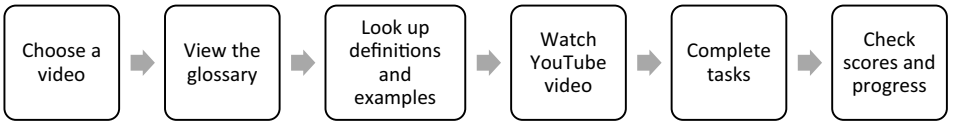


Figure 2. The workflow of learning tasks that learners need to complete on IdiomsTube

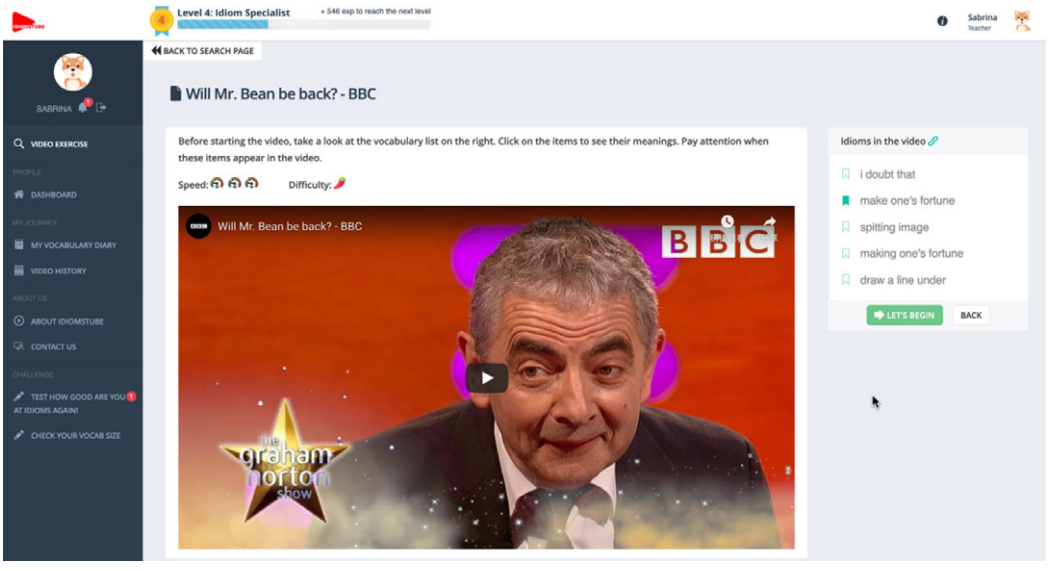


Figure 3. The video page showing a glossary of noteworthy FEs and hyperlinks to corresponding dictionary entries. Items in the glossary may be bookmarked for later reference. See also estimates of a video's difficulty and speech rate ratings above the video window

each video (see Figure 3) and enhance learners' comprehension of the video content. The IdiomsTube app is able to recognise 53,635 FEs present in YouTube videos. For every FE, the system generates a glossary with clickable hyperlinks to reputable and free online dictionaries (e.g. http://www.oxforddictionaries.com/, https://idioms.thefreedictionary.com/).

3.1.2 Consolidation tasks

The consolidation tasks consist of two parts (i.e. the exercises and the selected scene replay) that aim to raise learners' awareness of the ways the FEs have been used in context in the selected video and enhance their memory of their phonological and orthographic forms. Three types of exercises are available: fill-in-the-blanks, spelling, and pronunciation.

The fill-in-the-blanks exercise requires learners to complete the video's original lines by choosing the correct FEs (see Figure 4; note that Figures 4 to 19 are presented in the supplementary material). In this way, learners' attention can be drawn to the FEs' use in context. Should learners have difficulties recalling the correct FEs, they may click the *Hint* button to replay the exact scene in which each FE occurs.

The spelling task takes the form of the classic word game *hangman*, played in context, with the FEs as the answers learners must recall (see Figure 5). To excel at the game, learners need to accurately recall the orthographic form of each FE.

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

The pronunciation exercise requires learners to read aloud and record FEs using IdiomsTube's built-in audio recorder.[2] The FEs appear embedded within the video's original lines and are presented one by one. As shown in Figure 6, the *Hint* button is available to replay the scene in which each FE was used. These replays are a useful means of drawing learners' attention to the pragmatics of each FE, including its level of formality, its context of use, and, most importantly, the tone of voice appropriate for its pronunciation (see Lin, 2012, 2013, 2018a, 2018b, for an in-depth discussion of the special prosody of FEs).

3.1.3 Revision tasks

The revision tasks enable learners to easily monitor their FE learning progress (see Figures 7 and 8), to review their vocabulary diaries (see Figure 9) and revision flashcards (see Figure 10), and to address their needs and/or weaknesses. In the vocabulary diary, learners can see the full list of FEs encountered in previously watched videos and their mean score for each item, as well as items previously bookmarked. They then have several options. They can view the concordance lines showing the FEs used in all previously watched videos and click the hyperlinks to replay the corresponding scenes or conduct revision using the flashcards. A stack of 10 flashcards will be auto-populated and shown every time a learner logs in. Each flashcard reminds the learner of an FE learned from a previously watched video. The flip side of the card shows an example of how the FE was used in context in the video (see Figure 10).

