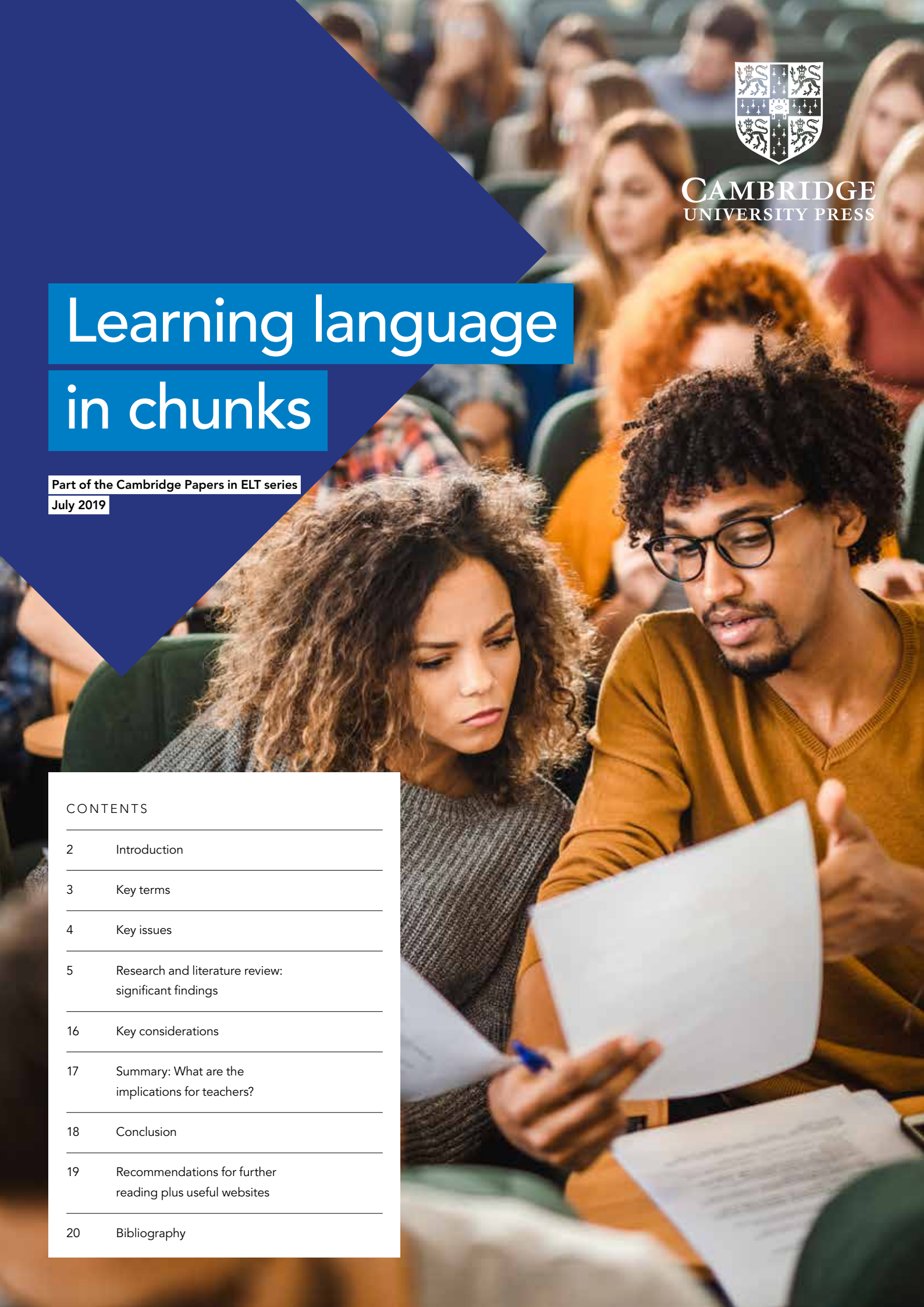


CAMBRIDGE
UNIVERSITY PRESS

Learning language in chunks

Part of the Cambridge Papers in ELT series
July 2019

CONTENTS

Introduction

It was over 25 years ago that Michael Lewis published *The Lexical Approach* (Lewis, 1993), prompting a radical re-think of the way that we view language – and, by extension, of the way that we teach it.

