

# Vocabulary Learning and Assessment: A Commentary on Four Studies for JALT Vocabulary SIG

Irina Elgort

*Victoria University of Wellington*

## Abstract

Four papers by Chie Ogawa, Haidee Thomson, Michael Holsworthm and Darrell Wilkinson were presented in the *Vocabulary Learning and Assessment* session at the Eighth Annual Japan Association for Language Teaching (JALT) Vocabulary Special Interest Group (SIG) Symposium, at the University of Niigata, Japan, in May 2020. The papers raised methodological questions, proposed approaches to assessing spoken and written word knowledge and fluency, and presented some experimental findings. It is my pleasure to discuss these papers in terms of the ideas proposed by the four researchers, their implementation and potential future directions.

## 1 Introduction: The Golden Age of L2 Vocabulary Research

Forty years ago, Paul Meara (1980) referred to vocabulary as a neglected area of language learning research. The situation has changed dramatically since then and L2 vocabulary research is now thriving. A new lexical studies conference, *Vocab@*, first held at the Victoria University of Wellington, New Zealand, in 2013, with two subsequent conferences, *Vocab@Tokyo* in 2016 in Japan and the most recent gathering, *Vocab@Leuven* in 2019 in Belgium, shows an impressive array of research topics, high quality of papers from prominent vocabulary researchers and, most encouragingly, many new, talented vocabulary researchers entering the field. Key themes represented at *Vocab@Leuven* included corpus studies, deliberate and incidental learning of words and multiword expressions, and tests of vocabulary knowledge. However, there are areas of vocabulary research that do not appear to enjoy the same level of attention as the research topics above; in particular, I want to draw our attention to two such topics: *fluency* development that, according to Nation (2007), is a key component of a balanced language learning curriculum, and the development of *non-declarative lexical knowledge*, including implicit, procedural and automatised knowledge (Godfroid, 2020). Recently, I have identified this gap in a methodological review of L2 studies in technology-mediated vocabulary development (Elgort, 2018) but, I believe, the situation is similar in L2 vocabulary research, more generally.

It is really exciting, therefore, to see that fluency development is considered in the majority of the papers presented in the *Vocabulary Learning and Assessment* session at this year's JALT Vocabulary SIG Symposium. In particular, Ogawa and Thomson consider questions related to speaking fluency and Holsworth considers

fluency of lexical processing in reading. Another encouraging trend in this year's papers is attention to multi-word expressions. A relative newcomer to L2 lexical studies, research into processing, using, learning and teaching of multi-word expressions has been rapidly gaining momentum in the last 10–15 years, no doubt, at least partially due to the vastly improved access to large-size language and learner corpora and a simultaneous associated boost in the number of accessible and user-friendly corpus and text analysis tools (such as Lawrence Anthony's *AntLab* [www.laurenceanthony.net](http://www.laurenceanthony.net); Mark Davies' *English-Corpora.org*; and #*LancsBox* – the Lancaster University corpus toolbox [corpora.lancs.ac.uk/lancsbox](http://corpora.lancs.ac.uk/lancsbox) spear-headed by Vaclav Brezina, and Tom Cobb's *Lextutor* tools [www.lextutor.ca](http://www.lextutor.ca), among others).

In the symposium, Ogawa's paper looks at approaches to teaching formulaic language, while Thomson brings to our attention challenges associated with assessing the knowledge of multi-word expressions in the classroom setting. An innovative take on vocabulary assessment is also considered in the paper by Holsworth, who proposes a set of measures for evaluating three components of lexical quality outlined by Perfetti and colleagues (e.g., Perfetti, 2007; Perfetti & Hart, 2001): recognition of orthographic form, access to semantic knowledge and accuracy of the spoken form production. Finally, Darrell Wilkinson considers the use of word cards – one of tried-and-true methods of deliberate word learning that is experiencing a resurgence as a result of the recent advent of online and mobile flashcard applications. Let us now consider each of the four papers. I will start with the two studies by Ogawa and Thomson that look at instructional approaches to facilitating productive use of multi-word expressions as a bridge to attaining production fluency. I will then turn to Holsworth's paper that considers fluency of visual word processing. After that, I will discuss Wilkinson's experiments that investigate the effectiveness of self-made word cards.