All of the aforementioned tasks are delivered through a simple and user-friendly interface (see Figure 1) following a specified workflow (see Figure 2). As the tool runs on the internet as a web-based application, it is compatible with any operating system and can be opened using the web browser on a personal computer or on any internet-enabled mobile devices. This ubiquity of access means that learners can enjoy self-directed FE learning using the app while on short breaks or on the go.

3.2 Video recommendation

The issue of video choice overload is tackled by the video recommendation component that assists EFL learners in finding comprehensible input among the tens of millions of English-captioned YouTube videos available. The component automatically assesses the difficulty level of all retrieved videos (see Figure 1) and instantly recommends videos based on EFL learners' progress within IdiomsTube (see Figure 11). In addition to recommendations based on language difficulty, the system also incorporates recommendations by the IdiomsTube community. Videos watched by other users (i.e. the trending videos) and videos recommended by teachers who use IdiomsTube with their classes (see Figure 12) also appear on each learner's personalised front page.

3.3 Additional special features

To offer a comprehensive and enjoyable FE learning experience, IdiomsTube also provides a range of special enhancement features such as gamification, idiom-of-the-day posts, the IdiomsTube Teacher's interface and IdiomsTube Facebook and Instagram pages.

Gamification is defined as "the application of videogame techniques such as rules, rewards, and levels to real-world situations ... in order to maximize the loyalty and motivation of participants" (Chandler & Munday, 2020). In IdiomsTube, learners earn experience points (with *exp* as the short form) for each video watched and each task attempted and are promoted to a higher level when a certain points threshold has been reached. As shown in Figure 13, learners are rewarded

[2]The audio recordings are processed using Python module SpeechRecognition 3.8.1 (https://pypi.org/project/SpeechRecognition/) which connects to speech recognition engine Wit.ai (https://wit.ai/).

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

with badges and prestigious titles (Level 1 Idiom Newbie → Level 2 Idiom Rookie → Level 3 Idiom Trainee → Level 4 Idiom Specialist → Level 5 Master of Idioms). The nicknames of top scorers also appear in the IdiomsTube's Halls of Honour (the *IdiomsTube World Leaderboard* and the *IdiomsTube School Leaderboard)*.

To connect the communities of IdiomsTube users, Facebook and Instagram pages have been set up. IdiomsTube users may also interact with each other, as well as the research team, through public posts or private messages. At the time of writing, the two pages had 1,884 followers/subscribers. In order to maintain activity on the pages, a picture or video post introducing an English FE is released every day. The same posts also pop up at every log in to the IdiomsTube website.

To enable school teachers and learners to incorporate IdiomsTube into their regular weekly or daily English language homework, the IdiomsTube Teacher's interface was created. This interface allows teachers to conduct multiple functions, including the automatic compilation of reports on the learning progress of every student and class, analysis of the vocabulary difficulty, speech rate and FE distribution of any English-captioned YouTube video, the growth of a personal repository of YouTube videos for use as teaching materials, and the ability to promote recommended, captioned YouTube videos to selected classes. (For further elaboration and demonstration of the Teacher's interface's functions, please see Figures 14–19 and https://tinyurl.com/yy6gzflb.)

4. The research underpinning the development of the IdiomsTube app

The design of the Idioms Tube app and its functions has been informed and supported by various theories and findings concerning effective FE learning. This section provides an account of the links between the methodological decisions made in the app development process and research into vocabulary acquisition.

4.1 The FE list for automatic extraction of English FEs

To automatically generate the learning tasks, the system first detects and extracts English FEs present in each YouTube video. This is done using a pattern-matching approach, which operates by searching for matching patterns in the caption based on a pre-defined dictionary of FE templates (see also Rayson, 2003). The advantage of this approach is that FEs can be identified even when they appear in variant forms. For example, variant forms of the FE template *put * to good use* (with * representing 0–3 words) can be extracted by the system, such as in:

the space has been put to good use
put it to good use
put the money to good use

At the time of writing, the IdiomsTube FE dictionary contained 53,635 templates, which have all been lemmatised (using Python's NLTK package). These entries range from two to 23 words in length, with a mean length of 3.94 words. They were collected by the author over the years from print and online sources, including dictionaries of idioms, proverbs, sayings, slang, and famous quotes. The idiomatic status of these items in the IdiomsTube dictionary can be verified by the fact that they also appear in reputable dictionaries.[3]

[3]For generating the pre-learning tasks, each FE needs to be hyperlinked to online dictionaries. These dictionaries include https://idioms.thefreedictionary.com (71.44%), https://www.cambridge.org (7.33%), https://www.collinsdictionary.com (7.15%), https://www.merriam-webster.com (5.68%), https://www.macmillandictionary.com (4.84%), en.wiktionary.org (1.35%), https://www.phrases.org.uk (1.06%), en.oxforddictionaries.com, https://www.brainyquote.com, https://www.thefreedictionary.com, https://www.yourdictionary.com, https://www.vocabulary.com, and https://www.urbandictionary.com (< 1%).

