Understanding Vocabulary Acquisition, Instruction, and Assessment: A Research Agenda

Norbert Schmitt

University of Nottingham, Nottingham, UK

norbert.schmitt@nottingham.ac.uk

Norbert Schmitt is Professor of Applied Linguistics at the University of Nottingham. He is interested in all aspects of second language vocabulary acquisition, use, pedagogy, and assessment, and has published widely in these areas. He has written a manual to guide vocabulary research (*Researching Vocabulary*), and is currently revising his 2000 volume *Vocabulary in Language Teaching* (along with Diane Schmitt). His current interests include a British Council project describing the acquisition order of the first 5,000 lemmas of English (with Barry O'Sullivan, Laurence Anthony, Karen Dunn, and Benjamin Kremmel) and developing the next-generation internet-based vocabulary size test (With Benjamin Kremmel).

Norbert Schmitt

School of English

University of Nottingham

Nottingham, NG7 2RD

UK

Abstract

This paper suggests six areas of vocabulary research which the author believes would be fruitful for future research. They include 1) Developing a practical model of vocabulary acquisition, 2) Understanding how vocabulary knowledge develops from receptive to productive mastery, 3) Getting lexical teaching/learning principles into vocabulary and language textbooks, 4) Exploring extramural language exposure and how it can best facilitate vocabulary acquisition, 5) Developing more informative measures of vocabulary knowledge, and 6) Measuring fluency as part of vocabulary competence. Nine tasks are suggested for how to research these six research directions, with advice on research design and how to set about carrying out the tasks.

Vocabulary is currently one of the most popular topics in applied linguistics research, with Rod Ellis (2009: 335) noting that 'It is probably true that to say that during my editorship of *Language Teaching Research* there have been more articles published on vocabulary teaching than on any other topic.' In a recent THINKING ALLOWED feature, Nation (in press) states that over 30 per cent of the research on L1 and L2 vocabulary learning in the last 120 years has occurred in the last 12 years. As a result of this research, we have a much better understanding of the nature of vocabulary, how much and which vocabulary is required to do things in a second language (especially English), and how to most effectively teach and learn the required vocabulary. Nevertheless, there are still large gaps in our knowledge of key aspects of vocabulary. In this article, I will discuss six issues which I feel deserve attention, and are worth addressing as a priority. If these points were taken up, I believe the resulting research would provide tangible improvements in vocabulary pedagogy and assessment, and so I suggest them as part of a vocabulary research agenda for the next 10 years.

1. Developing a practical model of vocabulary acquisition

Paul Meara first noted the lack of an overall theory of vocabulary acquisition in 1983, and this is still true today. Of course, there have been numerous theories which cover limited aspects of vocabulary learning. For example, the Revised Hierarchical Model (Kroll & Stewart 1994), posited that the psycholinguistic pathway to L2 meaning is initially through L1 translation equivalents. Ellis (2002) reviews how frequency partially drives language acquisition, including individual words and formulaic language. Brown and Payne (1994, cited in Hatch & Brown, 1995) proposed a five-step model of vocabulary learning, although it only dealt with form and meaning. In fact, the very few theories/models available have tried to explain how basic form-meaning connections are created (sometimes in relation to L1 lexicon entries), but there are still none that explain how the many different components of lexical mastery are developed. This is partly because vocabulary knowledge remains an extremely complex construct, which resists any single explanation. This consists of (ideally) very large numbers of both individual words with their inflections and derivatives, and of formulaic sequences. Every lexical item has its own characteristics which may make it relatively easier or more difficult for any particular learner to acquire, with L1 being a major factor (Laufer 1997). It is probably impossible to devise any explanation of acquisition which can reliably describe how every intrinsically-different lexical item is acquired by learners of various L1s.

The best-known and most widely-used framework is Nation's (2013) division of vocabulary knowledge into nine components of 'word knowledge' (e.g. spelling, word parts, meaning, grammatical functions, collocation). The framework has been extremely useful in describing the totality of WHAT learners need to know, but says nothing about any HIERARCHAL ORDERING (i.e. which components are typically learned before others, or should be taught before others). This limits its pedagogical value, because it is unclear how the various components relate to each other, and how to best prioritize the various components when teaching.

A limited number of studies have measured knowledge of multiple components from the framework for various research purposes (e.g. Schmitt 1998; Webb 2005; Chen & Truscott 2010), and generally found that the various measured components were interrelated, but that some appeared to be acquired before others. The most comprehensive of these multicomponent studies was carried out by González-Fernández (2018). She found that all of

the components she measured (form-meaning, derivatives, collocation, polysemy) were known to a recognition level before any of those aspects were known to a recall level. This suggests the key descriptor of vocabulary knowledge may not be the word knowledge components themselves (as implied by Nation's framework), but rather recognition vs. recall mastery of those components. She also found knowledge of the components conformed to an implicational scale, i.e. that some components were learned before others.

If her acquisition order is shown to be generalizable, this would give teachers and testers a blueprint for how (at least a portion of) vocabulary knowledge is acquired. But the research first needs extending to other word knowledge components and other L1s. This leads to my first two research suggestions:

Task 1. Carry out cross-sectional multi-component test batteries of vocabulary knowledge, measuring both receptive and productive mastery, to better understand the relationships between 1) receptive and productive mastery, and 2) the various word knowledge components.

