

transact information in a precise and efficient manner (*2-mile final* is a specific location in an aircraft landing pattern) (Schmitt and Carter, 2004). Formulaic sequences also realize a variety of conversational routines, gambits and discourse objectives (Coulmas, 1981) and occur at places of topic-transition and as summaries of gist (Drew and Holt, 1998). Likewise, Nattinger and DeCarrico (1992) argue that formulaic language fulfils the functions of maintaining conversations (*How are you?, see you later*), describing the purposes for which the conversations take place (*I'm sorry to hear about X, would you like to X?*), and realizing the topics necessary in daily conversations (*When is X? (time), How far is X? (location)*). In fact, practically every conventional activity or function in a culture has its associated phrasal vocabulary. (Cell phones now come ready equipped with a lexicon of phrases for users.) Because members of a speech community know these expressions, they serve as a quick and reliable way to achieve the desired communicative effect.

Formulaic sequences also often have a type of register marking function known as *semantic/collocational prosody* (Sinclair, 2004; Stubbs, 2002). For example, *bordered on X (bordered on the pathological, bordered on apathy)* often has a negative evaluation, while the collocations that form around the word *provide (provide information, provide services)* tend to be positive. Semantic prosody is one means of showing a speaker/writer's attitude, evaluation or stance; for example, the knowledge status of the proposition following the formulaic item (*I don't know if X indicates uncertainty about X*), their attitude towards an action or event (*I want you to X shows a positive attitude towards this action*), or their desire to avoid personal attribution (*it is possible to avoid a directly attributable suggestion*) (Biber et al., 2004). Likewise, the choice of formulaic sequences can reflect an author's style and voice, as well as encoding cultural ideas.

In addition to socio-functional motivations for formulaic language, there is also a commonly held psycholinguistic explanation for their existence, which has perhaps been articulated most clearly by Pawley and Syder (1983) and Kuiper and Haggis (1984), and tested by Kuiper (1996): formulaic sequences offer processing efficiency because single memorized units, even if made up of a sequence of words, are processed more quickly and easily than the same sequences of words generated creatively. In effect, the mind uses an abundant resource (long term memory) to store prefabricated chunks of language that can be used 'ready made' in language production. This compensates for a limited resource (working memory), which can potentially be overloaded when generating language online from individual lexical items and syntactic/

9

the acquisition of phrasal vocabulary

koenraad kuiper, georgie columbus and norbert schmitt

introduction

Corpus research has demonstrated that formulaic sequences are ubiquitous in both the written and spoken discourse of native speakers (Cowie, 1998; Biber et al., 1999; Erman and Warren, 2000; McCarthy and Carter, 2002; Kuiper, 2004; Biber et al., 2004). Although idioms have attracted perhaps the greatest research attention, there are many types of formulaic language, varying in degree of fixedness, institutionalism/conventionality and opacity/non-compositionality. This heterogeneity is reflected in the wide range of terminology in the area. Wray (2002, p. 9) found over fifty terms to describe the phenomenon of formulaic language, including *chunks*, *formulaic speech*, *multi-word units*, *collocations*, *formulas*, *prefabricated routines*, *conventionalised forms*, *holophrases* and *ready-made utterances*. The choice of term depends on the focus of the analysis, and many of them will be used in this chapter.

Although much of the research has been done on English, formulaic sequences have been found in a range of languages, including Russian, French, Spanish, Italian, German, Swedish, Polish, Arabic, Hebrew, Turkish, Greek and Chinese (Conklin and Schmitt, 2008). One reason for their pervasiveness may be that formulaic sequences serve a wide range of different functions in discourse. They can be used to express a concept (*get out of Dodge* = get out of town quickly, usually in uncomfortable circumstances), state a commonly believed truth or advice (*Too many cooks spoil the soup/broth* = it is difficult to get a number of people to work well together), provide phatic expressions which facilitate social interaction (*Nice weather today* is a non-intrusive way to open a conversation), signpost discourse organization (*on the other hand* signals an alternative viewpoint), and provide technical phraseology which can

discourse rules. In essence, the mind uses formulaic sequences to ease its processing load. This load can be considerable, especially in speech production, as de Bot (1992) illustrates:

When we consider that the average rate of speech is 150 words per minute, with peak rates of about 300 words per minute, this means that we have about 200 to 400 milliseconds to choose a word when we are speaking. In other words: 2 to 5 times a second we have to make the right choice from...[all the words in the productive lexicon]. And usually we are successful; it is estimated that the probability of making the wrong choice is one in a thousand. (p. 11)

The use of formulaic language seems particularly widespread and essential in spoken discourse that occurs under heavy time constraints, such as auctioneering and sports announcing, suggesting that hearer processing can be facilitated by the predictability of formulaic language (Kuiper, 1996, 2004).

The ubiquity of phrasal vocabulary has led some to suggest there may even be a larger number of phrasal items than single word vocabulary. Sinclair (1991) has gone so far as to argue that the dominant structuring feature of language is the *idiom principle*, rather than the rule-based *open-choice principle*. Pawley and Syder (1983, p. 213) suggest that the number of 'sentence-length expressions familiar to the ordinary, mature English speaker probably amounts, at least, to several hundreds of thousands'. Benson et al.'s (1986) collocational dictionary contains over 90,000 entries, and Moon (1997, p. 48) has suggested that 'the largest specialist dictionaries of English multi-word items...contain some 15,000 phrasal verbs, idioms and fixed phrases, but the total number of multi-word items in current English is clearly much higher'.

formulaic language in first language acquisition

Study of the acquisition of phrasal vocabulary in children was initiated by Peters (1983) (see Wray, 2002; Weinert, 1995), who argued that, for some children, knowledge of grammar and individual words comes to a large degree from the segmentation of phrasal (gestalt) chunks into smaller components. [See Peters, this volume: SFC.] Thus when a child realizes that the phrase *I wanna cookie* (previously used as a holistic unit) is actually *I wanna + noun*, he or she gains information about the way syntax works in the language, as well as the independent new word *cookie*. Children with these 'gestalt' tendencies make more use of formulaic sequences to communicate than children with more referential

tendencies (Nelson, 1973). In the course of acquisition, however, some sequences will cease to be used and are lost, while others are refined into adult-like forms.