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

4.2 The design of the learning activities

Learning a foreign language through listening alone is difficult. Due to the limited capacity of short-term memory and time pressure, the human brain is hardwired to prioritise the decoding of meaning from auditory input rather than remembering words verbatim. To facilitate verbatim memory of the FEs in the YouTube videos, IdiomsTube learning tasks (see section 3.1) repeatedly engage learners with the form, meaning, and usage of the FEs before, during, and after viewing.

4.2.1 The design of the pre-learning tasks

The pre-learning tasks, which include the glossary and the dictionary lookup, aim to (1) give learners a clearer idea about the particular vocabulary focus of the video; (2) facilitate the learners' noticing of the FEs (see Schmidt, 1990, for a discussion of the noticing hypothesis, which suggests that noticing is an essential starting point for acquisition); and (3) gloss the vocabulary items. There is substantial empirical evidence for the facilitative effect of glossing, glossary and dictionary lookup on incidental vocabulary learning (Hulstijn, Hollander & Greidanus, 1996; Webb, 2010). When attention to and engagement with noteworthy vocabulary is enhanced through providing a glossary and highlighting in bold, learners are more likely to remember the vocabulary than in cases where no glossing is provided. Input enhancement may be provided in various formats, although the most common format is to highlight the vocabulary in bold in the original written text and to provide a glossary with simple definitions in the first or second language in the margins (Ko, 2012; Yanguas, 2009). As IdiomsTube is open to English learners from all first language backgrounds, glossing is provided in this most common format in English as a second language.

4.2.2 The design of the consolidation tasks

The consolidation tasks include three types of exercises: fill-in-the-blanks, speaking, and spelling.[4] This is in addition to the selected scene replay, which can be repeated on demand by clicking the *Hint* button. They were designed to engage learners with different aspects of FE knowledge. The speaking and spelling exercises facilitate learners' *recall* of the phonological and orthographic forms respectively through *production*. The fill-in-the-blanks task, on the other hand, consolidates learners' *recognition* of the meaning of the FEs as they are used in context.

This task design of these consolidation tasks was guided by the *involvement load hypothesis* (Laufer & Hulstijn, 2001), which suggests that involvement in vocabulary learning is made up of three parts: *need*, *search*, and *evaluation*. To earn experience points for the fill-in-the-blank exercise, learners must remember the content of the video and correctly recall the FE that completes each original utterance. This represents the *need* component in Laufer and Hulstijn's (2001) framework. The process of choosing the right answer among the multiple choices and the re-watching of selected scenes if they are uncertain illustrates *search*. After filling in an FE, learners need to reread the completed utterance to determine whether the FE fits and if it is the best choice. This illustrates *evaluation*. When learners successfully complete all six questions in each exercise, they get immediate feedback in terms of scores and the correct answers so that improvements can be made.

The design of these consolidation tasks was also informed by previous research that examined the relationship between FE exercise format and learning outcomes. For example, the decision to schedule the three exercises after the video playback was informed by the finding that greater gains in FE knowledge can be generated by scheduling the exercises after rather than before the exposure (Strong & Boers, 2019). Furthermore, the decision to provide a list of intact FEs (e.g. *make a wish*) as the multiple choices in the fill-in-the-blank exercise was informed by Boers, Dang and Strong's

[4]See Szudarski and Conklin (2014) and Durrant and Schmitt (2010) for the importance of consolidation in FE learning, and Boers (2020) for a review of FE consolidation exercise designs.

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

(2017) finding that this practice is less likely to confuse learners than if learners are required to choose just the appropriate verb, for instance, to complete an FE (e.g. ____ *a wish*) embedded within a sentence.

Of the three tasks, the fill-in-the-blanks appears to be the most popular among current IdiomsTube users, who consider the multiple-choice fill-in-the-blanks format suitably challenging (Lin, under review). Although the speaking task and the spelling task are slightly less popular, they are important in terms of providing a variety of tasks through which learners may engage with different aspects of FE knowledge, including their phonological form, orthographic form, and composition. According to research in psycholinguistics on the role of the phonological loop in vocabulary learning (Baddeley, Gathercole & Papagno, 1998; see also Lin, 2021), engaging learners with the phonological form is particularly important for facilitating their memory of new vocabulary items. Some empirical studies found that learners who read new vocabulary aloud or silently in their heads seemed to remember the items more effectively than their counterparts who encountered the same vocabulary but were forbidden from reading the items aloud or silently (Baddeley *et al.*, 1998).