Task 2. Carry out the Task 1 test batteries in longitudinal studies to better understand how receptive/productive mastery and word knowledge components develop over time in individuals.

In an ideal world, it would best to measure all of the word knowledge components (both receptively and productively) in the same test battery. However, this seems virtually impossible in practical terms, as González-Fernández's battery of only four components took between 2.5-3.5 hours for most learners. González-Fernández's study could be usefully replicated (Porte 2012), but it should also be extended by measuring word knowledge components she did not (spelling, pronunciation, word parts, associations, grammatical functions, constraints of use). By also including one or two of the components she did measure as a comparison, it should be possible to relate any new study's results to the existing implicational scale. For example, González-Fernández found that recognition of the form-meaning link was typically mastered before recognition of derivatives. If a new study found that recognition of correct spelling was mastered before recognition of the form-meaning link, it could be inferred that recognition of correct spelling is also typically mastered before recognition of derivatives. The comparison would be further enhanced if González-Fernández's target words were used in the new studies. Another approach would be to measure all components, but with different target words, carefully controlling the target items to be similar across components. Although this has the limitation of not measuring the same words across the different components, the value of more components being measured concurrently may be worth the trade-off. It also eliminates any possibility of TEST CONTAMINATION (where exposure to a target word on one test in the battery may give hints to answering a subsequent test).

While cross-sectional studies can be informative, longitudinal studies are usually better at describing acquisition. This would require re-administering the test battery again after a semester or year, to see how the various components were enhanced and to what degree. If this proved difficult to arrange for groups of students, such a longitudinal study might be well-suited to a case study approach, if several very cooperative learners could be found.
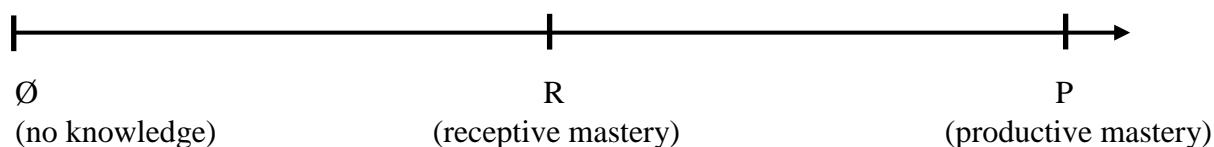
González-Fernández found the same results with Spanish learners of English (cognate language) and Chinese learners (noncognate language). But it would also be useful to extend

the research to other L1s, as previous research (e.g. Otwinowska & Szewczyk 2017) shows that L1 is an important factor in L2 vocabulary acquisition.

Eventually it should be possible to model which word knowledge components were typically learned before others (for most words and most learners), or alternatively, whether some components are inherently idiosyncratic regarding their learning burden (i.e. whether they are learned earlier or later depends mainly on the particular word or learner). It should also become clear whether recognition mastery always precedes recall mastery for all components, or just for the four that González-Fernández measured. The results would inform pedagogy in showing where most teaching effort needs to be applied: in moving vocabulary knowledge from Ø → RECEPTIVE, or from RECEPTIVE → PRODUCTIVE. It would also inform testing, because if an implicational scale could be identified for all/most components, it would only be necessary to test a few components to infer what other components were known.

2. Understanding how vocabulary knowledge develops from receptive to productive mastery

Point 1 is about understanding the various aspects of vocabulary knowledge and how they relate to each other. This point focuses more on acquisition, especially how to move learner knowledge to the more advanced productive level. There is plenty of research, along with teacher experience, to show that receptive mastery of a lexical item (ability to understand it when listening or reading) is generally stronger than productive mastery (ability to produce it in one's own speech or writing). Virtually all research which includes both receptive and productive measures show higher receptive scores (e.g. Melka 1997; Laufer & Goldstein 2004; Webb 2005). (See Point 5 below for more discussion on receptive/productive and recognition/recall measurements.) But research generally just reports the receptive and productive scores, without considering the relationship between the two. In simple terms with a continuum-based illustration, the relationship might be exemplified like this (although note that a 'states'-based conceptualization is also possible (e.g. Meara 1997)):

```
├────────────────────────┼────────────────────────┼───────────────►
Ø                        R                        P
(no knowledge)           (receptive mastery)      (productive mastery)
```

Many might think that the intervals (i.e. learning burden) is about equal between Ø→R and R→P for most words, as illustrated above. Others might think the major learning is in learning the word in the first instance to receptive mastery, and after that productive mastery follows on without too much trouble:

```
├───────────────────────────────────────┼───────────────┼───────────►
Ø                                        R               P
```
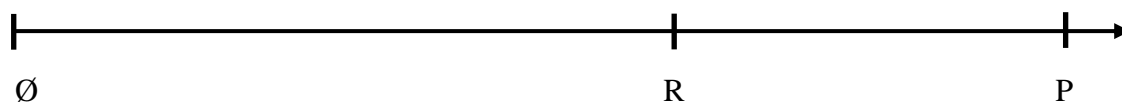
But I think the research indicates that the opposite is true: that learning most words to receptive mastery is relatively easy; it is enhancing that knowledge to productive mastery which is the real challenge:

```
├────────────────┼──────────────────────────────────────────────────┼──►
```
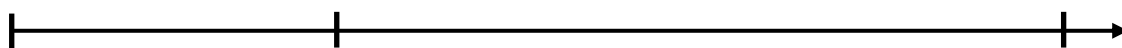
Ø   R   P

It is not difficult to understand why this might be so, as I explain in Schmitt (2014). In reading, to comprehend a word, it might be enough to be able to recognize the spelling of a word and remember its meaning. All or most of the other word knowledge components are already in the text (e.g. its collocation and derivative form), which may or may not be utilized to aid comprehension. But when writing, one must know and produce all of the various components independently without prompts. The same holds true for listening/speaking.