Wray (2002) argues that there is an interaction between the fate of early formulae and the rule system: developing formulaic sequences will only be retained if they appear to the learner to conform to the developing rule system. However, knowing whether to keep a formulaic sequence poses an interesting acquisition question. Children are exposed to many formulaic sequences in their input, but how do they decide what to analyse and what to keep at the holistic level? Wray (2002) suggests a 'needs-only analysis' mechanism. Rather than segmenting every sequence into the grammar system, children will operate with the largest possible unit, and only segment sequences when useful for social communication. Thus the segmentation process is driven by pragmatic concerns, rather than an instinctive urge to segment in order to push grammatical and lexical acquisition. The default would be to not analyse, and to retain holistic forms. Thus children maintain many formulaic sequences into adulthood, even though the components of those sequences are likely to be stored individually as well (perhaps being acquired from the segmentation analyses of other formulaic sequences). This suggests that dual storage, as both individual lexical items and as formulaic sequences, is the norm.

Relying on holistic versus analytical approaches to language acquisition and use has long been understood not to be an either/or proposition, and children use both approaches to varying degrees. However, Wray and Perkins (2000) and Wray (2002) suggest that the relative ratios between the approaches may change with age. Schmitt and Carter (2004) summarize Wray and Perkins' views:

During Phase 1 (birth to around 20 months), the child will mainly use memorized vocabulary for communication, largely learned through imitation. Some of this vocabulary will be single words, and some will consist of sequences. At the start of Phase 2 (until about age 8), the child's grammatical awareness begins, and the proportion of analytic language compared to holistic language increases, although with overall language developing quickly in this phase, the amount of holistically-processed language is still increasing in real terms. During Phase 3 (until about age 18), the analytic grammar is fully in place, but formulaic language again becomes more prominent. 'During this phase, language production increasingly becomes a top-down process of formula blending as opposed to a bottom-up

process of combining single lexical items in accordance with the specification of the grammar' (Wray and Perkins, 2000: 21). By Phase 4 (age 18 and above), the balance of holistic to analytic language has developed into adult patterns. (pp. 12–13)

According to Locke and Bogin (2006, p. 265), native adolescent language learning is characterized by the following features: the acquisition of more nuanced grammatical operations, and an increase in the acquisition of idiomatic phrases. Social talking becomes more sophisticated including gossiping in young women and teasing in young men, 'joking, deceiving, mollifying, negotiating and persuading, with increases in the use of sarcasm'. These 'facilitate achievement of two things that matter a great deal to adolescents and adults: status and relationships'. It is clear that all these can involve and help to facilitate increased acquisition of phrasal lexical items (Kuiper, 2006).

formulaic language in a second language

Given the importance of formulaic sequences in native speaker language it is not surprising that they are increasingly seen as an important aspect of language and language use in second language (L2) studies. Some research suggests second language learners handle this part of the language quite well. Schmitt, Dornyei, Adolphs and Durov (2004) found that L2 postgraduate students at the beginning of their study at a British university knew 16.84 out of 20 formulaic sequences (84 per cent) measured receptively with a multiple-choice test. The target items were formulaic sequences found frequently in academic writing, including the type of lexical bundles which Biber et al. (1999) describe (*in the long run, it is clear that, there is no evidence that*). The participants were also able to produce, on average, 12.83 (64 per cent) of the formulaic items in a modified C-test.

In a study using a multiple-choice test of formulaic sequences, Spöttl and McCarthy (2004) found that 14 Austrian multilingual learners were also relatively successful at selecting the correct formulaic sequence for the context. Although only eleven sequences were measured, over 70 per cent of the participants selected correct answers on seven items, and over 90 per cent on two items. This is further evidence that L2 learners have some knowledge of formulaic language. However, other studies have been less impressive. Nesselhauf (2003) extracted 1,072 English verb-noun combinations from 32 essays in the International Corpus of Learner English (ICLE) written by German

university students and found that almost one-quarter of these collocations were judged to be incorrect.

The learners in Spöttl and McCarthy (2004) had relatively good perceptions of their knowledge of the target formulaic sequences, although in some cases they tended to overestimate it. Other studies suggest such self-knowledge is not always as accurate. Phongphio and Schmitt (2006) found that 21 Thai university undergraduates were quite confident of their ability to recognize multi-word verbs when listening or reading, but scored only 55 per cent on a multiple-choice test. Indeed, there was little relationship between the self-rating scores and multiple-choice test percentages. Despite this, when given a context to guess the meaning of the verbs, the learners were able to produce a Thai definition 75 per cent of the time, suggesting they were able to use the context relatively successfully to infer the meanings of many of the unknown multi-word verbs. This is encouraging, as guessing from context is one of the strategies commonly promoted for learning individual words (Nation, 1990, 2001). When specifically asked what strategies they found helpful in discovering the meaning of unknown multi-word verbs (on a 5-point scale), the Thai students reported guessing from textual context (3.62) as one of their top three strategies, along with the use of bilingual dictionaries (3.67) and asking the teacher to translate or to give synonyms (3.52). However, this result must be tempered by the fact that the ratings given hovered around the non-committal middle of the scale.

Despite these modestly good results in written contexts, non-native speakers have been found not to do well when required to produce phrasal items orally. Even when asked to reproduce from dictation (Schmitt, Grandage and Adolphs 2004), only two out of 25 formulaic sequences were reproduced fully correctly by a majority of non-native participants. (Further details of this study are given below.) Similarly, Siyanova and Schmitt (2007) used a questionnaire of 26 multi-word/one-word verb pairs and found that non-native speakers were less likely to use multi-word verbs in informal spoken contexts. High frequency verbs like *make, look* and *do* are used in numerous formulaic sequences and are an important feature of informal spoken discourse. However, Altenberg and Granger (2001) found that EFL learners had great difficulty with formulaic sequences based on the verb *make*, especially the de-lexicalized uses, such as *make a distinction* and *make a decision*. (Biber et al., 1999).

Overall, the evidence is that spoken production of formulaic language is difficult for second language learners and tends to lag behind

other aspects of language development (Irujo, 1993). Some researchers find that learners often simply avoid using these forms (Laufer, 2000). One can assume that formulaic sequences are frustrating for learners given that some, such as multi-word verbs (*work out* at the gym), usually have single word alternatives (*train* at the gym), and therefore appear to be easy candidates for avoidance. However, the multi-word verbs tend to be colloquial in tone and are a particular feature of informal spoken discourse.