## **2 The Four Studies**

### ***2.1 Teaching Ideas of Improving Oral Performance through Formulaic Language Instruction by Chie Ogawa, Kyoto Sangyo University***

Chie Ogawa's paper suggests ideas on how to facilitate the learning of formulaic language by English as a Foreign Language (EFL) learners, as a way to improve their spoken fluency. As mentioned in the "Introduction", research into *formulaic language* – a cover term for different types of multi-word expressions (Schmitt, 2010) – is gaining momentum among L2 vocabulary researchers. However, so far, the bulk of this research has been focused on whether formulaic sequences are processed faster than individual words and whether non-native speakers are able to attain native-like knowledge of such sequences. There is less published research that looks at pedagogical interventions that facilitate the learning of multi-word expressions (but see the work of Frank Boers and colleagues on the topic). Therefore, Ogawa's paper is timely.

We know that the acquisition of formulaic language is challenging and even more proficient L2 speakers may have trouble producing multi-word expressions

(e.g., Laufer & Waldman, 2011). One of the reasons for this learning difficulty is that formulaic sequences are not salient in the input; in reading, for example, they are not visually different from individual words and thus are less likely to be explicitly noticed as multiword units, unless they have a completely different meaning from that of their component words, as may be the case for idiomatic expressions, such as *kick the bucket* or *another cattle of fish*.

Ogawa proposes three form-focused pedagogical treatments that could promote proceduralisation of the knowledge of formulaic language: text enhancement, peer feedback and repetition in language learning activities and tasks. The first approach, text enhancement (Sharwood Smith, 1991, 1993; Szudarski & Carter, 2016), aims to draw learners' attention to multi-word expressions, making them more salient in the input through externally induced focus. Ogawa refers to a meta-analytic synthesis by Lee and Huang (2008), suggesting that the effect of typographic enhancement on learning grammar from input is likely to be positive but small. In addition, only very few studies have examined how typographic enhancement affects proceduralisation and the development of implicit knowledge. In the first (to my knowledge) study of its kind, Sonbul and Schmitt (2013) did not see any effect of typographic enhancement on the development of implicit knowledge of lexical (medical) collocation, while learners' explicit knowledge was enhanced by this treatment. In a conceptual replication and an extension of this study, Toomer and Elgort (2019) confirmed the positive effect of orthographic enhancement on the development of explicit knowledge of L2 collocations and no effect on implicit knowledge. Nevertheless, Toomer and Elgort found that enriched input (nine contextual occurrences over 2 days) may facilitate the development of implicit knowledge of lexical collocations for advanced English as a Second Language (ESL) learners (measured using a collocation priming task) when the target collocations are *not* typographically enhanced in the text. Furthermore, Ogawa rightly points out that typographic enhancement may change learners' engagement with the rest of the text (Choi, 2017) and that learning from input alone may be challenging for lower proficiency learners.

Ogawa, thus, proposes the second treatment, peer feedback. The advantage of corrective peer feedback compared with teacher feedback is that learners are both feedback-givers and feedback-receivers, which doubles their opportunities for noticing the gap in their knowledge. With this in mind, Ogawa argues that it could be an effective way of increasing the noticing of multi-word expressions in the output, while warning that the provision of useful corrective peer-feedback is not assured by simply requesting it. Participants need to be willing to point out errors and inaccuracies in their peers' output, which could be a face-threatening act. For this reason, it is critical for the learners to be trained in the provision of peer-feedback in a constructive and polite manner, and for them to understand the benefits of giving and receiving peer-feedback. What's missing from this section, in my opinion, is concrete suggestions on how such peer feedback may be elicited, that is, what kind of activities and tasks could be designed to ensure that peer feedback is focused on multiword expressions and is useful in promoting their acquisition. Using peer-feedback to learn formulaic language may also be somewhat problematic: even if advanced L2 learners are having trouble with acquiring

formulaic language, would low and intermediate proficiency learners be able to provide useful corrective feedback to their peers? Another question is: what kind of teacher support would be most beneficial in helping students understand how to engage in peer-feedback activities in an effective way?

The third approach to improving the knowledge of formulaic language proposed by Ogawa and, in my opinion, the most promising one, is developing and using learning activities and tasks (such as 4/3/2, Nation, 1989) that create conditions for the proceduralisation of the established declarative knowledge and its eventual automatization, as argued by Segalowitz (2010) and DeKeyser (2003, 2015). However, as shown by Solovyeva and DeKeyser (2018) and Hui (n.d.), it may take a very long time to develop automaticity of vocabulary processing through input only. Moving beyond Nation's 4/3/2, it would have been helpful to see concrete examples of fluency development activities that are suitable for formulaic language.