I think it is safe to assume that most learners want to be able to employ their vocabulary in their speaking and writing. So for me, an under-researched area of particular interest is how to push learners' knowledge from receptive mastery to the point where they can independently use lexical items fluently and appropriately in their own output. This leads to my next research suggestion:

Task 3. Investigate various vocabulary-learning exercises and activities to determine which best improve vocabulary knowledge from receptive to productive mastery.

Note that this task is not about determining which vocabulary activities are best to *begin* the learning process, but which are best to enhance partially-known vocabulary. These questions will probably require a pretest-treatment-posttest design, with target lexical items being tested both receptively and productively. The targets will need to be known receptively (but not productively) at the beginning of the study. Because it is difficult to know how well individual lexical items are known, a range of items will probably have to be given in the pretest, and only those known receptively selected for inclusion in the study.

Most studies have used multiple-choice items as their receptive format, but these are probably not the best choice, because guessing will almost certainly inflate the scores to some unknowable extent (Gyllstad, Vilkaite, & Schmitt 2015). I suggest using MEANING RECALL formats for the receptive tests and FORM RECALL formats for the productive tests (see Point 5). This virtually eliminates guessing, and also better matches the lexical knowledge necessary for the receptive and productive skills (reading/listening; writing/speaking) (Schmitt 2010, 2014). Using L1 translations is a good way to operationalize these recall measures if one is testing a homogenous L1 group. With heterogeneous groups, it will be necessary to use alternative meaning-based prompts, such as L2 definitions or higher-frequency L2 synonyms.

Research suggests that it takes some time to build up to productive mastery, which prompts Read (2000: 154) to pose the very interesting question: 'Is there a certain minimum amount of word knowledge that is required before productive use is possible?' Answers to receptive-productive questions like this will probably require longer-term longitudinal studies. Just how long is probably a research question in itself, but studies with too short a treatment period (I speculate less than 6 months) will likely show little change from receptive to productive knowledge. (There is also the issue of APPROPRIACY of use, which I do not have space to address in this piece.)

Any one iteration of a learning activity will be unlikely to result in truly productive knowledge, so research should also enquire how many repetitions it takes to move knowledge to the productive level. The most effective methodology might entail a combination of activities, and so this should also be explored. Research suggests learners must practice vocabulary productively to reach productive mastery (e.g. Laufer 2005), so the activities

investigated will almost certainly require learners to produce output, rather than just practice receptively. It will only be possible to study limited numbers of activities in any one study, so I foresee this strand requiring a number of studies, and only by eventually synthesizing them will we get the bigger picture of the kind of vocabulary activities we should promote after initial (receptive) learning is in place.

3. Getting lexical teaching/learning principles into vocabulary and language textbooks

There has been enough research on vocabulary to suggest a number of principles of good vocabulary instruction. Hunt and Beglar (1998) proposed seven, including:

• Diagnose which of the 3,000 most common words learners need to study

• Provide opportunities for elaborating word knowledge

• Provide opportunities for developing fluency with known vocabulary

• Experiment with guessing from context

In Schmitt (2008), I added another six points, including:

• Learners need large vocabularies to successfully use a second language, and so high vocabulary targets need to be set and pursued.

• Vocabulary learning is a complex and gradual process, and different approaches may be appropriate at different points along the incremental learning process.

• Once this initial meaning–form link is established, it is crucial to consolidate it with repeated exposures.

• It is also important to begin enhancing knowledge of different aspects of word knowledge. Some of these may be usefully learned explicitly (e.g. knowledge of derivative forms), but the more 'contextualized' word knowledge aspects (e.g. collocation) are probably best learned by being exposed to the lexical item numerous times in many different contexts.

While principles like these are sound, and would surely lead to better vocabulary pedagogy, the problem is that most teachers do not have the time, expertise, or resources to consistently put them into practice. Take the idea of recycling (repeated exposures). For example, if a teacher taught 10 new words a class, it might be possible for her to conscientiously recycle all those words for awhile. But eventually, it will become unmanageable to recycle 100-200+ words in a systematic manner.

This task needs to be done by someone with the time and resources to carefully consider 1) which vocabulary is most beneficial for learners (largely high-frequency vocabulary), 2) which activities most effectively teach these lexical items (some activities may be better suited to certain lexical items than others), and 3) how to systematically build recycling and enhancement into a course longer-term. Syllabus designers, and particularly textbook writers (who typically take a year or more to write their books), are best positioned to organize this thoughtful development of vocabulary.