In contrast to the then prevailing structural account, in which language was viewed as comprising grammatical structures into which single words are slotted, Lewis argued that 'language consists of chunks which, when combined, produce continuous coherent text' (Lewis, 1997: 7).

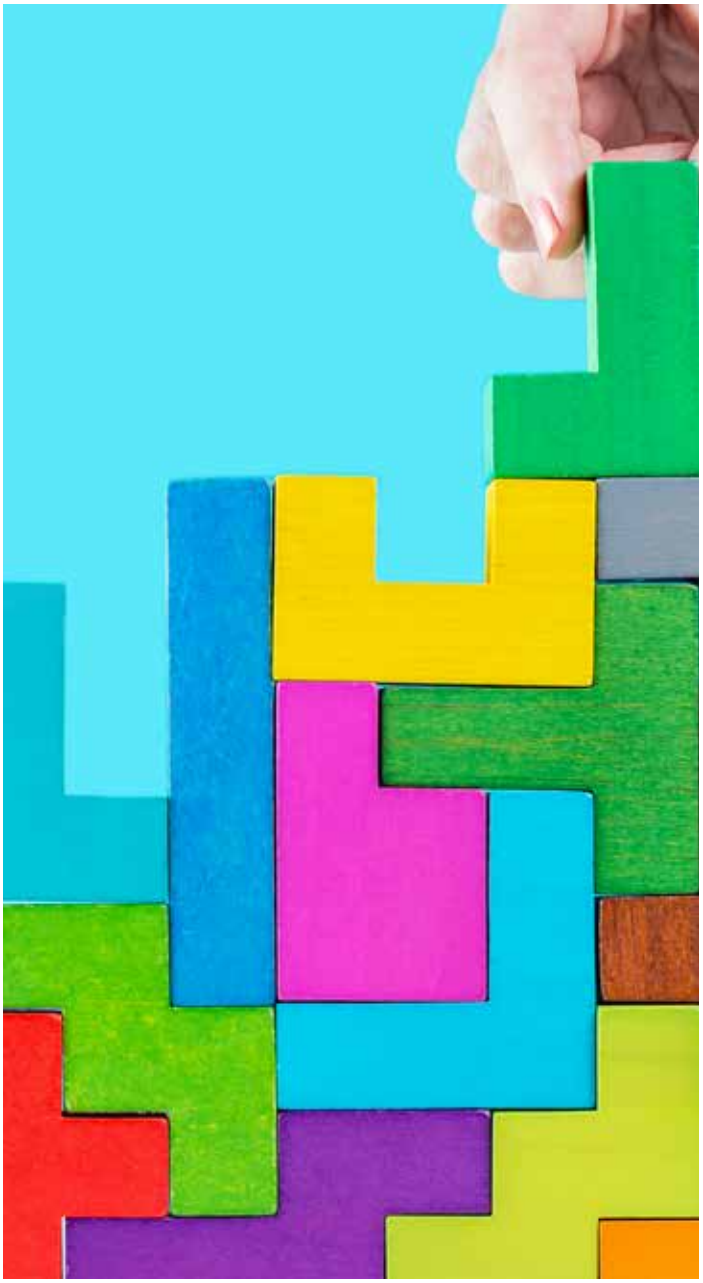
By 'chunks', Lewis was referring to everything from:

- collocations (*wrong way, give way, the way forward*)
- fixed expressions (*by the way, in the way*)
- formulaic utterances (*I'm on my way; no way!*)
- sentence starters (*I like the way…*)
- verb patterns (*to make/fight/elbow one's way…*)
- idioms and catchphrases (*the third way; way to go!*)

… everything, in fact, that doesn't fit neatly into the categories of either grammar (as traditionally conceived) or single-word vocabulary.