Here, I want to reiterate a key requirement for fluency development activities, that is, the target multi-word expressions used in such activities need to be familiar to their participants. This helps increase the rate of production without compromising accuracy. Prior knowledge of the target multi-word expressions will also increase the likelihood of their use in retelling and other free production tasks. This brings us to one of the challenges outlined in Thomson's paper, that is, how to ensure that the target language is used by the learners in free production.

There is one well-known caveat in teaching and learning formulaic language, that is, not all multi-word expressions are created equal (see Carrol & Conklin, 2019; Martinez & Schmitt, 2012; Siyanova-Chanturia & Martinez, 2015). Instructional approaches chosen to facilitate the learning of idioms, binomials, collocations and lexical bundles may need to be different. For example, in an eye-tracking study, Carrol and Conklin (2019) found that, beyond frequency, decomposability matters most for idioms, predictability and semantic association for binomials, and mutual information for collocations. This is why it is helpful to clarify what type or types of multiword expressions are targeted in the proposed instructional and learning treatments.

Overall, I enjoyed reading this paper because it took a careful and balanced approach in proposing research informed pedagogical treatments that could facilitate acquisition of multiword expressions. The author suggests that the combination of the three teaching ideas could promote better knowledge of L2 multi-word expressions and, overall, I agree. However, if fluency is, indeed, a key goal, I would prioritise multiple opportunities for processing and re-use over peer corrective feedback. Moreover, different stages of acquisition would require different instructional approaches: deliberate learning and explicit noticing through input enhancement and peer feedback may be particularly effective at the point of initial familiarisation and early learning, while increasing frequency of encounters (e.g., via input enrichment) and opportunities for re-use under time pressure may be more beneficial at later learning stages. These conjectures need to be tested and verified in future L2 learning studies with multi-word expressions.

## **2.2 The Challenges of Measuring or Assessing Multi-Word Expression Use in Conversation by Haidee Thomson, Hokusei Gakuen University and Victoria University of Wellington**

Haidee Thomson set out to investigate whether classroom interventions that require students to use and re-use four-word multi-word expressions, such as *I think I will* or *how do I get*, in structured learning activities and tasks promote their transfer into students' spontaneous interactions, such as those taking place in a role-play. In this study, she faced three key challenges: (1) how to score the accuracy of use, for example, whether or not to count incomplete or altered renditions of the original expressions as correct; (2) how to interpret the non-use of the target expression; and (3) technical and logistical challenges of assessing recorded spoken interactions in classroom research.

Challenges (1) and (2) are related. The solution to the challenge of scoring the accuracy of the target expressions in free production (e.g., *How can I go [to]*, instead of the target, *How do I get [to]*) and non-use (when the target expression is not produced), proposed by Thomson, is to follow free production with a cloze test, thus providing a form of triangulation. I agree that it is helpful to triangulate different measures, particularly so when a quantitative measure is accompanied by a qualitative one, in order to create a richer picture of the phenomenon or construct under investigation. It is important to think how the measures being triangulated are related to the phenomenon of interest. In a study that investigates fluency, Thomson proposes to use measures of controlled off-line production of multi-word expressions to triangulate measures of their free online production. Free spontaneous production relies, at least to some extent, on non-declarative (procedural, implicit) lexical knowledge, while a cloze test relies on declarative (explicit knowledge) and the use of task and linguistic strategies to come up with the right answer. Although the two knowledge types can support similar lexical processes, *fluent* speech is not possible when only explicit knowledge is used. So, even if it turns out that the results of the cloze test are correlated with the fluency measures, there is no guarantee that the cloze test scores would be *predictive* of fluent production, in general, or even the fluent production of the target multi-word expressions.

No doubt, measuring effectiveness of instructional and learning treatments in free production is hideously difficult. This is where psycholinguistic measures and approaches may help. Using behavioural (e.g., response time) measures of production under time pressure (e.g., using a shadowing paradigm), online access to the knowledge of multiword expressions can be tested under the conditions that limit participants' ability to use deliberate task and memory strategies, while the production of the target items is inevitable. Combing naturalistic free production with such behavioural production measures can get us closer to the desired triangulation that addresses the same kind of knowledge and access to this knowledge.