Unfortunately, most textbooks lack any obvious systematic approach to vocabulary. Many (most?) textbooks are built around some kind of reading passage, and the vocabulary highlighted largely depends on the topic of that passage. This vocabulary is seldom repeated

later as the topics for the next chapters change. This can lead to quite haphazard vocabulary selection (Schmitt & Schmitt 2014). For example, Hsu (2009) examined 20 international General English textbooks, ranging from low intermediate to advanced levels. She found little uniformity between the level of the textbook and the vocabulary required both within and across textbook series. For example, the advanced *Reading for Real* required 4,000–4,500 word families to reach 95% coverage, while the low intermediate *Reading for Success 2* required 7,000–7,500 families. Another ramification of unsystematic vocabulary selection is a lack of recycling. In one example of this, Matsuoka & Hirsh (2010) analyzed the vocabulary from the best-selling *New Headway Upper Intermediate* English textbook and found that of the 1,005 beyond-high-frequency word families, 66.4% occurred only once, and only 12.1% occurred five times or more. Finally, the vocabulary activities in textbooks are often quite limited in what they teach. Brown (2010) found that the nine General English textbooks he analyzed focused mainly on meaning and form, with some attention to grammatical function and spoken form, but did little to enhance knowledge of other types of word knowledge like collocation, derivative form, or constraints on use. I think it would be useful to find out why vocabulary teaching principles are generally not making it into textbooks:

Task 4. Interview publishers, commissioning editors, and textbook writers to determine why established vocabulary teaching principles are not generally incorporated into textbooks. What are the constraints writers face, and what can be done to make textbooks more pedagogically-sound from a vocabulary perspective?

I am not aware of any research which looks at how publishers and textbooks writers go about selecting the vocabulary for their books, and how they develop the activities for that vocabulary. Therefore, a logical first step is to interview them, to determine how much they take account of vocabulary research in their textbook development. Some obvious lines of questioning include the following:

• Are textbook writers aware of the vocabulary research in the first place?

• Is vocabulary development an essential component as they conceptualize their books?

• If so, how do they attempt to operationalize it in their textbooks?

• What are the constraints which keep them from more fully applying the principles in their books?

• If they do not believe it is important to highlight vocabulary, is this because they do not believe the end consumer wants it? Or is vocabulary simply too difficult to organize over time? Or are there other reasons?

• What guidance/guidelines, if any, do writers get from publishers concerning vocabulary content? What guidelines would they like to get that would make a more principled and systematic approach to vocabulary inclusion in textbooks possible?

In addition to this publisher-based research, it would be very useful to explore teachers' beliefs, attitudes, and usage of textbooks, to explore 1) how these might affect

publishers' attitudes, and 2) how the presently-available vocabulary textbook activities are being utilized.

Most schools and teachers rely heavily on textbooks for language content, so I believe it is essential that these books incorporate sound vocabulary principles in a way that the average teacher would never have the time to do properly, i.e. textbooks are the essential conduit for research to influence practice. It is time to find why this influence has been so meagre to date, and to think of ways of redressing the problem.

4. Exploring extramural language exposure and how it can best facilitate vocabulary acquisition

While many learners around the world struggle to learn 2,000 English words after schooling of hundreds, or even a thousand, hours (Laufer 2000), children in other countries come into school knowing substantial amounts of vocabulary. This is especially true in northern Europe. What makes these countries different? It probably has something to do with the social attitudes that English is useful and worth knowing (e.g. de Wilde & Eyckmans 2017). But the more important factor is likely to do with the exposure to English these children enjoy. Recent research has shown that children in some countries are exposed to English for several hours per week (e.g. Lindgren & Muñoz 2013; Jensen 2016; de Wilde & Eyckmans 2017). This has led to young learners having impressive vocabulary sizes for their age. For example, in Belgium, learners scored a mean of 66.20 out of 108 on the Peabody Picture Vocabulary test, despite not having had any formal English instruction when they took the test (de Wilde & Eyckmans, 2017). In Iceland, Lefever (2010) found that most children before the start of formal education could already understand basic spoken English, many could participate in simple conversations in English, and most were in the first stages of understanding written English. Clearly, these learners are acquiring considerable vocabulary and language outside the classroom, and some must come from media exposure (Kuppens 2010). There is even some evidence that out-of-class exposure has a larger effect than length of instruction (Peters 2018).

Three types of extracurricular exposure (sometimes referred to as EXTRAMURAL EXPOSURE) usually found to be important include: watching English-language television (with subtitles or captions) (e.g. Kuppens 2010; Lindgren & Muñoz 2013), playing computer/internet English games (e.g. Kuppens 2010; Sylvén & Sundqvist 2012; Jensen 2016), and particularly for older learners, consuming English-language reading material, whether on paper or on the internet (e.g. González Fernández & Schmitt 2015; Garnier & Schmitt 2016; Macis & Schmitt 2017). However, these studies do not typically indicate the relative importance of these various kinds of extramural exposure, nor examine in much detail the precise nature of the exposure (e.g. the prominent English features in the computer games being used). Having finer-grained detail about the nature of extramural exposure, and studying how this directly leads to L2 acquisition, would allow more concrete suggestions about how to promote the most effective use of extramural exposure in a range of contexts, and how to best integrate it with classroom instruction.

Task 5 Determine how to maximize the benefits of extramural exposure for vocabulary acquisition.

Task 6 Analyze computer games for the type of vocabulary they contain.