Others have found that learners overuse a small number of formulaic sequences (Granger et al., 2006); possibly because they stick with familiar and 'safe' sequences they feel confident using (Granger, 1998). Oppenheim (2000) found that the speech of her six non-native participants contained between 48 per cent and 80 per cent recurrent sequences with an overall mean of 66 per cent. This is further evidence that non-native speakers rely heavily on formulaic language in their efforts to produce fluent speech. Similarly, Bolander (1989) found that her learners of Swedish commonly relied on formulaic sequences in their speech. De Cock (2000), on the other hand, found some formulaic sequences were overused, some underused, and others simply misused by non-native speakers when compared to native norms.

Proficiency also seems to play a role in the use of formulaic sequences. Levy (2003) found that the more proficient of two groups studied in a university context were more likely to use bundles from the academic register, while the less proficient group used more bundles from the conversational register. The more proficient group also used more bundles to structure discourse and for pragmatic purposes; the less proficient group used bundles that were syntactically simpler and literal in meaning. The conclusion must be that while L2 learners are capable of producing mainly acceptable collocations, the underlying intuitions and fluency with collocations of even advanced learners do not seem to match those of native speakers.

The processing advantages of formulaic speech can be exploited by learners to increase fluency. For example, Dechert (1983) studied the spoken output of a German learner of English as she narrated a story from six cartoons. He found that some of her output was marked with hesitations, fillers and corrections, while other output was smooth and fluent. The fluent output was characterized by what Dechert labelled 'islands of reliability', which essentially describe formulaic language. Dechert suggests that islands of reliability may anchor the processes necessary for planning and executing speech in real time. Conklin and Schmitt (2008) found a similar advantage in reading by non-native

speakers who read formulaic sequences more quickly than non-formulaic phrases, even while the non-native speakers read more slowly overall.

The processing advantage of formulaic sequences for native speakers is argued to be that they are stored as ready-made wholes. However, some evidence suggests this may not be the case for non-native speakers. Schmitt, Grandage and Adolphs (2004) embedded recurrent sequences derived from a corpus analysis into a passage that was then used in a dictation task wherein the individual dictation bursts were deliberately long enough to overload working memory. This meant that the dictated language had to be reconstructed rather than being repeated rote from memory. Since the task was to repeat the dictated bursts exactly, it was assumed that the non-native participants would draw upon any of the target formulaic sequences they had stored in memory. Since they would be stored as wholes in memory, it was also assumed that they would be repeated fully intact, without hesitation, and with a normal stress profile. The results showed that many of the recurrent sequences were not repeated in such a manner, or even produced at all by the speakers. This suggests that these recurrent sequences may not in fact be stored as formulaic sequences in the minds of these participants.

the acquisition of phrasal vocabulary in L2

Non-native speakers, particularly adults who are classroom based, are more often expected to learn words rather than phrases, which is unlikely to give them the chance to learn and produce phrasal vocabulary. Even in naturalistic L2 learning contexts, 'there is little evidence... of a progression of the kind identified for first language acquisition, from using formulaic sequences as an aid to initial communication, through a process of segmentation, to nativelike abilities' (Wray, 2002, p. 176). With respect to collocational pairs, Wray suggests non-native speakers acquire separate words which they must then pair for correct collocation. Thus, while formulaic sequences seem to be useful for natives throughout the learning process, non-native speakers begin by focusing more on words than sequences because they are more manageable and give a feeling of control over the language. 'The consequence [of focusing on word-sized units in L2 learning] is a failure to value the one property of native-like input which is most characteristic of the idiomaticity to which the learner ultimately aspires: words do not go together, having first been apart, but, rather, *belong* together, and do not necessarily need separating' (Wray, 2002, p. 212).

While it is a common recommendation for students to memorize formulaic sequences as a basis for more fluent and idiomatic language, Wray (2004) found that this may not always be possible. Her case study subject was required to learn a number of Welsh formulaic sequences in a week in order to present a cooking show on Welsh television. In this case, the most efficient method would be just to learn the phrases and repeat them verbatim without any analysis. However, in interviews five and nine months after the learning, Wray found the subject had introduced many errors typical of an early stage learner of Welsh into the sequences as a result of faulty linguistic analysis. (See also Schmitt et al., 2004, described above.) This suggests that people may not have the capability to bypass linguistic analysis, even when it is in their interest to do so. It seems to be difficult to simply memorize formulaic sequences without them becoming involved in the larger language acquisition process.

When we ask what factors can facilitate the acquisition of formulaic sequences, we find, surprisingly, that the size of the mental lexicon in terms of individual word forms does not seem to be a strong predictor. Schmitt et al. (2004) correlated formulaic sequence gain scores over 2 to 3 months of intensive English for Academic Purposes pre-sessional study with the students' vocabulary size (in word families), and found no significant relationship. Similarly, correlations between the students' word family size and their phrasal item size showed only modest correlations, ranging mainly between .26 and .37, although the correlations between the 3,000 level vocabulary (Schmitt et al., 2001) and productive formulaic sequences ranged between .42 and .54. It seems that having acquired a larger vocabulary consisting of words does not necessarily translate into having acquired a large phrasal lexicon.

The Schmitt et al (2004) study also found that gains in formulaic language did not correlate with language aptitude, language motivation, or language attitudes, even though these factors have been shown to be generally important in language acquisition (Dörnyei, 2002; Dörnyei and Csizér, 2002). However, an in-depth case study has suggested that integration into the L2 environment and culture can make a difference. Dörnyei et al. (2004) interviewed and tested four 'good' formulaic sequence learners and three 'poor' learners over the course of six months. They found that the ability to integrate into the 'host national networks' was a key factor in learning formulaic sequences:

Success in the acquisition of formulaic sequences appears to be the function of the interplay of three main factors: language aptitude,

motivation and sociocultural adaptation. Our study shows that if the latter is absent, only a combination of particularly high levels of the two former learner traits can compensate for this, whereas successful sociocultural adaptation can override below-average initial learner characteristics. Thus, sociocultural adaptation, or acculturation, turned out to be a central modifying factor in the learning of the international students under investigation... (p. 105)

This finding is reminiscent of Wong-Fillmore's study (1976) of the naturalistic acquisition of a second language by elementary school children. She found that the child who had the most integrative orientation to the target culture used the most formulaic language and was creative with its use.