Thomson provides an excellent overview of the methodological issues and pitfalls that will be very useful for future studies that plan to audio-record students' classroom interactions. This experience underscores the importance of working out the study procedures and labelling approaches during the piloting stage. Researchers may also seek technological solutions to this problem, for example,

by using qualitative data analysis tools that link media files with participants' IDs, such as NVivo, or use multi-track recording software with separate input feeds for the two participants in dialogues (such as, Audacity).

Let us now return to the challenge of scoring partial responses. Thomson gives the following example: *How can I go (to)* produced instead of the target *How do I get (to)*. Because there were too few instances of complete expressions with all four words, she decided to score their production accuracy word-by-word. This presented its own challenge – whether to assign such an instance a score of 2 (because two out of the four words are the same as in the original lexical bundle) or a score of 4 (if the substitutions are counted as acceptable variants). In my opinion, there is a third option – not counting such expressions as correct, thus acknowledging that the classroom interventions used in this study did not have an observable effect on the development of free productive knowledge of the target expressions. Conceivably, *How can I go (to)* represents a free combination of four words put together in order to communicate a message, using individual known words. Indeed, is there a reason to assume that *How can I go (to)* is a pre-fabricated multi-word expression? This situation raises an important question: where do we draw the line between formulaic language and free word combinations and, at the risk of sounding radical, how important is it to have such a clear line? When teaching adults, in particular, it is just as important (if not more so), in my opinion, to have high quality of semantic knowledge of individual words, for example, *get* signifies a *completed* action (akin to the perfective aspect in other languages) and *go* signifies an *incomplete* action (akin to the non-perfective aspect). Presumably, such semantic/grammatical awareness would help learners to select *get* over *go* in the example provided. This thinking is aligned with Csomay's (2013) conclusion based on an analysis of the use of lexical bundles in university classroom discourse that, on the discourse level, there is a “strong relationship between grammar and lexis”.

Not to put in doubt the importance of having the knowledge of formulaic language, let me put it this way: if we know that formulaic language is resistant to explicit teaching and deliberate learning (especially if we are talking compositional, more transparent sequences, such as lexical bundles and grammatical collocations), why not simply encourage and facilitate learner exposure to comprehensible input (both spoken and written) that contains frequent lexical bundles and let statistical learning (Ellis, 2002) do its job? And, in the meanwhile, explicit teaching and deliberate learning time could be spent on the aspects of vocabulary knowledge that are more likely to benefit from this kind of learning, such as creating a robust form-meaning mapping for individual words (Elgort, 2011), idiomatic expressions (Obermeyer & Elgort, under review) and less compositional phrases, such as *not at all* (as a response to, *thank you*). This is where the teaching ideas proposed by Ogawa may also come handy; initially, input enhancement could promote noticing. Once learners start attempting these sequences in production, peer feedback could be helpful and, finally, as their accuracy of knowledge of these formulaic sequences develops, activities that facilitate fluency development become important. This would be in line with the *Four Strands* (Nation, 2007).

Now, let us briefly go back to the question of fluency. Both Ogawa and Thomson start their papers with the statement that fluency is an important goal

in second language learning, and that acquisition of formulaic language is a way to facilitate fluency attainment. So far, so good. However, the actual papers then proceed to the discussion of how best to facilitate the learning of multi-word expressions and evaluate the resulting knowledge, stopping short of making an explicit connection to fluency. Furthermore, the construct of fluency is not at all straightforward. Is fluency about a speed-up of some cognitive and motor processes or does it involve a qualitative change, a restructuring of lexical processing that optimises lexical access in comprehension and production (Anderson, 1982; DeKeyser, 2015; Segalowitz, 2000)? Another important question is the following: whose perception of fluency is accepted as the norm and what linguistic and non-linguistic features affect this perception? For native-speaker raters, for example, Saito, Ilkan, Magne, Tran, and Suzuki (2018) found articulation rate and pausing to be the distinguishing factors in fluency ratings of native English speakers and Japanese speakers of English of different proficiency levels. L2-speakers' perceptions of fluency, on the contrary, appear to be affected by a whole host of factors, in addition to the delivery speed and pausing (Magne et al., 2019), including lexical richness (the use of varied and sophisticated vocabulary) and the perception of speech as being *native-like*. Interestingly, an accurate online use of formulaic language could be seen as both a feature of lexical richness, sophistication (e.g., the use of idiomatic expressions) and of native-likeness (the use of lexical and grammatical collocations and lexical bundles).