Lewis was by no means the first to describe language in these terms: his singular contribution was to argue that, in order to accommodate this alternative description, it was language teaching that needed to be reformed – or, indeed, revolutionised.

This paper charts the extent to which the Lexical Approach, or 'learning language as chunks', as Lewis and subsequent scholars conceived it, is being applied a quarter of a century on, and the research that underpins such an approach.



Key terms

chunk:

an all-purpose word that embraces any formulaic sequence, lexical/phrasal expression or multi-word item.

cluster (or bundle):

any commonly occurring sequence of words, irrespective of meaning or structural completeness, e.g. *at the end of the, you know what.*

collocation:

two or more words that frequently occur together, e.g. *false eyelashes, densely populated, file a tax return.*

corpus:

a database of texts, typically authentic, which is digitally accessible for the purposes of calculating frequency, identifying collocations, etc.

formulaic language:

'a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar' (Wray, 2000: 465).

functional expression:

formulaic ways of expressing specific language functions, e.g. *Would you like...?* (for inviting).

idiom:

an expression whose meaning is not the sum of its individual words, i.e. it is 'non-compositional', e.g. *a wild goose chase, run out of steam, plain sailing.*

idiom principle:

the principle of language use whereby 'a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments' (Sinclair, 1991: 110). This contrasts with the 'open choice principle', where the only restraint on word choice is 'grammaticalness'.

idiomaticity:

the degree to which a particular wording is conventionalised in the speech community. For example, the time 3.20 is said as '*twenty past three*' or '*three twenty*', but not '*three and a third*' (as its Egyptian Arabic equivalent would be translated, for example).

lexical approach:

an approach to language teaching that foregrounds the contribution of vocabulary, including lexical chunks, to language use and acquisition.

lexical phrase:

one of many alternative terms to describe multi-word items.

MI (Mutual Information):

a statistical measure of the strength of a collocation based on the likelihood of its individual elements occurring together more frequently than would be expected by chance: *short + straw*, for example, has a higher MI score than *long + straw*, while *draw + the short straw* has a higher MI score than *choose + the short straw.*

n-gram:

a cluster defined in terms of its length, e.g. common 3-word n-grams are *I don't know, a lot of, I mean I …* (O'Keeffe et al., 2007: 66).

phrasal verb:

a combination of a verb plus a particle (either adverb or preposition) that is often idiomatic: e.g. *she takes after her father; the plane took off.*

phraseology:

a general term to describe the recurring features of language that are neither individual words nor grammatical structures.

Key issues

Lewis's Lexical Approach was strongly, even fiercely, argued, but was only sketchily supported by evidence. In the intervening years, researchers – directly or indirectly – have been investigating his claims, with a view to answering these key questions (among many others):

1. **To what extent does language consist of chunks?**
2. **How might the learning of chunks benefit language learning overall?**
3. **How can chunks be integrated into the second language curriculum?**
4. **How are chunks best learned and taught? What materials and activities might support their acquisition?**

This paper addresses each of these questions in turn.

Research and literature review: significant findings

1. To what extent does language consist of chunks?

In order to estimate the proportion of spoken or written text that is 'chunk-like', we first need to define 'chunk'. This is easier said than done: the number of terms that are used to capture the phenomenon is bewildering.

One writer (Wray, 2002) listed over 50, but current practice seems to favour, as an (uncountable) umbrella term, *formulaic language*, embracing different types of *multi-word units* (MWUs), or what most non-academic texts for teachers refer to simply as *(lexical) chunks*. (Krishnamurthy (2002: 289) prefers the term 'chunk' since, being relatively recent, it has 'less baggage associated with it'.) These, in turn, can be subdivided into such overlapping categories as *collocations, lexical phrases, phrasal verbs, functional expressions, idioms,* and so on (see the **Key terms** on page 3 for definitions).

What the items in these categories have in common is that:

- they consist of more than one word
- they are conventionalised
- they exhibit varying degrees of fixedness
- they exhibit varying degrees of idiomaticity
- they are probably learned and processed as single items (or 'holophrases').