The impact of language aptitude, motivation and socio-cultural adaptation was further explored by Adolphs and Durow (2004), who looked at the interview transcripts of one high-integration ($\approx 9,500$ words) and one low-integration student ($\approx 11,500$ words) from the Dörnyei et al. study. They compared the ten most frequent three-word sequences in the initial interview with the final interview given seven months later to see if the students progressed in using common sequences over time. It transpired that the high-integration student increased the percentage of these three-word formulas from 2.38 per cent to 3.53 per cent, while the low-integration student showed virtually no increase (1.34 per cent to 1.48 per cent). Furthermore, when three-word sequences of all frequencies were analysed (not just the top 10), the high-integration student produced a higher percentage (18.93 per cent to 20.98 per cent of the total text) compared to the low-integration student (9.55 per cent to 12.66 per cent). When the three-word sequences from the interviews were compared to a native spoken corpus (CANCODE), the high-integration student progressed towards using the more native-like forms, while the low-integration student did not. Overall, the analysis suggested that the student with better social integration improved more compared to the student who had difficulty integrating. Some potentially contradictory evidence (Siyanova and Schmitt, 2007) suggesting that the amount of exposure to native-speaking environments did not have an effect on the likelihood of using multi-word verbs could be resolved by acknowledging that it is not exposure *per se* that is important, but the kind of high-quality exposure that occurs in a socially integrated environment.

The availability of Cross Linguistic Influence (CLI) may also play a role in successful or less successful acquisition of formulaic sequences.

Spöttl and McCarthy (2004) found that their multilingual participants were largely able to transfer the meaning of formulaic items across L1, L2, L3 and L4 (see also Siyanova and Schmitt, 2008). Zalasiewicz (1991) carried out an analysis of 205 English idioms translated into Polish. Seven percent of the English idioms translated directly into Polish with full semantic and lexical equivalence, while 17 per cent could not be translated effectively at all. The majority of the idioms (75 per cent) could be effectively translated, because there was usually a semantic correlate available in Polish, but often the lexical form (words/syntax) was quite different. She interpreted the generally high level of similarity between Polish and English idioms in two ways:

Some instances, particularly involving close similarity, could be accounted for by a common origin, the result of shared historical or cultural roots. In most cases, though, the parallels may be the result of a general tendency to construct idioms about the same general topics, with reference to a common set of metaphors. The overall differences were largely the result of reference to specifically English national, cultural or historical phenomena. (p. 1)

One of the few studies to explore whether formulaic sequences can be successfully taught was carried out on university pre-sessional students. Jones and Haywood (2004) highlighted formulaic sequences during a ten week course, and found that they were largely successful in raising their students' awareness of formulaic sequences, but that this awareness did not translate into any substantial increase in the usage of the sequences in the student output. The researchers did note however, that while there was no definite improvement in group performance, 'there were instances where individual students used phrases accurately and appropriately in their own unsupported writing' (p. 289). This suggests that it may not be easy to increase the number of formulaic sequences produced by students. On the other hand, instruction may have more effect on the accuracy and appropriacy of use of formulaic sequences. This is one area which clearly requires more research.

Along these lines, Phongphio and Schmitt (2006) looked at Thai university undergraduates and their impressions of learning strategies and formulaic language. Their 21 participants considered seven strategies which are potentially useful for discovering the meaning of new multi-word verbs. Overall, the students appeared rather lukewarm towards strategies for discovering the meaning of these verbs, with an average rating very near to the non-committal middle of a 5-point rating scale.

Even the top three strategies failed to gain particularly strong ratings: the use of bilingual dictionaries (3.67), guessing from textual context (3.62) and asking the teacher to translate or to give synonyms (3.52). Similarly, the participants were asked to consider strategies potentially useful for consolidating or enhancing knowledge of multi-word verbs which are already partially acquired. Again, the average score was not overwhelmingly positive (3.48), with the exception of 1) connecting multi-word verbs to personal experience (4.19), and 2) using the target multi-word verbs which have just been learnt in real conversation (3.90). Given these relatively modest scores, it seems important to increase learners' awareness of the value of such strategies.

a study of native speaker and L2 acquisition of formulaic vocabulary

It is clear from the foregoing discussion that many questions relating to the acquisition of formulaic vocabulary remain. These include the difficulty of estimating the size of a typical native speaker's phrasal vocabulary and of recognizing what the statistical threshold is after which corpus evidence would establish that an expression is formulaic as opposed to analytic (Altenberg, 1998; Moon, 1998). Currently, taxonomic theories of the linguistic properties of phrasal lexical items predominate (Burger, 2003; Kuiper, 2007) and few established theories exist as to how multi-word lexical items are stored and retrieved from the mental lexicon (Cutting and Bock, 1997; Sprenger et al., 2006; Kuiper et al., 2007). Since theories of language acquisition rely on an understanding of what is acquired and how it is stored and retrieved, these are serious shortcomings.

Cloze testing, however, holds out promise as a method for investigating whether or not a speaker has acquired a formulaic lexical item. Asking a reader to complete an expression (cloze testing) assumes that context and some of the constituents can activate the expression in the mental lexicon, allowing the speaker to supply the missing words (Jackendoff, 1995). Thus, if a respondent can produce the missing word(s) which conventionally fill(s) the gap, then they can be taken to know the expression, i.e. to have acquired it and stored it in their mental lexicon.

An appropriate model of lexical access for phrasal lexical items is superlemma theory (Sprenger et al., 2006; Kuiper et al., 2007). Part of a larger model of lexical access (Levelt, 1989; Levelt and Meyer, 2000), superlemma theory presupposes that a phrasal lexical item has a single

lexical concept. If this is activated, then its superlemma node, in which are found all the idiosyncratic syntactic properties of the item, is activated, including, for example, whether it is flexible under movement, and whether it can take internal modification and so forth. The activation of the superlemma node in turn activates the lemma nodes of all its constituent words. These nodes contain the syntactic properties of individual words on which, for example, the phrase structure of the expression, in part, depends. In turn, the lemma nodes are related to and thus activate their phonological, phonetic and graphological form(s). Under superlemma theory formulaic expressions are unitary, i.e. stored holistically, with a single associated concept in the mental lexicon and in their superlemma representation, rather than being stored as their constituent words.