### **2.3 Assessing Low-Level Cognitive Processes of Word Recognition by Michael Holsworth, Kyoto Sangyo University**

Michael Holsworth's paper considers three components of word recognition: orthographic, semantic and phonological knowledge, and proposes a battery of tests for measuring these components. Similar to Ogawa and Thomson, Holsworth is interested in the fluency and automaticity of access to lexical knowledge, but, this time, instead of spoken production, the focus is on visual word recognition and processing, such as that involved in L2 reading. Holsworth refers us to the early work by Charles Perfetti (Perfetti, 1985; Perfetti & Hart, 2001), suggesting that quality and efficiency of lower level word processing affects readers' ability to allocate cognitive resources to higher level processes, such as the construction of the global and local meaning of the text, integration of the new and old information and making inferences, which are needed for reading with understanding.

The author does not appear to make a distinction between knowledge and access to knowledge, often referred to as declarative and procedural knowledge in the L2 literature. Instead, if I understand correctly, Holsworth sees processing and access to knowledge as inherent in the quality of the knowledge itself. Taken one step further, this means that word learning is a process of improvement of the quality of lexical representations (Perfetti, 2007), understood as the establishment of (1) precise, fully specified orthographic representations, (2) redundant, word-specific phonological representations and (3) generalised and flexible semantic representations, accompanied by an increasingly stronger binding of these three types of representations.

To help language teachers evaluate the quality of L2 word knowledge of their students, Holsworth adopted and extended three tests of the constituent components of lexical quality. To measure orthographic knowledge, he used a lexical decision (LD) test that required participants to make word–non-word decisions; to measure semantic knowledge, he used an antonym-pairs test that required participants to identify pairs of items as antonyms or non-antonyms. According to Holsworth, both these tests had been originally developed by Matsuo (2017) and extended by Holsworth to include two additional sets of parallel items for longitudinal testing purposes. The items used in the tests were selected from the most frequent 4,000 words of the British National Corpus (BNC) corpus. The scoring of these tests only concerned response times (with incorrect responses excluded), as far as I understand. The third, pronunciation test, was based on the speaking section of the Wide Range Achievement Test (WRAT) (Wilkinson & Robertson, 2006) and extended by Holsworth to have an additional set of items. In this test, the scoring concerned the accuracy but not the speed of responses.

Overall, I really like the idea of evaluating the three components of lexical quality to guide EFL teachers' understanding of their students' lexical knowledge and planning of instructional interventions to extend this knowledge. However, I also have a number of questions and concerns about the proposed tests related to their validity, item selection, scoring procedures and the interpretation of their results. Firstly, let us consider the validity of the measures used to evaluate each of the components. Personally, I am not convinced about the use of response times in the LD task as a measure of the precision of L2 *orthographic* representations. The speed of LDs is not a very good measure of orthographic knowledge; rather, LDs measure lexical access. It has been shown that making an LD will not only involve accessing formal–lexical representations, but will also engage lexical–semantic representations of the word stimuli (Joordens & Becker, 1997; Masson, 1995; Neely, 1991). This is precisely why studies interested in orthographic representations and orthographic processing usually use orthographic priming (e.g., Andrews & Hersch, 2010; Forster & Veres, 1998) rather than non-primed LDs. Furthermore, higher *accuracy* of LDs may point to *more precise lexical representations*, while *faster responses* are more likely to indicate higher ease/fluency of lexical (but not, specifically, orthographic) processing. So, if the LD test described in this paper was, indeed, “designed to measure the orthographic knowledge of a reader” using response times, then the construct validity of the test may be violated. In fact, orthographic knowledge may be better measured by a dictation or speeded spelling judgment/verification test.

In addition, using Reaction times (RTs) in LDs as a measure of *lexical* processing assumes that LDs can be made at a certain level of accuracy (as pointed out in the paper). This does not appear to be the case for lower proficiency EFL speakers, who are the target audience for this test (based on the fact that only the first 4000 frequency-level words are tested). Grainger and Jacobs (1996) explain that both local and global lexical activation processes are involved in making ‘yes’ decisions in the LD task. The local activation criterion is met when the activation level reaches a threshold for a particular word, while the global criterion is a function of the overall level of activation in the lexical processor. In L1 LDs, the

local activation criterion is likely to be reached before the global one (provided the overall level of activation in the lexical processor is set high by using word-like non-words); that is, the decisions are likely to be based on the activation of the *lexical representations of the individual stimuli rather than on their word-likeness*. However, with EFL participants who may only have very weak lexical representations for L2 (English) words, the global criterion of *word-likeness* may be reached prior to the local one. This means that the LD may not be indicative of the knowledge of the test items. One of the signs that L2 lexical representations are unstable for lower proficiency EFL participants is their high rate of false alarms – “yes” responses to word-like non-word distractors.