4.2.3 The design of the revision tasks

The revision tasks include the personal bookmarks, exercise history, video history, and revision flashcards. As second language vocabulary studies have shown, repeated and sustained exposure to new vocabulary is important because the process of vocabulary acquisition is incremental (see Schmitt, 2010, for further discussion). Learners need to encounter a new item eight to 20 times to remember and learn the item (Pigada & Schmitt, 2006; Rott, 1999; Waring & Takaki, 2003). On IdiomsTube, these repeated encounters accumulate through participation in the pre-learning and consolidation tasks, in addition to the viewing of personal learning and video diaries as part of the learning progress management routine. When learners view their progress and exercise history, they can see which items and questions they have answered (in)correctly and replay selected scenes in which the items are used. This allows them to recall their memory of the form and usage.

The revision flashcards are yet another strategy to assist learners in systematising and increasing their encounters with the new vocabulary. They are auto-populated from the list of FEs that learners have encountered through their recently viewed YouTube videos. According to previous studies on vocabulary acquisition (e.g. Bahrick, Bahrick, Bahrick & Bahrick, 1993), repeated exposure to new vocabulary at regular time intervals produces significantly higher knowledge retention than massed exposure. This spacing effect has been described as "one of the most robust phenomena in experimental psychology" (Ellis, 1995: 118) and "is too powerful for instructional programmes to ignore" (Ellis, 1995: 120). Therefore, the IdiomsTube system monitors the dates on which each learner encounters each FE. Informed by Bahrick *et al.*'s (1993) experimental findings, the system has been set to provide 13 repetitions at 14-day intervals. In other words, any FEs that learners have not encountered in the videos or through any of the tool's learning activities in the past 14 days will appear as a flashcard.

4.3 Assessing video difficulty

The quality of language input has long been an area of concern in education. To facilitate the conversion of input to intake, language learners need input at the right level of difficulty so that it is comprehensible enough (Zhao, 1997). This type of input, which, according to input hypothesis (recently renamed comprehension hypothesis; Krashen, 2009) is one level above that of the language learner, is called *comprehensible input*.

Finding comprehensible input among hundreds of thousands of videos on YouTube, however, is not easy. Until now, there has been no available system able to automatically assess the difficulty

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

of online videos from a language input point of view. The approach currently adopted by the IdiomsTube app to automatically estimate the difficulty of a video from a language point of view considers two well-researched factors of second language listening comprehension, namely lexical difficulty and speech rate (see Lin, in preparation).

4.3.1 Lexical difficulty

Over the past decades, substantial research has been conducted on the impact of lexical difficulty on text comprehension (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt, Jiang & Grabe, 2011). Results have shown that, to gain adequate comprehension, learners need to know at least 95% to 98% of the words in a *written* text (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt *et al*., 2011). In the second language vocabulary acquisition literature, this is known as *lexical coverage*. In the context of *spoken* communication, adequate comprehension may be gained at 90% lexical coverage (Adolphs & Schmitt, 2003). It should be noted, however, that lexical coverage also varies by genre (see Nation, 2013).

To estimate the lexical difficulty of an English-captioned YouTube video, the IdiomsTube app analyses the distribution of words in the video based on Nation's BNC/COCA list (Nation, 2012). Within the list, word families are categorised based on their frequency in British and American English combined. The 1,000 most frequent word families in British and American English (e.g. *air*, *big*, *chair*) are put in the 1k bin, the second 1,000 most frequent word families (e.g. *edit*, *insist*, *typical*) are put in the 2k bin, the third 1,000 most frequent word families (e.g. *potential*, *republic*, *portrait*) are put in the 3k bin, and so on.

Following the initial lexical distribution analysis, the app recommends a lexical difficulty rating based on the bin of word families that need to be known in order to achieve 85% lexical coverage. A video is given a lexical difficulty rating of 1 if knowledge of word families in the 1k bin is sufficient to cover 85% of the word tokens in the video. A video is given a lexical difficulty rating of 2 if knowledge of word families in both the 1k and 2k bins is needed to cover 85% of the tokens in the video. A video is given the highest lexical difficulty rating of 5 within the IdiomsTube system if knowledge of word families in the 1k, 2k, 3k, 4k, and 5k bins (or even higher) are all needed to cover 85% of the tokens. A lexical coverage level of 85% was chosen as the system threshold because of the assumption that the comprehension of videos may be achieved at a lower percentage of lexical coverage with the presence of contextual, visual, and auditory cues. This assumption awaits empirical testing.

4.3.2 Speech rate

Speech can be defined as a continuous stream of sounds that, given the limited capacity of short-term memory, needs to be processed under time pressure. The difficulty increases when processing spoken input in a second language. This is likely to be the reason why, when asked about the challenges of English listening, most learners complained about the speech rate of the input (Zhao, 1997).