In responding to a cloze test, a respondent accumulates evidence as to what (s)he is reading incrementally in the normal way through eye fixations at various points along the lines of print. The graphic forms activate their associated lemmas and then the associated concept of each lemma. Parsing proceeds on the basis of the syntactic information contained in the lemma, and the accumulated syntactic properties the parser creates. In the case of a phrasal lexical item, the reader has two options, the lemmas can either add up to a freely produced expression or a lexicalised expression. The activation of a superlemma node, where one is available for the sequence of words being perceived, aggregates the syntactic information gained from the parser until such a point that the activation level is higher than that of the homophonous, non-lexicalized counterpart. Once the superlemma is sufficiently activated, all its constituent lemmas are also activated, thus allowing the respondent to retrieve the missing word. Note this latter process moves from perception to production since the entering of a filler in the gap is a production process.

The situation can be more complex, as indicated earlier, where there is more than one conventional option in the gap. The formulaic expression *enter the fray* has *join* also available as the initial verb. It is an open question as to which of the two is more likely to be selected since both may be activated when the superlemma for *enter/join the fray* is activated. These processes are of course not failsafe, and slips of the pen, misreadings and misunderstandings can occur. And if the respondent does not know the formulaic expression, i.e. if it is not represented in his/her mental lexicon, they must attempt to guess what an appropriate gap filler would be from the context. This will be a best guess on the basis of the compositional evidence available either side of the gap.

Thus if one did not know the expression *worship the ground someone walks on* and the gap was the word *ground*, then any plausible walking surface such as *earth*, *floor* or *carpet* might be selected.

The study to be reported below used cloze testing to explore when formulaic expressions are learned by native speakers and how native and non-native speakers differ in their acquisition. The gapped expressions were all familiar to mature native speakers and were designed to provide subjects with the best linguistic conditions for 'success'. Significant clues were present as to the meaning of the whole expression while at the same time providing contexts for more than one plausible filler. We selected only VPs as test cases and in all cases it was the verb that was gapped in line with Mel'čuk's (1998) proposal that, in the case of lexicalized predicate argument structures, it is often the predicate which is idiomatic while the argument retains its literal meaning. Gapping the same item in each case also created syntactic parallelism and thus clear syntactic priming (Smith and Wheeldon, 2001).

subjects

Respondents for the age grading experiment were selected into four sets of ten: 16 ± 1 years of age, between 20 and 30, between 40 and 50, and over 65. All were New Zealanders. The high-school students were selected from an academic stream while the adult participants were generally university educated. No attempt was made to control for gender or socio-economic status on the (falsifiable) assumption that this would not be relevant.

For the non-native subjects, two groups of EFL speakers were selected. The first group were a set of ten German secondary students aged around 16 who had been studying English at secondary school. These were selected for comparison with the high-school cohort of native speakers. The second group were native speakers of a variety of languages other than English; all were adult EFL-speakers over the age of 20 selected for comparison with the adult native speakers between 20 and 50 from the age grading experiment.

procedure

Respondents were asked to read a story about a social event written in a vernacular style such as one might find in a popular magazine. The aim was to maintain stylistic homogeneity throughout the task and provide sufficient narrative interest to encourage respondents to continue to the end of the story. The context for each expression being investigated was thus extensive. The instructions to participants were as follows: 'Thank

Word lists (Nation, 2000) and the discussion in (Nation and Waring, 1997). The phrasal lexical items themselves could not be ordered by frequency, because, as indicated earlier, reliable measures of frequency are currently unavailable. The four categories were classified as shown in Table 9.2.

Table 9.2 Categorization of verbs

Category	Frequency criterion
High frequency light verbs (HL)	appearing in the top 1–3,000 words in the MFW lists (as words). Note that light verbs are also higher in frequency than the other high frequency verbs.
High frequency lexical verbs (H)	appearing in the top 1–3,000 words in the MFW lists (as words)
Medium frequency lexical verbs (M)	appearing in the 3,000–5,000 word list in the MFW lists (as words)
Low frequency lexical verbs (L)	not appearing in any lists

you for taking the time to help us with our research. Please read the story and when you find a gap (—), write in the verb you think should go there. For example: Wow! It's really raining cats and dogs out there! Don't worry if you can't think of a word straight away, just put in your best guess.' Appendix A provides the first part of this story.

The expressions being targeted were given in bold type. Nothing was said as to why they were in bold but the aim was to provide a visual clue that the gap was related to the bolded sequence of words which was usually the complement of the verb. The test items are shown in Table 9.1, ordered by the frequency of their head verb.

The cloze items were selected on the basis of their membership in four categories: light (or de-lexicalized) verbs (Grimshaw, 1990), non-light high frequency verbs, mid-frequency verbs and low frequency verbs. The frequencies were derived as follows. A random selection of phrasal lexical items were selected from the *Syntactically Annotated Dictionary of Idioms* (Kuiper et al., 2003), and the verbs were divided into three frequency bands and four categories based on a combined evaluation of verb frequency rankings in the Brown Corpus (accessed via edict.com.hk), Kilgariff's BNC rankings (Kilgariff, 1995), the Most Frequent

Table 9.1 Test items by descending order of frequency of the head verb

1. do things by halves
2. make tracks (for)
3. take a fancy to NP
4. give NP the creeps
5. keep a straight face
6. let NP into a secret
7. join/enter the fray
8. drive NP to drink
9. act the goat
10. avoid NP like the plague
11. wipe NP off the map
12. tighten NP's belt
13. seal NP's fate
14. spare no expense
15. scrape the bottom of the barrel
16. worship the ground NP walks on
17. wring NP's neck
18. pluck/summon up courage
19. goad/spur NP into action
20. toe the company line

Note: NPs are noun phrases which were filled in the story. The verbs were gapped.