Secondly, I have noticed some seeming inconsistencies in the selection of items used in the test of semantic representations, that is, the antonym pairs test. Some of the pairs appear to be much easier to judge than others, based on the strength of the semantic relationships between the two items. For example, it is very easy to make a decision that *right* and *left* are antonyms and that *laugh* and *drive* are not, but the decisions to word pairs, *guest* – *manager* and *kilometre* – *earth* are not straightforward. This means that the measure of response time in this semantic classification task does not only reflect the participants’ knowledge of and access to the semantic features of the test items but also the presence – absence of similar semantic features in the pairs of items and the level of ambiguity in the critical semantic feature. For example, for *late* – *early*, the critical feature is located on the time continuum and is in common for the two words in the pair, that is, *early* means before some expected time or near the beginning of a time period, while *late* means after some expected time or near the end of a time period; in other words, the participants are comparing  $(-)n$  with  $(+)n$ , where  $n$  is the set feature in common. The same analysis would work for *life* – *death*, that is,  $(-) \textit{being alive}$  versus  $(+) \textit{being alive}$ . For *laugh* – *drive*, the decision is also relatively straightforward (provided the meanings of the two words are known) because there is no semantic feature in common between the two items with a  $+/-$  sign. However, this fast decision-making approach will not work for *guest* – *manager* and, therefore, some degree of deliberation may be involved in making the antonym/non-antonym decision. For example, in a hotel, *guest* is a visitor and *manager* is someone who receives visitors; therefore, these two words could be antonyms; on the contrary, if the hotel schema is not invoked as part of the decision-making process, the two items could be perceived as not having semantic features in common and a “no” decision could be made. Thus, response times in this task reflect more than semantic knowledge; they reflect conscious explicit decision-making processes that are task specific and unrelated to the quality of semantic representations of the items involved. For a speeded antonym pairs test to be a valid measure of semantic knowledge, therefore, pairs of items need to have a consistent, unambiguous semantic relationship that would allow test-takes to identify them as antonyms or non-antonyms, without the need to engage metacognitive strategies.

Thirdly, the author states that the proposed pronunciation test was “designed to measure participants’ phonological awareness”. To me, this framing of the test is confusing; this is because, to measure phonemic or phonological awareness, some form of *recognition* or *segmentation* test is usually envisaged; however, instead, the test proposed by Holsworth involves a word pronunciation (i.e., production) task.

I would also question the rationale for specifically selecting the WRAT-4 speaking section to measure phonological awareness because the word reading section of this test was designed to measure “word decoding through letter identification and word recognition”, rather than phonological awareness (online information available from [www.pearsonclinical.com.au/products/view/567](http://www.pearsonclinical.com.au/products/view/567)).

I also want to comment on the practical aspects of conducting and scoring these three tests by language teachers. Because the measure used in the *lexical decision* and *antonym pairs* tests is response times (rather than response accuracy), the test will need to be taken in a quiet (ideally, sound-proof) computerised environment because millisecond-level response time differences need to be recorded. In addition, because Holsworth proposes not to remove extreme outliers (very slow responses), the integrity of the data analysis and, consequently, test results would be threatened by the presence of such outliers. In relation to the pronunciation test, the scoring appears to be quite laborious and would require seeking and training inter-raters for the purposes of test scoring. In my opinion, using a recognition or segmentation test format to measure phonological awareness would eliminate these requirements and make the test much easier to score and administer.

Finally, similar to the two papers discussed above, clear conceptualisations and/or operationalisations of *automaticity* and *fluency* are missing in this paper. Holsworth correctly points out the need for the lower level processes, such as word recognition and processing during reading to be “automatic and efficient”. However, the three tests proposed in the paper may not assess automaticity of word processing as conceptualised by Segalowitz and colleagues (e.g., Segalowitz & Hulstijn, 2005; Segalowitz & Segalowitz, 1993). Since the development of fluency and automaticity of low-level lexical processing is, indeed, a very important goal of language learning, I would recommend to explicitly operationalise them, clearly state what criteria would be indicative of their attainment (or different levels of attainment) and show how the proposed tests speak to these criteria. An idea of developing a battery of tests that assess lexical quality (as defined by Perfetti, 2007) goes a long way to address this; these tests need to be technically appropriate to target well-defined constructs representing specific knowledge dimensions while, at the same time, remaining accessible to language teaching practitioners. This is a very challenging task and Holsworth’s paper facilitates our thinking on how we can get there.