In their study on radio news broadcasts, Pimsleur, Hancock and Furey's (1977) proposed definitions of fast and slow English speech:

Fast	= above 220 words per minute (wpm)
Moderately fast	= 190 to 220 wpm
Average	= 160 to 190 wpm
Moderately slow	= 130 to 160 wpm
Slow	= below 130 wpm

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

These reference values are widely known in the ELT circle. However, their use as guidelines in ELT material development has been criticised by Tauroza and Allison (1990) on the basis that the values were derived based on just one particular variety of (scripted) speech. To fill this gap for empirically valid reference values for IdiomsTube users, a controlled experiment involving 87 EFL learners was recently conducted (Lin, in preparation). In the experiment, subjects listened to 80 naturally occurring utterances varying in speech rates and density of unknown words and were asked to repeat and translate what they heard under time pressure. The resulting best-fitting linear-mixed effects model, which takes account of a video's density of unknown words and second language proficiency when recommending a speech rate rating for a video, will be incorporated into IdiomsTube in the next major app update.

In this interim period, Pimsleur *et al.*'s (1977) reference values are used to assess and rate the difficulty of the speech rate in YouTube videos. For example, a video with a speech rate below 130 wpm is given a speech rate difficulty rating of 1; a video with a speech rate between 130 and 160 wpm is given a rating of 2. The maximum difficulty rating of 5 is given at 220 words or more per minute. The speech rate of a video is calculated by taking the median of the speech rates of all individual lines in the caption of the video. To calculate the speech rate of each individual line in the caption, the duration of the line (as indicated by the time codes) is divided by the number of tokens in the line.

If the lexical difficulty or speech rate difficulty of a video is awarded at a rating of 4 or above, the system will show a pop-up window reminding learners they can adjust the speech rate of the video playback. This aims to give learners more time to process the auditory input and engage more with the language content.

5. Challenges facing the development of apps for facilitating FE learning from online videos

IdiomsTube was developed to facilitate independent FE learning from any user-chosen English-captioned YouTube video. Before the creation of IdiomsTube, many commercial websites existed, providing English learning activities based around a repository of preselected videos for a fee. A notable example is the website *English Central* (https://www.englishcentral.com). However, IdiomsTube differs because, as a research project, it focuses on developing the *technology* to automatically generate various types of English learning activities for any English-captioned video on YouTube rather than to simply provide a platform for disseminating activities developed by teachers. Once developed, this automatic technology will also be applicable to other online streaming platforms.

As the first-ever project to develop this type of technology for computer-assisted language learning, several technical challenges have arisen. This section elaborates on these challenges so that creative solutions may be found through constructive academic discussion.

5.1 The length of videos affects the number of FEs found

The identification of sufficient, noteworthy FEs from each video is crucial to the normal functioning of the IdiomsTube app. The number of FEs identified in a video, however, depends on multiple factors including the comprehensiveness of the IdiomsTube FE dictionary, the ability of the app's algorithm to recognise the relevant FEs and their variant forms, and the duration of the video. While continual improvements are being sought in terms of the content of the FE dictionary and the app's ability to recognise FEs, we hope to highlight the underexplored effect of video duration on the number of FEs being identified.

Learners may expect IdiomsTube to identify the majority of the FEs present in a video. However, the reality is that the number of FEs identified is dependent upon the duration of the video chosen. In our experience, there are far more short (under 20 minutes) than long (over

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

20 minutes) videos on YouTube. This is likely to be for two possible reasons. First, most viewers expect to find short, entertaining, and/or instructional videos on YouTube. How-to videos, for example, are typically five minutes long. TED Talks are typically 15 to 20 minutes long. If viewers wish to view a longer video, such as a whole movie, they are more likely to find these on alternative video-streaming websites such as Netflix and Amazon. The second reason is that most YouTube content providers are ordinary home producers (as opposed to production companies), who produce videos with limited human and financial resources. There are pros and cons for using home-produced videos as English learning resources, but in the long term and to increase the chance of finding longer videos with many FEs present, it may be worth extending the technology developed in this project to other online streaming platforms, which provide longer videos.

5.2 The priority at which each FE should be taught

In cases where many (say 20) FEs are present in a video, further questions arise concerning the value or necessity of including all the FEs present in the automatically generated pre-learning, consolidation, and revision tasks. This question requires in-depth discussion from both a theoretical and a technical point of view. Considering previous research on the limited human attention span (Young, Robinson & Alberts, 2009), it seems likely that teaching all 20 FEs may not be any more beneficial than teaching a careful selection of five FEs. This is because there is only so much information that a learner can absorb at once. Too much input may only overwhelm or confuse the learner. In psychology and cognitive science, the magic number 7, plus or minus 2, has been proposed to describe the number of items of information that a human brain can focus on at any one time (Miller, 1956). If it is possible to apply this notion to the context of computer-assisted English FE learning, then IdiomsTube should aim to teach a selection of five to nine FEs, even if the actual number present in the video is higher. The selection process for these five to nine items, however, is a separate theoretical and technical issue.