Most research into extramural exposure has used surveys or interviews to determine the category (e.g. watching TV or playing computer games in the L2) and the extent of exposure, and then matched the answers with learner scores on vocabulary and other language tests. As shown above, this methodology has demonstrated that extramural exposure of various types is related to better L2 proficiency. What is less clear is how maximize the benefits of this exposure for vocabulary acquisition. Research on several approaches could be useful. The first involves marrying supplementary explicit instruction to the extramural exposure. Two small studies (Miller & Hegelheimer 2006; Ranalli 2008) have shown that supplementary materials (e.g. word lists of the vocabulary in SIM games and vocabulary exercises) lead to better vocabulary learning when playing the games. Potential research studies could match extramural exposure with a variety of supplementary materials provided in language classrooms to determine which types of material are the most effective in promoting vocabulary growth. Some obvious candidates include: lists of words from the extramural exposure, a variety of explicit exercises focusing on those words, and strategy training exercises to help learners manage and learn new words (e.g. exercises which train learners to concurrently focus on both L2 audio and L1 subtitles when watching subtitled media (Kuppens 2010)). This approach could be informative about which supplementary materials are most effective with particular types of extramural exposure, or whether some combination is best. It would be useful to include three conditions: one in which learning gains from the extramural exposure alone were measured, one where the learning from just the explicit materials was measured, and a combined condition. This would illustrate how much learning accrues from the interaction of exposure and materials, and how much simply comes from the explicit materials themselves.

Another approach would involve using multiple vocabulary tests to determine what vocabulary knowledge comes from different types of extramural exposure. By using both recognition/receptive and recall/productive tests, it should be possible to better describe the degree of mastery that extramural exposure typically leads to.

Gaming is an important type of extramural exposure for many learners, and it would be useful to better understand what kind of games are most beneficial. Jensen (2016) is a model of how finer-grained analyses can prove informative. She classified the games into the following categories: games with both oral and written English input, with oral but no written English input, with written but no oral English input, with oral English input and written Danish input, and Danish oral input and written English input. She found that gaming with 'both spoken and written English input' was significantly related to scores on the Peabody Picture Vocabulary Test, and to a much lesser extent, gaming with 'only written English input'. Likewise, different games might be better for boys vs. girls. Sylvén and Sundqvist (2012) found that not only did boys engage in gaming more than girls, but they tended to play different types of games; they preferred first-person shooter or multiplayer games, while girls tended towards single-player simulation games. This makes a difference because Sylvén and Sundqvist consider multiplayer games (particularly Massively Multiplayer Online Role-playing games, such as *EverQuest 2* and *World of Warcraft*) 'highly beneficial for L2 acquisition because they provide learners with opportunities for engagement with rich target language input [and output] as well as for scaffolded interaction' (p. 315).

Further finer-grained research into games could explore whether different types of game promote different types of vocabulary (e.g. Do shooter games highlight 'action' verbs?, Do multi-player games promote the vocabulary of commanding or negotiating?). An analysis of the vocabulary in games could show what words are being presented, e.g. the percentage of high-/mid-/low-frequency words through a Lextutor analysis (www.lextutor.ca),

vocabulary that realizes language functions through referral to references such as Nattinger and DeCarrico (1992), and specialist vocabulary. A lexical analysis could also show whether words are being repeated enough to make incidental learning viable (Cobb 2007), although this would also require the additional step of determining the number of repetitions required for this kind of input, as it might differ from reading, on which most incidental learning research has been carried out. Other research directions include exploring the role of gaming visual input in vocabulary learning (e.g. in what ways does the visual input support the acquisition of vocabulary?), and the nature of gaming interaction (e.g. what kinds of interaction are most facilitative for vocabulary learning?).

5. Developing more informative measures of vocabulary knowledge

Any research into vocabulary acquisition is only as good as the tests used to measure the learning, and most vocabulary tests are not validated to any great degree (Schmitt, Nation, & Kremmel in preparation). Most studies have used only a single measure, so in Points 1 and 2, I suggested the value of using test batteries, including measures of multiple types of word knowledge, and at receptive and productive masteries (see also Webb 2005). Furthermore, most studies to date have measured only some aspect of the form-meaning link, often with a multiple-choice format (i.e. at the 'recognition' level of mastery), so I have also argued that it is better to measure form-meaning knowledge at the levels of meaning-recall and form-recall (Point 2).

Until now, I have used the terms RECEPTIVE and PRODUCTIVE quite loosely when speaking about testing, as have most commentators. What we really want in vocabulary measurement is the ability to infer what learners can DO with the target words. (Nobody interprets test scores as simply words that learners can answer on a vocabulary test!) Receptive knowledge entails knowing a lexical item well enough to extract communicative value from speech or writing. Productive knowledge involves knowing a lexical item well enough to produce it when it is needed to encode communicative content in speech or writing. That is, receptive/productive knowledge of vocabulary is usage-based, and should presumably be measured with skill-based instruments. However, it is hardly ever measured this way.