Word combinations are *conventionalised* if they occur together with more than chance frequency. Corpus linguistics has exponentially enhanced our knowledge of what combinations of words are significantly frequent. Thus, the sequence *no way of [+ -ing]* can be completed with virtually any verb, but only one – *knowing* – is significantly frequent. (It is more than ten times more common than the next most frequent combination: *no way of telling*.) Moreover, according to *The Corpus of Contemporary American English* (Davies, 2008) it is relatively common across a variety of registers: spoken language, fiction, news and academic writing.

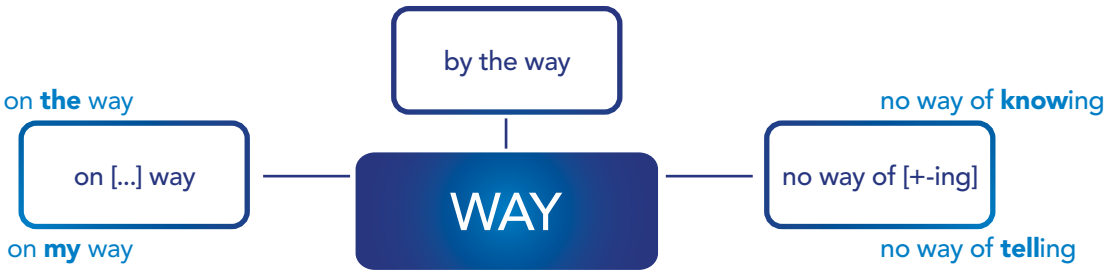


Diagram showing some examples of the fixedness of key phrases used with ***way***.

With regard to *fixedness*, an example of a fixed chunk is *by the way*, which, as a discourse marker, allows no variation: **by a way*, **by the ways*. *On the way*, however, allows some variation, e.g. *on my way*. *By and large* is an example of a chunk that is not only fixed but also idiomatic, i.e. it is 'non-compositional': its composite meaning cannot be inferred from its individual words. *By the way*, on the other hand, is less idiomatic since – even in its sense of marking a new direction in the discourse – its meaning is relatively easily derived from its parts.

In terms of their psycholinguistic status, that is, the way that they are mentally stored and accessed, there is growing evidence, e.g. from eye-tracking and read-aloud studies (see, for example, Ellis *et al*., 2008), that chunks are processed holistically, rather than as a sequence of individual words. This is attributed to frequency effects: the more often a sequence (of morphemes or words) is encountered, the more likely it is that it is represented and retrieved as a single unit (Siyanova-Chanturia & Martinez, 2014). However, it would be unwise to assume that what corpus data reveal about recurring sequences necessarily reflects the way that these sequences are mentally organised. Using dictation and delayed recall tasks, Schmitt *et al* (2004: 147) found that neither their native nor their non-native speaker informants consistently retrieved chunks as whole units, leading them to conclude that 'it is unwise to take recurrence of clusters in a corpus as evidence that those clusters are also stored as formulaic sequences in the mind'.

Nevertheless, and regardless of how chunks are defined, their pervasiveness is a fact of life: one frequently cited estimate is that nearly 60% of spoken language (slightly less of written) is formulaic to some degree. Biber et al (1999: 990), using slightly different criteria, found that 45% of the words in their extensive corpus of conversational English (but only 21% in academic prose) occurred in what they called 'bundles', i.e. 'recurrent expressions, regardless of their idiomaticity, and regardless of their structural status'. Some high frequency four-word bundles in spoken English include: *I don't know what, I don't think so, I was going to, do you want to*. On the other hand, they found many fewer instances of idiomatic phrases, of the type: *in a nutshell, a piece of cake, fall in love* or a *slap in the face*. These tend to occur more often in fiction than in spoken language. Idiomatic expressions that do occur with high frequency in both spoken and written language are verb + noun phrase combinations with *have, make* and *take* (as in *have a look, make sense, take time*) and phrasal verbs, especially (in spoken language) those with *come, go* and *put*.