Determining the priority at which FEs should be taught presents a substantial theoretical and technical challenge because, so far, no empirical investigations have been conducted to reveal the ways in which a range of factors (e.g. length of FE, frequency of occurrence, phonological parallelism, semantic transparency) interact to affect the learnability of English FEs. Furthermore, no published research exists to provide information on the frequencies of occurrence of a long list of FEs in British and American English. The longest existing list of English FEs, developed by Lei and Liu (2018), contains only 9,049 items. Therefore, there is very little available information on dealing with the practical problem of determining the relative priority at which the 53,635 FEs in the IdiomsTube dictionary should be presented in the learning tasks.

In this project, a crude approach was attempted to determine the teaching priority of FEs based on three factors, namely the length of the FE, the difficulty of the lexical words constituting the FE (based on the BNC/COCA list by Nation, 2012), and the frequency of occurrence within our trial corpus of 40,000 randomly sampled YouTube videos. The algorithm prioritises the teaching of FEs that are long, consist of lexical words from the higher level frequency bins in Nation's list, and have a low frequency of occurrence in our corpus of 40,000 randomly sampled YouTube videos. This algorithm allows IdiomsTube to focus on teaching FEs that are more challenging in terms of their composition (i.e. they are longer and made up of more difficult words) and, therefore, may be easier to miss due to their infrequent appearance on YouTube. Nevertheless, this algorithm clearly has room for improvement in terms of sophistication. For example, as it only takes account of three factors, many FEs in the list of 53,635 FEs share the same priority ranking. Looking ahead, we plan to conduct experimental research and big data analysis to create a more empirically robust system. As the number of IdiomsTube users continues to grow, we are beginning to see which FEs are more difficult to learn by analysing existing users' in-app behaviour. For example, it is becoming clearer which FEs are most often encountered by IdiomsTube users given their collective YouTube video preferences, which FEs learners are most likely to make errors within

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

the consolidation tasks, and which FEs learners are more likely to request hints for. When the dataset gets sufficiently large, we will apply machine learning algorithms to develop a more robust statistical model.

5.3 The distinction between metaphoric and non-metaphoric use

Another technical challenge facing IdiomsTube, as well as other software programs that identify FEs based on their form, is the inability to discern between metaphoric and literal use of FEs. Some FEs such as *put somebody's foot down* and *cooking with gas* can be used both metaphorically and literally depending on context. These FEs will be identified and included in the learning tasks even though they may have been used literally within the video. Solutions to this technical challenge are needed to improve IdiomsTube's accuracy in terms of FE identification. One of the possible, albeit computationally sophisticated, solutions is to perform semantic tagging on every word in the video and, at the same time, enrich IdiomsTube's FE dictionary so that it stores not only the FEs but also each item's context(s) of use. If there is a match in terms of contexts of use (e.g. *put your foot down* is used metaphorically in the context of decision-making), then the usage will be included in the automatically generated tasks.

5.4 Caption formats

YouTube is a great resource for independent language learning because it is a platform where home and commercial producers come together to produce an unprecedented variety of English videos. Due to this, every learner can find videos that they enjoy watching and through which they can learn English. The disadvantage of having so many home and commercial producers on YouTube, however, is the difficulty of standardising and/or controlling the quality of the captions, from which all the IdiomsTube learning tasks are generated. For example, there is no standard on spelling, punctuation, the use of abbreviations and contracted forms, the possible positions of line breaks, the length of each line of caption, when time codes should be inserted, and so on. Although some producers will not transcribe some audible words (such as song lyrics, laughs, and murmurs), other producers even insert notes about non-verbal behaviour (such as claps and exhalation, presumably for the hard of hearing). This lack of uniformity in captioning practice has greatly complicated the steps for automatically generating the learning tasks.

There are also other issues surrounding YouTube captioning, such as transcription accuracy. YouTube has a built-in auto-caption service that uses automatic speech recognition to transcribe the video in real time. The option to switch on auto-caption appears to be available for all English videos in which no user-uploaded caption is available. The technology also seems reasonably accurate for videos that include slow, careful, and formal speech (as in the case of some National Geographic documentary videos we tested). However, mis-transcriptions are noticeable with other types of English videos. To address this issue, the system has currently disabled the generation of learning tasks for videos that provide auto-caption only. When learners search for videos on the IdiomsTube front page, only videos with user-uploaded English captions will be displayed in the results section. This solution provides a temporary fix only and it is not without problems. For instance, there is still no way to spot inaccuracies in user-uploaded captions, meaning that these inaccuracies may then appear in the learning tasks. Going forward, improved solutions are needed, including a more accurate built-in auto-caption service from YouTube and more innovative ways of spotting inaccuracies in user-uploaded captions such as running spell-checkers and grammar checkers on video transcripts before generating the learning tasks.