This is because it is very difficult to measure vocabulary knowledge in context. All skills require more than just vocabulary knowledge. A target word interacts with the other words in a text in sematic, grammatical, morphological, and phraseological ways, and thus it is difficult to measure understanding or production of a single lexical item in context without the context becoming part of what is being measured. This is why vocabulary is often measured in isolation, and usually at just the form-meaning link level. Typical tests involve measuring target words with either a multiple-choice or matching test format which learners RECOGNIZE the correct form or meaning, or with the meaning given and the L2 word form needing to be RECALLED (form recall). Alternatively, the form can be given, with the meaning needing to be RECALLED (meaning recall).

Receptive tests are easier to develop because test writers are in control. They can select the target items, and embed them in contexts which are not too informative, so that that the meaning cannot be inferred from context. But just because a learner can understand a lexical item in one non-defining test context, does that mean they will be able to understand it in a variety of real-world contexts? Then there is the issue of oral/written language. Does the ability to comprehend a lexical item on a written test imply the ability to understand it when

listening? Or vice versa? These issues suggest the need to better understand how receptive test formats work:

Task 7. Determine what receptive test formats show about the ability of learners to comprehend target words in reading and listening.

One approach to this task is to investigate current item formats in terms of the information they give about the ability to understand words when reading or listening. A recent study showed that different existing formats do have explanatory power in predicting receptive skills, in this case, OVERALL VOCABULARY SIZE in predicting general reading proficiency. Laufer & Aviad-Levitzky (2017) found that both meaning recognition and meaning recall vocabulary measures correlated with reading scores at .91-.92. But it would be even more interesting to discover what the test formats showed about the ability to understand the PARTICULAR TARGET WORDS when reading. Target words could be placed in existing test formats, and then also embedded in multiple authentic reading and listening contexts. Comprehension questions would then test understanding that was directly reliant on knowledge of the target words. It would also be important to control and limit the surrounding context so answers could be not arrived at through lexical inferencing strategies. The ability of learners to comprehend the target words in a variety of reading/listening texts would be the benchmark of what learners can do with the target words. This benchmark could then be compared the scores from the various item formats to see how well each format reflects the reading/listening comprehension of the target words. This type of CONCURRENT VALIDATION would be very useful in knowing how to interpret the scores from the various test formats. If a quick and simple item format was shown to reliably predict the ability to comprehend a word, this would be a very good result, as large numbers of words could be tested, and practitioners and researchers could be confident whether learners actually knew the words well enough to understand them in real-world language.

However, it may turn out that existing formats are unable to adequately describe comprehension. A second approach would be to develop new formats. Space limitations prohibit discussion of the many possible formats, but good places to look for inspiration are *Tools for Researching Vocabulary* (Meara & Miralpeix 2017) and Paul Meara's *lognostics* website (*www.lognostics.com*), which illustrate a number of interesting experimental formats. But whether Meara's formats or completely original ones are explored, they will need the same kind of concurrent usage-based validation evidence as discussed above.

Turning to productive tests, things get trickier. Most measures are post-hoc computerized analyses of learner output (e.g. type-token ratios, lexical sophistication, lexical density – see Coh-Metrix (www.cohmetrix.com) and TAALES (www.kristopherkyle.com/taales.html). But these merely describe the lexical items produced, not a learner's complete productive vocabulary. Just because a learner produces one word in speaking or writing (e.g. *difficult*), it does not mean that they do not know other possibilities, like *onerous* or *hard*; they just happened to select *difficult* in that instance. Any output a learner produces will inevitably contain only a small percentage of the words they were capable of producing. This makes it almost impossible to DIRECTLY measure the complete range of a learner's productive vocabulary. However, there are ways in which it might be INDIRECTLY measured, and I think these are worth pursuing.

Task 8. Develop a test of productive vocabulary knowledge through establishing the ratio of receptive-to-productive knowledge.

One indirect approach which I think holds promise (and is relatively doable) entails potentially establishing the ratio between receptive and productive knowledge. There have been many studies which compared 'receptive' and 'productive' knowledge, but they have mainly looked at only the form-meaning link (e.g. Fan 2000; Laufer & Goldstein 2004). My suggestion involves sampling a large, representative number of words, either from frequency lists or from a forthcoming list of the best-known lemmas in English (Schmitt, O'Sullivan, Anthony, Dunn, & Kremmel in preparation).[1] The researcher would then have learners say/write sentences for each of these words. From my experience, participants often produce sentences which do not really show the meaning/usage of the target word (including one very naughty example 'I like the word *raspy*'), so it might be necessary to do the task one-on-one to probe further and ask for new sentences if necessary. Developing a set of criteria for acceptable answers would also be important. This approach would be very time-consuming, and would probably have to be carried out as case studies. But the end result should be a good estimate of the learners' productive vocabulary. This estimate could then be compared to scores from one or more receptive tests. If the productive/receptive ratio was relatively constant across learners, then only a receptive test need be given, and a learner's productive vocabulary size could be calculated using the established ratio. Even if the ratio proves not to be stable across learners, the results of this study would still be a valuable contribution towards understanding the relationship between productive and receptive vocabulary.

