NORBERT SCHMITT

JS: Could you please tell us a little bit first about your background, and your main areas of interest?

NS: I started my career quite late, as a lot of applied linguists do. When I was 32, I decided I wanted a change in life and I wanted to travel, so I went to Japan without any teaching experience or any real knowledge, and basically learned on my feet, and I ended up really liking it. So I did a Masters at the Osaka branch of Temple University. It was during the bubble economy in Japan so they were able to bring in all the big names. There was a monthly visiting speaker. It was the most fabulous place in the world. At some point when I had finished my Masters, I started becoming interested in vocabulary. I just fell in love with it and the questions of how people learn words and phrases.

JS: You specialize in vocabulary. Do we know how many words learners need to function in English?

NS: Yes, we have a rough idea, but the correct answer to any applied linguistics question is: *It depends*. With productive vocabulary, it's really hard to tell. If you're speaking and you don't know a word, you can paraphrase, you can somehow get around it. So people can have strategies for getting meaning across even if they don't know the best word. People who are good at that will require fewer words and people who are not so good might just get stumped. With receptive vocabulary we have a better idea, because we can analyze the range of vocabulary in reading texts and listening passages. You don't have to know every word. Sometimes we can skip a word and don't lose any meaning, but sometimes we have to guess meaning from context. So the research seems to show that if you know 95-98% of the words, you're normally able to get the meaning. What does that mean in terms of the vocabulary size you need? Our best estimates now are you need 8,000-9,000 word families to be able to read a wide variety of texts. For speaking and listening, speech is less dense than written discourse, so 2,000 or 3,000 word families will get you a long ways in everyday conversation.

JS: The English language is estimated to have between one and two million word forms. How do we even begin to identify the vocabulary that is worth focusing on for learners?

NS: That's a really good question because nobody knows all the words in English. Native speakers don't know them. The best tool we have is frequency. We can look at how often words occur. High frequency words like *table*, *look*, and *love* occur over and over again because we need them a lot.

JS: Many writers and publishers focus on the 2,000 or 3,000 words that are the most frequent according to corpora. Is this the most effective way?

NS: That's the best way we have. It's not perfect. Just

because a word is in the 2,000 or 3,000 word list doesn't mean that everybody uses it all the time, but the probability is that for most people, most of the time, these high-frequency words will be useful, and will probably be useful for anybody learning English because these are about everyday, basic ideas. From about 3,000-5,000 words, you get into common mid-frequency words. They're not going to come up all the time, but they're really useful. Beyond that, maybe from about 5,000-10,000, these are still useful mid-frequency words, like *brandy*, which will only be useful if you're feeling like a nice drink in the evening, but that happens sometimes, doesn't it?

JS: What does it mean to know a word? Does greater vocabulary size relate to greater depth of vocabulary knowledge?

NS: Yes it does. One of the things that it's important to realize when we talk about vocabulary size is that we're actually measuring it in some way. Every measurement we have is in some way a depth measurement. So, are we going to measure size in terms of how many words you can recognize and understand the meaning of? That's one way to measure it. But we could also measure it by asking how many words you can use in a sentence with the right collocations and very appropriately. That would also be size, but it would be many fewer words. So depth and size are absolutely related because however we measure size, there's going to be depth. From my own research what I've found is that for the high frequency words you typically know a lot about them because you see them a lot. For low frequency words maybe all you know is the meaning, but maybe you can't pronounce them right, maybe you don't know the right collocations, maybe you can't use them appropriately. It's much less likely that you're going to have full mastery.

JS: Could you please explain the difference between incidental learning and explicit intentional learning, and what each adds to vocabulary knowledge?

NS: Think about learning a language in your mother tongue. It's almost all incidental. Children in their first language learn almost totally incidentally. At school, children learn the written form for words they already know orally. When it comes to second language learning, we need both incidental and explicit learning. Incidental learning does help, but it is slow, not very sure, and you can forget quite easily. For mother tongue learning, there's so much exposure, so much repetition, that it's always reinforced.

Explicit learning is stronger, more powerful, just because you're focusing on the words. The problem is you can't study every word. The punchline for me is we need both. Explicit instruction focusing on the most useful words is necessary, and then we need to have lots of incidental learning to reinforce and learn more about those words

that we've already studied. It's really difficult for teachers to recycle enough in the classroom, so maximizing exposure outside the classroom is key. We can do this by extensive reading, watching TV, anything that gives more exposure, in different contexts that will enhance or elaborate the meanings of the words that students have started learning in the classroom.

JS: What about extensive reading and graded readers?

NS: Extensive reading is really valuable to give more repetition and seeing the words in different contexts. One thing many teachers and programmers forget about is that it's not just about learning more words. It's also about enhancing words you already know. I'm a vocabulary guy so I look at it from a vocabulary standpoint and I know that extensive reading really helps vocabulary learning, but what's really sweet about it, is that it helps all of language learning. Of course, it helps your reading, it helps you become more familiar with grammar, to understand how English is structured for pragmatic purposes. So extensive reading is probably the best value way of enhancing your learning beyond what you get in the classroom, in a very enjoyable way. Everybody needs learning beyond the classroom and extensive reading is a really valuable way of doing that.