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

6. Conclusion

This paper presented IdiomsTube, a computer tool developed for independent English FE learning through YouTube videos. It discusses the ideas behind the design of its functions as well as the theoretical and empirical bases of the project. At this stage, the app is capable of automatically generating three types of learning tasks to facilitate the learning of English FEs present in any English-captioned videos posted on YouTube, automatically assessing the difficulty level of the videos, and sharing and highlighting video recommendations from other learners and EFL class teachers. The app also provides gamification features and a teacher's interface so that it can be used to support school-based as well as independent English learning.

Further improvements in the design and functions of IdiomsTube are needed, particularly in terms of deciding the priority level at which each FE is taught, distinguishing between the metaphorical and literal forms of FEs, and overcoming the problems arising from the lack of uniformity in user-uploaded English captions. These practical decisions should be informed by empirical research evidence, and these gaps are being filled through experiments in the lab as well as analyses of log files showing users' in-app behaviours and patterns (see, e.g., Lin, 2021, under review, in preparation). Filling the empirical research gaps and tackling these technical challenges are crucial for IdiomsTube as well as all future automatic tools designed to facilitate independent English FE learning from any other video-streaming platforms.

In this age of the internet, video streaming is a part of many young people's daily routines. Looking ahead, there is a need for more computer-assisted tools to be developed so that this extensive exposure to authentic English input may be fully exploited as a sustainable foreign language learning opportunity.

Acknowledgements. This project was co-funded by the Hong Kong Research Grants Council (Ref: PolyU 156020/17H) and the Standing Committee on Language Education and Research of the Hong Kong SAR Government (Ref: EDB(LE)/P&R/EL/164/1).

Ethical statement. The author is unaware of any conflict of interest in this research.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/S0958344021000252

References

Adolphs, S. & Schmitt, N. (2003) Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4): 425–438. https://doi.org/10.1093/applin/24.4.425

Baddeley, A., Gathercole, S. & Papagno, C. (1998) The phonological loop as a language learning device. *Psychological Review*, 105(1): 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S. & Bahrick, P. E. (1993) Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4(5): 316–321. https://doi.org/10.1111/j.1467-9280.1993.tb00571.x

Bishop, H. (2004) The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In Schmitt, N. (ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 227–248. https://doi.org/10.1075/lllt.9.12bis

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

Boers, F., Dang, T. C. T. & Strong, B. (2017) Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, 21(3): 362–380. https://doi.org/10.1177/1362168816651464

Chandler, D. & Munday, R. (2020) *A dictionary of media and communication* (3rd ed.). Oxford: Oxford University Press. https://doi.org/10.1093/acref/9780198841838.001.0001

Durrant, P. & Schmitt, N. (2010) Adult learners' retention of collocations from exposure. *Second Language Research*, 26(2): 163–188. https://doi.org/10.1177/0267658309349431

Ellis, N. C. (1995) The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, 8(2–3): 103–128. https://doi.org/10.1080/0958822940080202

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

Hu, M. & Nation, P. (2000) Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1): 403–430.

Hulstijn, J. H., Hollander, M. & Greidanus, T. (1996) Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *The Modern Language Journal*, 80(3): 327–339. https://doi.org/10.1111/j.1540-4781.1996.tb01614.x

Iyengar, S. S. & Lepper, M. R. (2000) When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6): 995–1006. https://doi.org/10.1037/0022-3514.79.6.995

Ko, M. H. (2012) Glossing and second language vocabulary learning. *TESOL Quarterly*, 46(1): 56–79. https://doi.org/10.1002/tesq.3

Krashen, S. (2009) The comprehension hypothesis extended. In Piske, T. & Young-Scholten, M. (eds.), *Input matters in SLA*. Bristol: Multilingual Matters, 81–94. https://doi.org/10.21832/9781847691118-007

Laufer, B. & Hulstijn, J. (2001) Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, 22(1): 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B. & Ravenhorst-Kalovski, G. C. (2010) Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1): 15–30.

Lei, L. & Liu, D. (2018) The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, 23(2): 216–243. https://doi.org/10.1075/ijcl.16135.lei

Lin, P. (2018a) Formulaic language and speech prosody. In Siyanova-Chanturia, A. & Pellicer-Sánchez, A. (eds.), *Understanding formulaic language: A second language acquisition perspective*. Abingdon: Routledge, 78–94. https://doi.org/10.4324/9781315206615-5

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

Lin, P. (2019) *Vocabulary learning in the age of internet television*. Paper presented at the 54th RELC International Conference & 5th Asia-Pacific LSP & Professional Communication Association Conference, Southeast Asian Ministers of Education Organisation (SEAMEO), Singapore, 11–13 March.

Lin, P. (2021) In search of the optimal mode of input for the acquisition of formulaic expressions. *TESOL Quarterly*. Advance online publication. https://doi.org/10.1002/tesq.3055

Lin, P. (in preparation) Textual factors and individual differences in L2 listening comprehension: A controlled experiment.

Lin, P. (under review) Computer-assisted learning of English formulaic expressions from YouTube videos. In Reynolds, B. (ed.), *Vocabulary learning in the wild*. Springer, 1–24.