6. Measuring fluency as part of vocabulary competence

Being able to employ vocabulary in the four skills involves more than just knowledge of lexical items. It is also necessary to be able to listen and speak in real time, or interlocutors will soon tire of the halting, disfluent communication. Likewise, if learners are unable to read at a sufficient rate, the slow word-by-word decoding makes it difficult to understand the constantly-developing meaning structure of an extended text. For fluent reading, words need to be recognized quickly, automatically, and accurately (Grabe 2009). This leads Daller, Milton, & Treffers-Daller (2007) to propose *fluency* as one of the three basic components of vocabulary knowledge (in addition to size and depth). However, while there has been considerable research into vocabulary size and some into depth (see Schmitt 2014 for an overview), there has been very little into measuring the fluency with which vocabulary can be employed, or how it can be acquired. But if fluency is seen as an essential requirement for using vocabulary, we should begin measuring it in our vocabulary research.

Task 9. Explore the Degree to which Vocabulary Activities Develop Fluency in Use.

This task could be seen as an extension of Point 5's encouragement to develop and use more informative measures of vocabulary. But how can we measure fluency in vocabulary use? The trick is measuring the fluency of the individual lexical items of interest, rather than just overall reading/listening/speaking/writing speed. I think there are number of methodologies that might prove useful for this. For reading, eye-tracking can show the fluency in which individual words and phrases are read, in terms of the number and duration of the eye fixations on the target items (e.g. Pellicer-Sánchez 2016). Timed lexical decision

tasks have also been widely used for the measuring fluency at the individual word level. In writing, keystroke-logging software can show how fluently the target items are typed, in terms of duration and corrections (e.g. Miller, Lindgren, & Sullivan 2008). For listening, measuring comprehension of ideas realized by target vocabulary in real-speed speech may be workable. In speaking, psycholinguistic/technical measurement options are available (e.g. the use of PRAAT [www.fon.hum.uva.nl/praat/] for the analysis of speech), but for the purposes of measuring how well vocabulary is produced orally, using a panel of judges may well be just as good and far more practical.

Once a suitable technique for measuring fluency has been selected for a particular skill, a fluency test could be added to the more typical vocabulary measures. A number of vocabulary learning activities and exercise types could be explored in a pretest-treatment-posttest design with a battery of tests: minimally recognition/recall tests, better yet measures of vocabulary in use (Point 5), and best of all also including a measure of fluency. This would give a much more comprehensive view of the type and amount of learning that accrues from different task types. I would guess that some activities/exercises are good at promoting some aspects of lexical proficiency, but less useful for other aspects. To give you some ideas, the following studies have looked at the effect of vocabulary activities on the development of written fluency: Snellings, van Gelderen, & de Glopper (2002), Fukkink, Hulstijn, & Simis (2005), Elgort (2011), and Pellicer-Sánchez (2015). Note that there has been relatively little fluency research into auditory recognition or spoken lexical production.

Fluency would probably be most useful for charting the incremental improvement of both receptive and productive levels of mastery, i.e. from the slow, halting ability to comprehend/produce lexical items to quick and automatic ability. It would also allow us to reconceptualize the vocabulary learning timelines in Point 2 to something more nuanced, which might be key to understanding the movement from receptive to productive mastery:
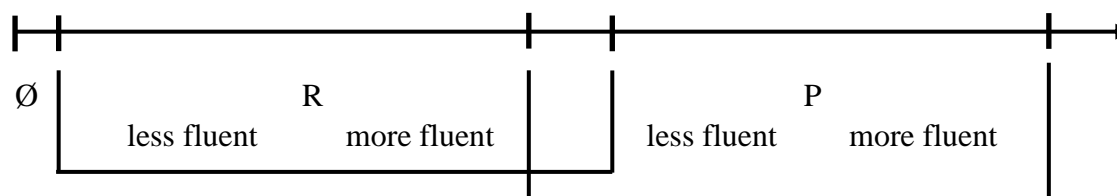
7. Conclusion

I believe the six research directions outlined above could provide real benefits to vocabulary pedagogy and assessment. But I would not want to leave the reader with the impression that these are the only topics worth investigating. Before writing this article, I surveyed 23 vocabulary scholars about what they would like to see researched, and received a varied list of over 36 research topics! (The most repeated (4X) was the need for more longitudinal studies.) Also, I regret not having the space to include the very important issue of formulaic language. I see my six topics as 'bigger issue' priorities, but hope to see vocabulary research advancing on many other fronts as well in the upcoming decade.

Note

1. The List of Rank Knowledge of Vocabulary (LORKOV) will provide a list of the best-known 5,000 lemmas in English, based on test results of thousands of Spanish, German, and Chinese learners of English. It is a British Council project with the following researchers: Norbert Schmitt, Barry O'Sullivan, Laurence Anthony, Karen Dunn, and Benjamin Kremmel.

References

Brown, D. (2010). What aspects of vocabulary knowledge do textbooks give attention to *Language Teaching Research* 15.1, 83–97.

Chen, C. & J. Truscott. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics* 31.5, 693–713.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology* 11.3, 38–63.

Daller, H., J. Milton & J. Treffers-Daller. (2007). *Modelling and assessing vocabulary knowledge.* Cambridge: Cambridge University Press.

De Wilde, V. & J. Eyckmans. (2017). Game on! Young learners' incidental language learning of English prior to instruction. *Studies in Second Language Learning and Teaching* 7.4, 673–694.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning* 61.2, 367–413.

Ellis, N.C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition* 24, 143–188.

Ellis, R. (2009). Editorial. *Language Teaching Research* 13.4, 333–335.

Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal* 31.2, 105–119.