A number of studies have attempted to compute the frequency of chunks compared to that of single words, and have established that there are many chunks that are as frequent as, or more frequent than, the most frequent individual words. In one study (Shin & Nation, 2008), using the 10 million word spoken section of the British National Corpus (BNC), the researchers identified the most frequently occurring collocations, and found that 84 collocations qualified for inclusion in the top 1,000 word band (examples being: *you know, I think, a bit, as well, in fact…*) – with increasing numbers of collocations eligible for inclusion in successive bands.

A number of studies have attempted to compute the frequency of chunks compared to that of single words, and have established that there are many chunks that are as frequent as, or more frequent than, the most frequent individual words.

In a similar fashion, Martinez & Schmitt (2012), using slightly different criteria, identified over 500 'phrasal expressions' that would qualify for inclusion in a list of the top 5,000 word families in the BNC, both written and spoken. Examples include: *after all, as soon as, make sure, once again*.[1]

Other researchers have investigated not just the frequency but also the distribution of lexical chunks (specifically n-grams) in different registers of both spoken and written text. Biber *et al* (2004) for example, conclude that their patterns of use are not accidental, but that they form the 'building blocks' of discourse and often correlate with specific textual functions, such as expressing the speaker/writer's attitude, or foregrounding new information. As such they represent an important

1 The complete list is available by clicking on PHRASE list at: https://www.norbertschmitt.co.uk/vocabulary-resources

indicator of a text's register as well as a measure of the speaker or writer's command of that register.

2. How might the learning of chunks benefit language learning overall?

There are at least three reasons that have been put forward for prioritising the learning of lexical chunks, each of which we shall look at in turn:

- they facilitate fluent processing
- they confer idiomaticity
- they provide the raw material for subsequent language development.

Fluency

As long ago as 1925, Harold Palmer anticipated the Lexical Approach by posing the question, 'What is... the most fundamental guiding principle [to conversational proficiency]?' He answered this question with: *'Memorize perfectly the largest number of common and useful word-groups.'* (Palmer, 1925: 187, emphasis in original). It took another seven decades for Pawley & Syder (1983: 214), in a seminal paper that sought to solve two of the 'puzzles' of native-like proficiency, to reach a similar conclusion: 'Lexicalised sentence stems, and other memorized strings, form the main building blocks of fluent connected speech'. In other words, the possession of a memorized store of 'chunks' allows more rapid processing, not only for production but also for reception, since 'it is easier to look up something from long-term memory than compute it' (Ellis *et al.*, 2008: 376). Or, as another scholar neatly put it, 'speakers do as much remembering as they do putting together' (Bolinger, 1976: 2).

Subsequent research supports these intuitions. For example, Towell *et al* (1996) compared the spoken fluency of advanced speakers of French before and after an extended stay in France. They found that the more fluent speakers were able to speak faster and with less hesitation due to the effective use of chunks.

Boers *et al* (2006) conducted a study where two groups of learners were taught the same curriculum, but only one group was given 'lexical-phrase oriented pedagogy' as well. Both groups were assessed on a speaking task: the experimental group was generally judged more fluent *and* their fluency correlated with the number of chunks they used. The researchers also noted that the more confidently chunks were used, the more they contributed to the perception of fluency (Boers & Lindstromberg, 2009: 36).

> [Towell et al.] found that the more fluent speakers were able to speak faster and with less hesitation due to the effective use of chunks.

Idiomaticity

The possession of a store of formulaic language helps resolve another of the 'puzzles' of native-like proficiency, alluded to by Pawley and Syder (1983), that of native-like selection, or idiomaticity, i.e. the capacity a language user has to distinguish 'those usages that are normal or unmarked from those that are unnatural or highly marked' (194). For example, given all the possible ways of performing a particular speech act, such as expressing regret, speakers of English typically choose to say *I'm [very/so] sorry*, rather than, say, *It causes me pain* (as in German *Es tut mir leid*) or *It tastes bad* (as in Spanish *Me sabe mal*), both of which would be grammatically well-formed in English.