Lin, P. M. S. (2012) Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics*, 33(3): 342–347. https://doi.org/10.1093/applin/ams017

Lin, P. M. S. (2013) The prosody of formulaic expression in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics*, 18(4): 561–588. https://doi.org/10.1075/ijcl.18.4.05lin

Lin, P. M. S. (2014) Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System*, 42: 164–176. https://doi.org/10.1016/j.system.2013.11.010

Lin, P. M. S. & Siyanova-Chanturia, A. (2014) Internet television for L2 vocabulary learning. In Nunan, D. & Richards, J. C. (eds.), *Language learning beyond the classroom*. Abingdon: Routledge, 149–158.

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

Miller, G. A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63(2): 81–97. https://doi.org/10.1037/h0043158

Nation, I. S. P. (2012) *The BNC/COCA word family lists*. https://www.wgtn.ac.nz/__data/assets/pdf_file/0005/1857641/about-bnc-coca-vocabulary-list.pdf

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

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

Peters, E. (2012) Learning German formulaic sequences: The effect of two attention-drawing techniques. *Language Learning Journal*, 40(1): 65–79. https://doi.org/10.1080/09571736.2012.658224

Pigada, M. & Schmitt, N. (2006) Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1): 1–28.

Pimsleur, P., Hancock, C. & Furey, P. (1977) Speech rate and listening comprehension. In Burt, M., Dulay, H. & Finocchiaro, M. (eds.), *Viewpoints on English as a second language in honour of James E. Alatis*. New York: Regents, 27–34.

Rayson, P. (2003) *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison*. Lancaster University, unpublished PhD.

Rott, S. (1999) The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, 21(4): 589–619. https://doi.org/10.1017/S0272263199004039

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press

Schmidt, R. W. (1990) The role of consciousness in second language learning. *Applied Linguistics*, 11(2): 129–158. https://doi.org/10.1093/applin/11.2.129
Schmitt, N. (2010) *Researching vocabulary: A vocabulary research manual*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230293977
Schmitt, N., Jiang, X. & Grabe, W. (2011) The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1): 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x
Spöttl, C. & McCarthy, M. (2004) Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In Schmitt, N. (ed.), *Formulaic sequences: Acquisition, processing and use*. Amsterdam: John Benjamins, 191–226. https://doi.org/10.1075/lllt.9.11spo
Strong, B. & Boers, F. (2019) The error in trial and error: Exercises on phrasal verbs. *TESOL Quarterly*, 53(2): 289–319. https://doi.org/10.1002/tesq.478
Szudarski, P. & Conklin, K. (2014) Short- and long-term effects of rote rehearsal on ESL learners' processing of L2 collocations. *TESOL Quarterly*, 48(4): 833–842. https://doi.org/10.1002/tesq.201
Tauroza, S. & Allison, D. (1990) Speech rates in British English. *Applied Linguistics*, 11(1): 90–105. https://doi.org/10.1093/applin/11.1.90
Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M. & Tice, D. M. (2008) Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology*, 94(5): 883–898. https://doi.org/10.1037/0022-3514.94.5.883
Volz, A., Higdon, J. & Lidwell, W. (2019) *The elements of education for teachers: 50 research-based principles every educator should know*. New York: Routledge. https://doi.org/10.4324/9781315101002
Waring, R. & Takaki, M. (2003) At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, 15(2): 130–163.
Webb, S. (2010) Pre-learning low-frequency vocabulary in second language television programmes. *Language Teaching Research*, 14(4): 501–515. https://doi.org/10.1177/1362168810375371
Webb, S., Newton, J. & Chang, A. (2013) Incidental learning of collocation. *Language Learning*, 63(1): 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x
Wible, D. (2008) Multiword expressions and the digital turn. In Meunier, F. & Granger, S. (eds.), *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins, 163–181. https://doi.org/10.1075/z.138.13wib
Wojcicki, S. (2015) Keynote speech at VidCon 2015 conference, Anaheim Convention Center, California, July 23–25.
Yanguas, I. (2009) Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, 13(2): 48–67.
Young, M. S., Robinson, S. & Alberts, P. (2009) Students pay attention! Combating the vigilance decrement to improve learning during lectures. *Active Learning in Higher Education*, 10(1): 41–55. https://doi.org/10.1177/1469787408100194
Zhao, Y. (1997) The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics*, 18(1): 49–68. https://doi.org/10.1093/applin/18.1.49

About the author

Phoebe Lin is an assistant professor at the Department of English and a member of Research Centre for Professional Communication in English (RCPCE), The Hong Kong Polytechnic University. Her research interests include vocabulary acquisition, speech prosody, experimental research, second language acquisition, multimodal corpora, natural language processing, data science, crowdsourcing, software development, and computer-assisted language learning.

Author ORCiD. Phoebe Lin, https://orcid.org/0000-0001-7175-2099

https://doi.org/10.1017/S0958344021000252 Published online by Cambridge University Press