Table 9.3 Frequency data and band allocation of head verbs

Verb	Rank number	Total occurrences	Frequency band
do	18	559,596	HL
make	46	217,268	HL
take	54	179,220	HL
give	76	131,417	HL
keep	189	50,092	HL
let	330	29,768	H
join	594	17,331	H
drive	618	16,477	H
act	654	15,620	H
avoid	866	11,750	H
wipe	3,122	2,367	M
tighten	4,178	1,548	M
seal	4,249	1,512	M
spare	5,457	1,023	M
scrape	6,011	865	M
worship	no rank	0 (as verb)	L
wring	no rank	0	L
pluck	no rank	0	L
goad	no rank	0	L
toe	no rank	0 (as verb)	L

Table 9.4 BNC frequencies of alternate verbs

Verb	BNC freq. per million words
<i>join</i>	173.91
<i>enter</i>	142.69
<i>summon</i>	14.02
<i>pluck</i>	6.29
<i>spur</i>	3.97
<i>goad</i>	1.55

Twenty items were then selected using these criteria to create the short story, five from each category. These 20 verbs were subsequently checked against Kilgariff's lemmatized BNC frequency list, and the results showed that the original allocations were largely accurate. Items ranked highly were largely in the 'High' category, and with a high total number of occurrences. The frequency bands were also checked against frequency data from the CELEX database (Baayen et al., 1993) and the BNC corpus data (www.natcorp.ox.ac.uk/). With minor ranking exceptions the allocation of verbs to frequency bands was confirmed. (For example, in the BNC data *worship* ranked one higher than *scrape* while *seal* ranked one higher than *tighten*.)

The initially employed frequency data and band allocation of each head verb are contained in Table 9.3.

As can be seen in Table 9.1, some cloze items allowed for more than one conventional option. There were three selected for investigation here: *enter/join the fray*, *goad/spur NP into action* and *pluck/summon up courage*. The alternates appeared in the same frequency bands, i.e. they were relatively close in their frequency as is shown in their BNC frequency (Table 9.4).

results

Differences among the four native speaker groups as a function of age were not large as shown in Figure 9.1.

Perhaps this result is not surprising considering that the items chosen were vernacular, non-specialized items, likely to be known to individuals from teenage onwards. However, given that enculturation is lifelong, the acquisition of often low frequency phrasal vocabulary items could be expected to continue, though presumably at a decreasing rate, throughout the life course. One would also expect that lexical retrieval becomes slightly less efficient after 65.

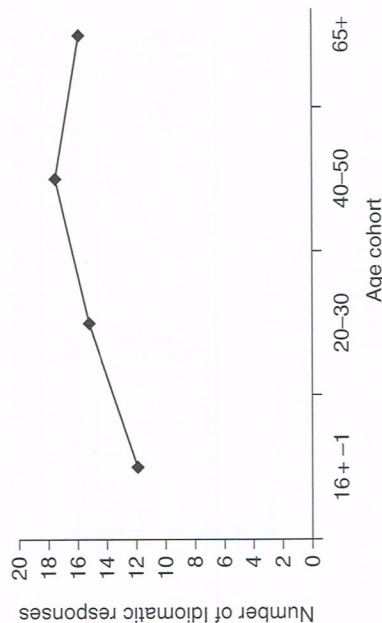


Figure 9.1 Native speaker recall of verb heads of phrasal lexical items by age group

The cloze tests of the native and EFL adolescents and adults yielded the comparisons in Table 9.5. These data suggest that EFL speakers' acquisition of vernacular phrasal vocabulary items is an order of magnitude lower than that of native speakers. Note, however, that given the very large phrasal vocabulary of native speakers, non-native speakers even in adolescence have still acquired a sizable phrasal vocabulary. Note too that the cloze procedure is likely to underestimate non-native speaker recognition of phrasal vocabulary; a multi-choice methodology would be anticipated to yield a significantly larger number of recognized phrases.

The impact of the frequency of the gapped head verb was consistent across all the age-graded native speaker respondents with more successful retrieval of light and higher frequency verbs as compared to lower frequency verbs (Figure 9.2).

It was to be expected that retrieval of mid- and low-frequency heads would show the attrition in the older age cohort it appears to. Non-native speakers show similar preference for clozing on high frequency verb heads as is shown in Figure 9.3.

The cloze testing process has, therefore, corroborated previous hypotheses that acquisition of phrasal lexical items is age graded and that non-native speakers have a lower rate of acquisition. It has also corroborated the hypothesis that the frequency of a head word which is gapped affects its recall, the more frequent items being easier to recall than low frequency items. There were, however, some exceptions. Some lexicalized VPs seem to have high saliency, notwithstanding the low frequency of their head verb. A clear case is *worship the ground NP walks*

on which was correctly clozed by 10/10 of the over-65-year-old native speakers, 9/10 of the 40–50-year-olds, 9/10 of the 20–30-year-olds and 8/10 of the 16-year-olds.

The verbs which were scored as being correctly recalled were those which are the conventional heads of the phrasal lexical item. In three cases there was more than one of these. Cloze testing of native speakers revealed the selectional patterns favouring particular heads of phrase, as Table 9.6 shows.

These selectional preferences are not exclusively related to the relative frequency of the respective verbs in text, however, as shown in Table 9.7.

We can therefore infer that phrasal lexical items with alternative selectional preferences may have those alternatives independent of the frequency of their head, as shown by *enter vs. join* and *pluck vs. summon*. Such preferences may be dialectally determined. However, cloze testing, by forcing a single response, cannot show acquisition of the alternatives. On the other hand from this small data set it would seem that speakers differ in their preference for particular words from a selection

Table 9.6 Selection of alternative idiomatic heads

	Word 1	Word 2	Total
Times selected	enter the fray 3 goad NP into action	join the fray 27 spur NP into action	30/40
Times selected	5 pluck up the courage to	10 summon up the courage to	15/40
Times selected	9	4	15/40

Table 9.7 BNC frequencies of alternative idiomatic heads

	Word 1	Word 2
BNC fr	enter 142.69	join 173.91
BNC fr	goad 1.55	spur 3.97
BNC fr	pluck 6.29	summon 14.02

Table 9.5 Native and EFL adolescent and adult cloze test results.