Wray (2000) makes the useful distinction between chunks that are speaker-oriented, i.e. that enable fluent production, and those that are *hearer-oriented*, i.e. that achieve social and interactional purposes, such as polite formulae (e.g. *I wonder if you'd mind…?*) or expressions that assert group identity (e.g. 'teen talk': *can't even; yeah right*).

> The use of chunks can help students to be perceived as idiomatic language users, disposing of a relatively impressive lexical richness and syntactic complexity.

For language learners (and assuming native-like fluency is their goal), knowing how things are done in the target language is clearly an advantage. As Boers and Lindstromberg (2009: 37) note, 'The use of chunks can help students to be perceived as idiomatic language users, disposing of a relatively impressive lexical richness and syntactic complexity'.

Arguably, the emphasis placed on learning and testing phrasal verbs in many ELT courses reflects their iconic status as markers of idiomaticity. There is some evidence that memorized chunks do confer idiomaticity. For example, in a study of three exceptional Chinese learners of English (Ding, 2007), it was found that all three drew on the vast number of texts they had memorized as part of their schooling, and were able to extract idiomatic phrases from these:

> W. said that she was still using many of the sentences she had recited in middle school. For instance, while other students used 'Family is very important,' she borrowed a sentence pattern she had learned from Book Three of New Concept English: 'Nothing can be compared with the importance of family.' This made a better sentence, she said.

Language development

Another argument in favour of a lexical approach is that 'lexical phrases may also provide the raw material itself for language acquisition' (Nattinger & DeCarrico, 1989: 133). That is to say, the phrases are first learned as unanalysed wholes. 'Later, on analogy with many similar phrases, [the learners] break these chunks down into sentence frames that contain slots for various fillers' *(ibid.)*.

Lewis (1997: 211) himself made a similar point: 'The Lexical Approach claims that, far from language being the product of the application of rules, most language is acquired lexically, then "broken down"… after which it becomes available for re-assembly in potentially new combinations'. Proponents of 'usage-based' theories of language acquisition, such as Ellis (1997: 126), concur: 'Learning grammar involves abstracting regularities from the stock of known lexical phrases'. However, attempts to research this claim have met with mixed results, one problem being the difficulty of deciding if a learner's utterance is the result of holistic learning or of (re-)analysis and subsequent creativity – or a combination of both.

Wong Fillmore (1979), in a study of five Spanish-speaking learners of English, found evidence that formulaic sequences were gradually analysed, so that their constituent elements were available to be re-combined creatively: 'The formulas the children learned and used constituted the linguistic material on

which a large part of the analytical activities involved in language learning could be carried out' (212).

On the basis of this and similar studies (mainly with young learners), Ellis and Shintani (2014: 71) accept that 'the prevailing view today is that learners unpack the parts that comprise a sequence and, in this way, discover the L2 grammar. In other words, formulaic sequences serve as a kind of starter pack from which grammar is generated'.

Other researchers are less convinced. Granger (1998: 157-8), for example, analysed a corpus of adult L2 learner language and concluded that 'there does not seem to be a direct line from prefabs to creative language … It would thus be a foolhardy gamble to believe that it is enough to expose learners to prefabs and the grammar will take care of itself'. Wray (2002: 201) suspects that what might work for young learners does not necessarily work for literate adults, whose inclination is to unpack formulaic expressions – not for their syntax, but for their words. She concludes that 'there is very mixed evidence regarding the effectiveness of teaching formulaic sequences'. Finally, Scheffler (2015) argues that – even if these 'unpacking' processes apply in first language acquisition – the sheer enormity of the input exposure required in order to 'extract' a workable grammar is simply unfeasible in most L2 learning contexts.

It has also been argued that over-reliance on formulaic language may have adverse effects, contributing to the premature stabilization of the learner's developing language system. As Swan (2006: 5-6), for example, puts it, 'Much of the language we produce is formulaic, certainly; but the rest has to be assembled in accordance with the grammatical patterns of language [...]. If these patterns are not known, communication beyond the phrase-book level is not possible'.

> [It] remains highly plausible that formulaic sequences are supporting the acquisition process.