JS: Could you please explain exactly what formulaic language is, and why it is important?

NS: Very simply, formulaic language is vocabulary that has a meaning, but that meaning is attached to more than one word. For example, let's take the idea of dying. How can we express that meaning? We can use a single word **die**, we can have a sort of euphemism **pass away**. That still means die, but there are two words, so that's formulaic language. **Kick the bucket** would be an expression, that's three words, but they're not three separate words really. They look like it, but they're really one word that has one meaning - to die. So formulaic language is multi-word units that have a single meaning. For me, *formulaic language* is the umbrella term for this idea, but there are different categories, like collocations, phrasal verbs, and idioms.

When we do research we find that somewhere between 25-50% of language is actually formulaic language. So we need to get away from the idea that language is individual vocabulary words that are held together with grammar. Some scholars estimate that native speakers know as many formulaic sequences as they do individual words, and I think they are probably right.

JS: What criteria should be used to incorporate formulaic sequences into teaching?

NS: The honest answer is, I don't really know. We know that formulaic language is important, but we're still trying to figure the pedagogy out. We're at the point where we

know that formulaic sequences exist and that they're important and we're starting to describe what they are, but we're not at the point yet that we can make very clear statements or give clear help to the teachers about what to do. At the moment, the best I can do is to say that we need to make students aware that vocabulary is not just individual words and maybe teach prototypical examples. I think the main way of looking at it is meaning-based. If we want to teach students how to say certain things, in many cases those will be formulaic sequences. So maybe instead of teaching collocations as if they were just two words added up, like A+B, if we want to teach the idea that I want coffee with nothing in it, instead of teaching *coffee* and then the adjective *black* to go with it, we just teach *black coffee* as a single chunk to express that idea. So start teaching some multi-word units the same way as individual vocabulary.

JS: Can we rely on incidental acquisition of formulaic language or does it require targeted teaching?

NS: Incidental learning does help and that's one reason why extensive reading, for example, is really important, because you're going to get exposed to these collocations and formulaic sequences in a way that we can't really address in an instructed manner. So probably the best way is to encourage massive exposure. As we begin to understand the learning of formulaic sequences better, we might be able to develop a more explicit, targeted teaching approach, but at the moment we don't really have it.
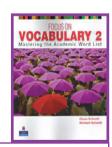
JS: Could you please explain the selection framework based on frequency and transparency suggested by you and Ron Martinez?

NS: I mentioned that we're struggling with actually how to give teachers help. So my former PhD student, Ron Martinez, and I tried to take a really practical approach. We can't teach all formulaic sequences, but which ones might be most useful for students? Frequency is important because formulaic sequences come up all the time. But they're still too many. So how are we going to make the list smaller so that teachers could teach it? And so we came upon the idea of transparency. If it's a collocation, for example like *pick up* – pick up your

pencil – that's pretty clear. Students probably will be able to guess that, so maybe we don't need to teach that meaning sense. But how about *might as well*? What does that mean? You can't really figure it out from the three words. –We might as well get married next week–. This is something you actually have to study. So what we did was we took the highest frequency multi-word units that were not transparent, non-guessable and we built a list of 505 formulaic sequences. Since they had a special definition based around frequency and opaque meaning, we had to find a new name for that so we called them PHRASal Expressions, or PHRASE for short. The PHRASE List is available on my website where you can download it for free. This is the first step toward giving teachers some help in deciding which formulaic sequences are most valuable to teach. It's not the only list. It's not the perfect list, but we feel like it's a good start for teachers who want to include this in their classroom, but really don't have any idea how to proceed.

JS: What do you enjoy most about your job?

NS: Two things I enjoy most. First is I really enjoy research and writing. The second is mentoring my PhD students. I've got world class PhD students who are going to be the next generation of vocabulary researchers. I learn as much from them as they learn from me. It's a real collaboration. I've seen some of my former students going out and becoming established names and are already very influential, so it's a real pleasure to mentor these future stars and help them get a quick start and see them take on their own PhD students. Now I have academic grandchildren, my students' students. That's what I get most pleasure from.

The interviewee

Norbert Schmitt is Professor of Applied Linguistics at the University of Nottingham. He is interested in all aspects of second language vocabulary and has lectured and published widely on this topic, including 8 books, and more than 50 research articles. This includes the best-selling textbook on academic vocabulary entitled *Focus on Vocabulary: Mastering the Academic Wordlist* by Pearson Press, co-authored with his wife, Diane Schmitt. His personal website (*www.norbertschmitt.co.uk*) provides many research and teaching resources, including vocabulary wordlists and tests. It also includes most of his research articles available for free download.