Native adolescents	EFL adolescents
11.9	0.8
Native adults	EFL adults
16.4	1.8

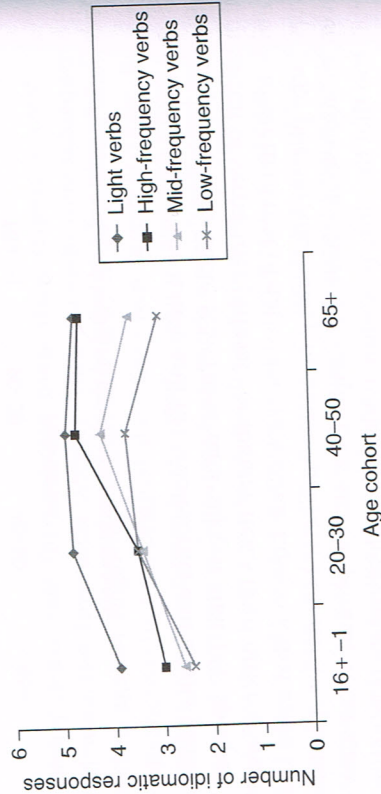


Figure 9.2 Native speaker recall of heads of phrasal lexical items (Frequency plotted against age)

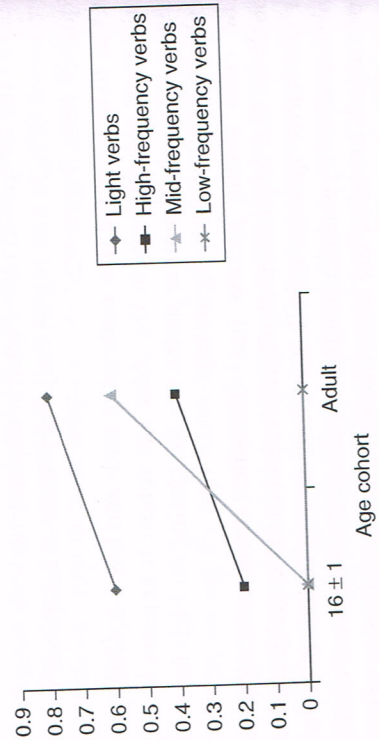


Figure 9.3 Non-native speaker recall of heads of phrasal lexical items (Frequency plotted against age)

set, and that populations of speakers can, through cloze testing, reveal normative preferences. It is an interesting question whether such preferences are also found in the frequencies with which the alternatives are manifest in large corpora.

conclusion

In addition to revealing learners' knowledge of phrasal vocabulary at a particular point in time, as here, cloze testing could be used iteratively to explore acquisition over a period. One might use it, for example, to see if early phrasal vocabulary learning undergoes a learning burst at some stage in language development and if acquisition order is related to the acquisition of the head which would imply that verb phrases with low frequency heads would be acquired in the order of the frequency of their heads. These and other questions will have to await further research.

appendix a

Shannon walked into the vast, badly decorated function room and looked around for faces she knew. Tom's hand waved frantically at her from a table near the coat check. *Thank goodness!* she thought to herself, *I'm not the only one from Data Entry here!* Shannon normally these kind of events like the plague, but her pushy new flatmate had convinced her a night out might be in order, especially considering she'd only been at the firm for a few weeks. She walked towards Tom and was further relieved to see familiar faces from Accounts, which had its offices on the same floor as her department. On closer inspection, she realised she knew a few others at the table – Jenny, who had a tendency to the goat at inter-departmental health and safety meetings, pulling faces and telling stupid jokes; Annabel, who always looks like a startled deer when you ask her anything that isn't work-related, her face going blotchy at the prospect of real conversation; and Jonno, who every woman on the floor the creeps with his fake smile and lame innuendo.

references

Adolphs, S. and Durow, V. (2004). Social-cultural integration and the development of formulaic sequences. In Schmitt, N. (ed.), *Formulaic Sequences* (pp. 107–26). Amsterdam: John Benjamins.

- Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word-combinations. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis and Applications* (pp. 101–22). Oxford: Oxford University Press.
- Altenberg, B. and Granger, S. (2001). The grammatical and lexical patterning of *make* in native and non-native student writing. *Applied Linguistics* 22, 2, 173–94.
- Baayen, R. H., Piepenbrock, R. and van Rijn, H. (1993). The CELEX Lexical Database.
- Benson, M., Benson, E. and Ilson, R. (1986). *The BBI Combinatory Dictionary of English*. Amsterdam: John Benjamins.
- Biber, D., Conrad, S. and Cortes, V. (2004). Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25, 377–405.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hylltenstam and L. Obler (eds), *Bilingualism across the Lifespan* (pp. 73–86). Cambridge: Cambridge University Press.
- Burger, H. (2003). *Phraseologie: Eine Einführung am Beispiel des Deutschen* (2nd edition). Berlin: Erich Schmidt Verlag.
- Conklin, K. and Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, March, 29: 72–89.
- Coulmas, F. (1981). *Conversational Routine*. The Hague: Mouton.
- Cowie, A. (ed.) (1998). *Phraseology: Theory, Analysis and Applications* (pp. 125–43). Oxford: Oxford University Press.
- Cutting, J. and Bock, K. (1997). That's the way the cookie bounces: syntactic and semantic components of experimentally controlled idiom blends. *Memory and Cognition* 25: 1, 57–71.
- de Bot, K. (1992). A bilingual production model: Levelt's 'speaking' model adapted. *Applied Linguistics*, 13, 1–25.
- De Cock, S. (2000). Repetitive phrasal chunkiness and advanced EFL speech and writing. In C. Mair and M. Hundt, M. (eds), *Corpus Linguistics and Linguistic Theory* (pp. 51–68). Amsterdam: Rodopi.
- Dechert, H. (1983). How a story is done in a second language. In C. Faerch and G. Kasper (eds), *Strategies in Interlanguage Communication* (pp. 175–95). London: Longman.
- Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (ed.) *Individual Differences and Instructed Language Learning* (pp. 137–58). Amsterdam: John Benjamins.
- Dörnyei, Z. and Csizér, K. (2002). Some dynamics of language attitudes and motivation: results of a longitudinal nationwide survey. *Applied Linguistics*, 23, 421–62.
- Dörnyei, Z., Durow, V. and Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 87–106). Amsterdam: John Benjamins.
- Drew, P. and Holt, E. (1998). Figures of speech: figurative expressions and the management of topic transition in conversation. *Language in Society*, 27, 495–522.