Wray (2000: 472) goes so far as to suggest that 'formulaic sequences may be used by some adult learners as a means of actually *avoiding* engaging with language learning' (emphasis added). That is, they are used as communication strategies at the expense of the development of a productive grammar. However, in a later book on the subject (2002: 188), she is more conciliatory: 'Despite such reservations, it remains highly plausible that formulaic sequences are supporting the acquisition process, whether this be simply by maintaining in the learner a sense of being able to say something, even when there is only a small database to draw on, or by providing a wealth of stored native-like data for later analysis'.

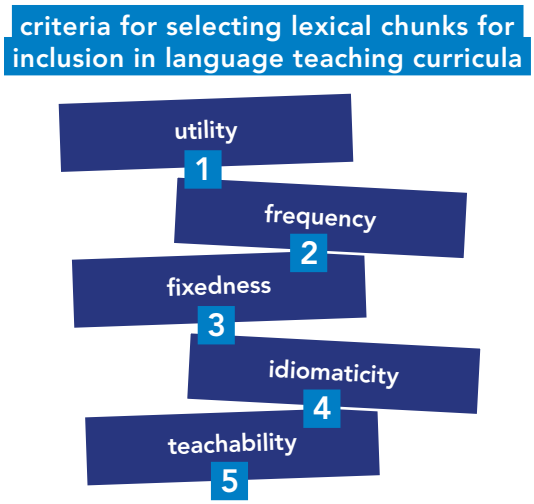
In short, the jury is still out on the role that chunks play in overall language development, and an exclusive focus on chunks may even be counter-productive, given the risk of over-reliance and consequent fossilisation. Nevertheless, their key role in facilitating fluency and, to a lesser extent, idiomaticity, is uncontroversial.

3. How can chunks be integrated into the second language curriculum?

(And, specifically, which chunks should be learned when?)

Since the advent of the Lexical Approach, there has been a perceptible shift in ELT course book design in the direction of including a more prominent focus on formulaic language. Nevertheless, there is a common perception, amongst corpus linguists in particular, that ELT materials have not kept pace with developments in the field. Hunston (2002: 38), for example, observes that 'unfortunately, phrases tend to be seen as tangential to the main descriptive systems of English, which consist of grammar and lexis'. In a similar vein, Granger & Meunier (2008: 251) complain that 'vocabulary teaching today is still too often exclusively word-based'. Where chunks, such as collocations, are included in course materials their selection appears somewhat arbitrary, 'largely grounded on the personal discretion and intuition of the writers' (Koprowski, 2005: 330).

In an attempt to rectify this situation, a number of researchers have proposed criteria for selecting lexical chunks for inclusion in language teaching curricula. These criteria include:



Utility

With regard to selecting chunks on the basis of their utility, or usefulness, one legacy of the functional-notional syllabuses of the 1970s has been the inclusion of functional language in the form of formulaic expressions associated with different speech acts, such as asking directions, making requests, etc. This was the direction advocated by Nattinger (1980: 342), an early proponent of a focus on chunks, 'since patterned phrases are more functionally than structurally defined, so also should be the syllabus… In that way, the items we select to teach would not be chosen on the basis of grammar but on the basis of their usefulness and relevance to the learners' purpose in learning'.

While functions are no longer the primary organising feature of mainstream courses, course designers are now better equipped in terms of identifying the most common exponents of different functions. *The Functional Language Phrase Bank* (Cambridge University Press)[2], for example, is compiled on the basis of corpus data, and typical exponents of common functions, such as giving advice, offering help or expressing agreement are tagged for frequency. And the practice of teaching useful classroom language in the form of fixed expressions (e.g. *What does X mean? How do you say Y? Can you spell that?*) is consistent with this functional imperative.

Frequency

Taking frequency as a starting point, Willis (2003: 166) notes that 'many phrases are generated from patterns featuring the most frequent words of the language'. This echoes Sinclair's earlier advice, to the effect that 'learners would do well to learn the common words of the language very thoroughly, because they carry the main patterns of the language' (1991: 79). Willis (*ibid.*) goes on to say that 'learners should be given the opportunity early on to recognise the general use of words, such as *about* and *for*, paving the way for the recognition and assimilation of patterns at a later stage'.