#### **2.4 Deliberate Vocabulary Learning from Word Cards by Darrell Wilkinson, Tampere University, Tampere, Finland**

The paper by Darrell Wilkinson takes another look at using word cards as a deliberate word learning method. He goes over some key reasons why the use of word cards may be more effective than using word lists, highlighting the importance of retrieval of form and meaning, opportunities for implementing different spacing regimes, learner control over the selection of items and flexibility in terms of the time and place of practice. He then shares the findings from his two experimental studies that evaluated the benefits of learners making their own word cards. In the first study, Wilkinson confirmed that the sheer act of making

word cards results in knowledge gains, and that these gains are greater on passive than on active recall. Unfortunately, some details of the study are not provided in the paper, for example, whether the target words were provided to the learners or whether learners chose their own learning targets, what measures were used to evaluate passive and active recall, and whether the initial knowledge of the target items was measured prior to the word card intervention. The second experiment compared the learning from pre-made and self/learner-made word cards. Wilkinson found that, on all measures (passive and active recall and retrieval), the short-term learning from pre-made word cards was superior compared to that from the self-made cards. However, long-term, this advantage of the pre-made cards dissipated and, in the end, the learning outcomes from these two approaches were not different.

These results are reassuring and suggest that both approaches to making word cards can be used, as and when appropriate. For example, teachers can make electronic word cards (using Quizlet) for the words students need to learn as part of their foreign language curriculum, and students can create their own word cards, that are relevant to their interests and learning goals and add them to their personal card decks and practice routines. This line of research could be further extended to check whether the use of audio recordings made by the teacher or by students themselves would better support word learning.

### 3 Conclusions and Recommendations

The four papers presented at the Eighth Annual JALT Vocabulary SIG Symposium raise important issues and are representative of some key direction in vocabulary research, namely, teaching, learning and acquisition of multi-word expressions, the development of fluency and automaticity in processing and production, and approaches to measuring and testing aspects of word knowledge. The authors demonstrate keen interest in drawing on and applying interdisciplinary insights and methods from cognitive psychology, psycholinguistics and education/learning science to applied linguistics research, which is laudable. This is in line with recent calls for interdisciplinarity in SLA research, for example, Rebuschat, Meurers, and McEnery (2017) in *Language Learning*; Suzuki, Nakata, and DeKeyser (2019) in *The Modern Language Journal* and our upcoming Special Issue of *Second Language Research on Issues in L2 lexical acquisition and processing: Setting an interdisciplinary research agenda* (with Anna Siyanova). Importantly, these issues are considered with practical outcomes in mind, that is, how to optimise foreign language learning and teaching practices. It has been a real privilege to discuss the ideas put forward by the four papers and suggest how they may be fine-tuned and extended. Thank you for the opportunity.

### References

- Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, 139(2), 299–318. doi:10.1037/a0018366

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369–406. doi:10.1037/0033-295X.89.4.369
- Carrol, G., & Conklin, K. (2019). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*. doi:10.1177/0023830918823230
- Choi, S. (2017). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*, 21, 403–426. <https://doi.org/10.1177/1362168816653271>
- Csomay, E. (2013). Lexical bundles in discourse structure: A corpus-based study of classroom discourse. *Applied Linguistics* 34(3), 369–388. doi:10.1093/applin/ams045
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). New York, NY: Blackwell.
- DeKeyser, R. (2015). Skill acquisition theory. In B. Van Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). New York, NY: Routledge.
- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61(2), 367–413. doi:10.1111/j.1467-9922.2010.00613.x
- Elgort, I. (2018). Technology-mediated vocabulary development in a second or foreign language: A critical review. *CALICO Journal*, 25(1), 1–29. doi:10.1558/cj.34554
- Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143–188. doi:10.1017/S0272263102002024
- Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning Memory and Cognition*, 24, 498–514. doi:10.1037/0278-7393.24.2.498
- Godfroid, A. (2020). Sensitive measures of vocabulary knowledge and processing: Expanding Nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). New York, NY: Routledge.
- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103, 518–565. doi:10.1037/0033-295X.103.3.518
- Hui, B. (n.d.). Processing variability in intentional and incidental word learning: An extension of Solovyeva and DeKeyser (2018). *Studies in Second Language Acquisition*, 1–31. doi:10.1017/S0272263119000603
- Joordens, S., & Becker, S. (1997). The long and short of semantic priming effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1083–1105. doi:10.1037/0278-7393.23.5.1083

- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*, 647–672. doi:10.1111/j.1467-9922.2010.00621.x
- Lee, S., & Huang, H. (2008). Visual input enhancement and grammar learning: A meta-analytic review. *Studies in Second Language Acquisition, 30*, 307–331. doi:10.1017/S0272263108080479
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*(2), 325–343. doi:10.3758/s13428-011-0146-0
- Matsuo, T. (2017). *The Roles of Lexical Size, Depth, and Automaticity of Word Recognition on Reading Comprehension* (Unpublished doctoral dissertation), Tokyo, Japan: Temple University.
- Meara, P. 1980. Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching and Linguistics: Abstracts, 13*, 221–246. doi:10.1017/S0261444800008879
- Magne, V., Suzuki, S., Suzukida, Y., Ilkan, M., Tran, M., & Saito, K. (2019). Exploring the dynamic nature of second language listeners' perceived fluency: A mixed-methods approach. *TESOL Quarterly, 53*(4), 1139–1150. doi:10.1002/tesq.528
- Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. doi:10.1093/applin/ams010
- Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 3–23. doi:10.1037/0278-7393.21.1.3
- Nation, P. (1989). Improving speaking fluency. *System, 17*(3), 377–384. doi:10.1016/0346-251X(89)90010-9
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching, 1*(1), 2–13. doi:10.2167/illt039.0
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.
- Obermeyer, A. & Elgort, I. (2020). *Figurative expressions at the interface: The effects of deliberate and contextual learning on semantic processing*. Manuscript submitted for publication.
- Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford Press.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383. doi:10.1080/10888430701530730
- Perfetti, C. A., & Hart, L. (2001). The lexical bases of comprehension skill. In D. Gorfien (Ed.), *On the consequences of meaning selection* (pp. 67–86). Washington, DC: American Psychological Association.
- Rebuschat, P., Meurers, D. & McEnery, T. (2017). Language learning research at the intersection of experimental, computational, and corpus-based approaches. *Language Learning, 67*, 6–13. doi:10.1111/lang.12243

- Saito, K., Ilkan, M., Magne, V., Tran, M. N., & Suzuki, S. (2018). Acoustic characteristics and learner profiles of low-, mid- and high-level second language fluency. *Applied Psycholinguistics*, *39*, 593–617. doi:10.1017/S0142716417000571
- Schmitt, N. (2010). *Researching vocabulary*. Basingstoke: Palgrave.
- Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 200–219). Ann Arbor, MI: University of Michigan Press.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York, NY: Routledge.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford, UK: Oxford University Press.
- Segalowitz, N. S., & Segalowitz, S. J. (1993). Skilled performance, practice, and the differentiation of speed-up from automatization effects: Evidence from second language word recognition. *Applied Psycholinguistics*, *14*, 369–385. doi:10.1017/S0142716400010845
- Sharwood Smith, M. (1991). Speaking the many minds: On the relevance of different types of language information for the L2 learner. *Second Language Research*, *7*(2), 118–132. doi:10.1177/026765839100700204
- Sharwood Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*(2), 165–179. doi:10.1017/S0272263100011943
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36*(5), 549–669. doi:10.1093/applin/amt054
- Solovyeva, K., & DeKeyser, R. (2018). Response time variability signatures of novel word learning. *Studies in Second Language Acquisition*, *40*(1), 225–239. doi:10.1017/S0272263117000043
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge acquisition of collocations under different input conditions. *Language Learning*, *63*, 121–159. doi:10.1111/j.1467-9922.2012.00730.x
- Suzuki, Y., Nakata, T. & DeKeyser, R. (2019). Optimizing second language practice in the classroom: Perspectives from cognitive psychology. *The Modern Language Journal*, *103*, 551–561. doi:10.1111/modl.12582
- Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, *26*, 245–265. doi:10.1111/ijal.12092
- Toomer, M. & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning*, *69*, 405–439. doi:10.1111/lang.12335
- Wilkinson, G. S., & Robertson, G. (2006). *Wide range achievement test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.