Fukkink, R. G., J. Hulstijn, & A. Simis. (2005). Does training in second-language word recognition skills affect reading comprehension? An experimental study. *Modern Language Journal* 89, 54–75.

Garnier, M. & N. Schmitt. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System* 59, 29–44.

González Fernández, B. & N. Schmitt. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL International Journal of Applied Linguistics* 166, 94–126.

González-Fernández, B. (2018). *A theory of vocabulary acquisition: how do second language learners acquire the different types of word knowledge?* PhD thesis, University of Nottingham.

Grabe, W. (2009). *Reading in a second language*. Cambridge: Cambridge University Press.

Gyllstad, H., L. Vilkaite & N. Schmitt. (2015). Assessing vocabulary size through multiple -choice formats: Issues with guessing and sampling rates. *ITL International Journal of Applied Linguistics* 166, 276–303.

Hatch, E. & C. Brown, C. (1995). *Vocabulary, semantics, and language education*. Cambridge: Cambridge University Press.

Hsu, W. (2009). College English textbooks for general purposes: A corpus-based analysis of lexical coverage. *Electronic Journal of Foreign Language Teaching* 6.1, 42–62.

Hunt, A. & D. Beglar. (January 1998). Current research and practice in teaching vocabulary. *The Language Teacher* 22.1, 7–11. Available online at <http://jalt -publications.org/tlt/articles/1914-current-research-and-practice-teaching -vocabulary>.

Jensen, S. H. (2016). Gaming as an English language learning resource among young children in Denmark. *CALICO Journal* 34.1, 1–19.

Kroll, J. F. & E. Stewart. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33, 149–174.

Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology* 35.1, 65–85.

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press, 140–155.

Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of 'involvement'. *Selected papers from AILA '99 Tokyo*. Tokyo: Waseda University Press, 47–62.

Laufer, B. (2005). Focus on form in second language vocabulary learning. *EUROSLA Yearbook* 5, 223–250.

Laufer, B. & T. Aviad–Levitzky. (2017). What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *Modern Language Journal* 101.4, 729–741.

Laufer, B. & Z. Goldstein. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning* 54.3, 399–436.

Lefever, S. (2010). English skills of young learners in Iceland: "I started talking English when I was 4 years old. It just bang… just fall into me." *Ráðstefnurit Netlu – Menntakvika, Menntavísindasvið Háskóla Íslands*.

Lindgren, E. & C. Muñoz. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism* 10.1, 105–129.

Macis, M. & N. Schmitt. (2017). Not just 'Small Potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research* 21.3, 321–340.

Matsuoka, W. & D. Hirsh. (2010). Vocabulary learning through reading: Does an ELT course book provide good opportunities? *Reading in a Foreign Language* 22.1, 56–70.

Meara, P. (1983). *Vocabulary in a second language, Vol. 1*. Specialised bibliography 3. London: CILT.

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press, 109–121.

Meara, P. & I. Miralpeix. (2017). *Tools for researching vocabulary*. Bristol: Multilingual Matters.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (eds.), *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press, 84–102.

Miller, K.S., E. Lindgren & K. P. H. Sullivan. (2008). The psycholinguistic dimension in second language writing: Opportunities for research and pedagogy using computer keystroke logging. *TESOL Quarterly* 42.3, 433–454.

Miller, M. & V. Hegelheimer. (2006) The SIMs meet ESL. Incorporating authentic computer simulation games into the language classroom. *Interactive Technology & Smart Education* 3.4, 311–328.

Nation, P. (in press). Fill in the details here please.

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

Nattinger, J. R. & J. S. DeCarrico. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Otwinowska, A. & J. M. Szewczyk. (2017). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*. DOI: 10.1080/13670050.2017.1325834.

Pellicer-Sánchez, A. (2015). Developing automaticity and speed of lexical access: the effects of incidental and explicit teaching approaches. *Journal of Spanish Language Teaching* 2.2, 1–15.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition* 38, 97–130.

Peters, E. (2018). The effect of out-of-class exposure to English language media on learners' vocabulary knowledge. *ITL International Journal of Applied Linguistics* 169.1, 141–167.

Porte, G. (2012). Introduction. In G. Porte (ed.), *Replication research in applied linguistics*. Cambridge: Cambridge University Press, 1–18.

Ranalli, J. (2008). Learning English with The Sims: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning* 21.5, 441–455.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Schmitt, N. (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning* 48.2, 281-317.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research* 12.3, 329–363.

Schmitt, N. (2010). *Researching Vocabulary*. Basingstoke: Palgrave.

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64.4, 913–951.

Schmitt, N., P. Nation & Kremmel, B. (in preparation). Minimum validation requirements for vocabulary tests.

Schmitt, N., B. O'Sullivan, L. Anthony, K. Dunn & B. Kremmel. (in preparation). The LORKOV List.

Schmitt, N. & D. Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching* 47.4, 484–503.

Snellings, P., A. van Gelderen & K. de Glopper. (2002). Lexical retrieval: An aspect of fluent Second language production that can be enhanced. *Language Learning* 52.4, 723–54.

Sylvén, L. K. & P. Sundqvist. (2012). Gaming as extramural English L2 learning and L2 proficiency among young learners. *ReCALL* 24.3, 302–321.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition* 27, 33–52.