- Erman, B. and Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 29–62.
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (ed.), *Phraseology: Theory, Analysis and Application* (pp. 145–60). Oxford: Oxford University Press.
- Granger, S., Paquot M. and Rayson P. (2006) Extraction of multi-word units from EFL and native English corpora. The phraseology of the verb 'make'. In A. Häcki Buhofer and H. Burger (eds), *Phraseology in Motion I: Methoden und Kritik. Akten der Internationalen Tagung zur Phraseologie* (Basel, 2004) (pp. 57–68). Baltmannsweiler: Schneider Verlag Hohengehren.
- Grimshaw, J. (1990). *Argument Structure*. Cambridge, MA: MIT Press.
- Irujo, S. (1993). Steering clear: avoidance in the production of idioms. *International Review of Applied Linguistics in Language Teaching*, 31, 205–19.
- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk and R. Schreuder (eds), *Idioms: Structural and Psychological Perspectives* (pp. 133–66). Hillsdale, NJ: Erlbaum.
- Jones, M. and Haywood, S. (2004). Facilitating the acquisition of formulaic sequences. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 269–300). Amsterdam: John Benjamins.
- Kilgarriff, A. *BNC Database and Word Frequency Lists* 1995 [cited 24 August 2006]. Available from www.kilgarriff.co.uk/BNC_lists/lemma.al.
- Kuiper, K. (1996). *Smooth Talkers*. Mahwah, NJ: Erlbaum.
- Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 37–54). Amsterdam: John Benjamins.
- Kuiper, K. (2006). Knowledge of language and phrasal vocabulary acquisition. *Behavioral and Brain Sciences*, 29, 291–2.
- Kuiper, K. (2007). Cathy Wilcox meets the phrasal lexicon. In J. Munat (ed.), *Lexical Creativity, Texts and Contexts* (pp. 93–112). Amsterdam: Benjamins.
- Kuiper, K. and Haggio, D. C. (1984). Livestock auctions, oral poetry and ordinary language. *Language in Society*, 13, 205–34.
- Kuiper, K., McCann, H., Aitchison, T. and van der Veer, K. (2003). *S.A.I.D.: A Syntactically Annotated Idiom Database*. Linguistics Data Consortium, University of Pennsylvania.
- Kuiper, K., van Egmond, M.-E., Kempen, G. M. and Sprenger, S. (2007). Slipping on superlemmas: multiword lexical items in speech production. *The Mental Lexicon*, 2: 3, 313–57.
- Laufer, B. (2000). Avoidance of idioms in a second language: the effect of L1-L2 degree of similarity. *Studia Linguistica*, 54, 186–96.
- Levelt, W. J. M. (1989). *Speaking: from Intention to Articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. and Meyer, A. S. (2000). Word for word: multiple lexical access in speech production. *European Journal of Cognitive Psychology*, 12: 4, 433–52.
- Levy, S. (2003). *Lexical Bundles in Professional and Student Writing*. Unpublished Ph.D. dissertation, University of the Pacific, Stockton, CA.
- Locke, J. L. and Bogin, B. (2006). Language and life history: a new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, 29, 259–325.

- McCarthy, M. and Carter, R. (2002). This that and the other: multi-word clusters in spoken English as visible patterns of interaction. *Taunga* (Yearbook of the Irish Association for Applied Linguistics), 21, 30–52.
- Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*. Oxford: Clarendon Press.
- Moon, R. (1997). *Vocabulary connections: multi-word items in English*. In N. Schmitt and M. McCarthy (eds), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 40–63). Cambridge: Cambridge University Press.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A corpus based approach*. Oxford: Clarendon Press.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. (2000) *Most Frequent Words* [cited 8 August 2006. Available on www.edict.com.hk/lexiconindex/frequencylists/words2000.html].
- Nation, P. and Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt and M. McCarthy (eds), *Vocabulary: Description, acquisition, pedagogy*. New York: Cambridge University Press.
- Nattinger, J. R. and DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nelson, K. (1973). *Structure and Strategy in Learning to Talk*. Monographs of the Society for Research in Child Development, Serial no. 149, nos 1–2.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 2, 223–42.
- Oppenheim, N. (2000). The importance of recurrent sequences for nonnative speaker fluency and cognition. In H. Riggensbach (ed.), *Perspectives on Fluency* (pp. 220–40). Ann Arbor: University of Michigan Press.
- Pawley, A. and Syder, F. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. Richards and R. Schmidt (eds), *Language and Communication* (pp. 191–225). London: Longman.
- Peters, A. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Phongphio, T. and Schmitt, N. (2006). Learning English multi-word verbs in Thailand. *Thai TESOL Bulletin*, 19, 2, 122–36.
- Schmitt, N. and Carter, R. (2004). Formulaic sequences in action: an introduction. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 1–22). Amsterdam: John Benjamins.
- Schmitt, N., Dornyei, Z., Adolphs, S. and Durow, V. (2004). Knowledge and acquisition of formulaic sequences. In N. Schmitt (ed.), *Formulaic Sequences*. Amsterdam: John Benjamins.
- Schmitt, N., Grondage, S. and Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (ed.), *Formulaic Sequences* (pp. 127–51). Amsterdam: John Benjamins.
- Schmitt, N., Schmitt, D. and Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test, *Language Testing*, 18.1, 55–88.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (2004). *Trust The Text: Lexis, Corpus, Discourse*. London: Routledge.
- Siyanova, A. and Schmitt, N. (2007). Native and nonnative use of multi-word vs one-word verbs. *International Review of Applied Linguistics*, 45, 109–39.
- Siyanova, A. and Schmitt, N. (2008). L2 learner production and processing of collocation: a multi-study perspective. *Canadian Modern Language Review*, 64, 3, 429–58.
- Smith, M. and Wheeldon, L. (2001). Syntactic priming in spoken sentence production – an online study. *Cognition*, 78, 123–64.
- Spöttl, C. and McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (ed.), *Formulaic Sequences*. Amsterdam: John Benjamins. pp. 190–225.
- Sprenger, S., Levelt, W. J. M. and Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–84.
- Stubbs, M. (2002). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: a review. *Applied Linguistics*, 16, 180–205.
- Wong-Fillmore, L. (1976). *The Second Time Around*. Doctoral dissertation, Stanford University.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2004). 'Here's one I prepared earlier': Formulaic language learning on television. In N. Schmitt (ed.), *Formulaic Sequences* (pp. 249–68). Amsterdam: John Benjamins.
- Wray, A. and Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language and Communication*, 20, 1–28.
- Zalasiewicz, K. (1991). English and Polish Idioms: A Comparative Study. Unpublished MA dissertation: University of Nottingham.