Hence, one way of organising a syllabus of phrases might be to peg it to the most common words – a principle that underpinned one of the first course books both to adopt a lexical syllabus and to be informed by corpus data, The Collins COBUILD English Course (Willis & Willis, 1988). The principle is perpetuated in course materials that focus on 'key words' (such as *take, get* or *way*) and diagram their common collocations.

Fixedness and idiomaticity

As a basis for selection, frequency, however, is problematic – as Boers & Lindstromberg (2009: 14) point out: 'Below the small group of highly frequent chunks, frequency distribution rapidly levels off, confronting the learner [and the teacher and the course book author] with a great many chunks of medium frequency'. In order to choose between these, they suggest enlisting criteria of fixedness and of idiomaticity. They argue that chunks which are relatively fixed in their form (such as *first and foremost*, by *leaps and bounds*) are, once learned, easier to deploy as a single item, and therefore more facilitative of productive fluency. And expressions that are idiomatic – hence semantically opaque, such as *every so often* and *by and large* – are likely to cause problems in comprehension, and therefore should be prioritised over those chunks that are relatively transparent.

2 https://languageresearch.cambridge.org/functional-language

In similar fashion, Martinez (2013) advocates selecting chunks on the basis, not just of frequency, but also of transparency (or lack thereof). Thus, an expression such as *take time* is both frequent and transparent – and, as such, probably doesn't merit detailed attention. On the other hand, an expression such as *take place* (meaning 'occur') is frequent but not transparent, and is therefore likely both to impede comprehension, and to be under-used in production. Hence it merits an instructional focus. Similar principles inform the PHRASE list of Martinez and Schmitt, mentioned on page 6.

In compiling their 'Academic Formulas List', Simpson-Vlach & Ellis (2010) used both quantitative and qualitative data in order to select those chunks that, in an academic register (both spoken and written), have 'high currency and functional utility'. Items were initially derived from a corpus based on raw frequency of occurrence plus their MI score (a statistical measure of the probability of co-occurrence). Twenty experienced EAP teachers were then asked to rate a sub-set of these items according to their 'formula teaching worth', i.e. how formulaic, how functional and how teachable they were. The results were extrapolated to the larger set of items to produce a definitive list, which was then sub-divided into functional categories, such as:

QUANTITY SPECIFICATION	CAUSE AND EFFECT	VAGUENESS MARKERS	HEDGES
a list of	*as a consequence*	*and so on*	*to some extent*
all sorts of	*the reason why*	*blah, blah, blah*	*it might be*
there are three	*as a result of*	*and so forth*	*a kind of*

Teachability

While 'teachability' is a somewhat elusive criterion, Boers and Lindstromberg (2009) argue that idiomatic expressions can be made more memorable, and hence more teachable, once their 'mnemonic potential' has been unlocked through teacher 'elaboration'. For example, when learners understand the sporting references that are encoded in expressions like *jump the gun, neck and neck*, or *on the ball*, they are more likely to remember them.

Likewise, drawing attention to the commonly used phonological repetition in expressions like *make-or-break, short and sweet, fair and square, time will tell,* etc., can enhance their memorability – 'for although these chunks have considerable mnemonic potential, relatively few learners will unlock it without prompting or guidance' (2009: 123). All things being equal then, it makes sense to select those chunks that are not only relatively frequent but are also teachable, in the sense that their mnemonic potential can be unlocked.

All things being equal then, it makes sense to select those chunks that are not only relatively frequent but are also teachable, in the sense that their mnemonic potential can be unlocked.

Other criteria for selection of lexical chunks that have been proposed include 'prototypicality' (Lewis, 1997: 190) and 'generalisability'. On the assumption that memorized chunks provide the 'raw material' for the development of the second language grammar, then there is a case for choosing to teach chunks that embed prototypical examples of the target grammar. For example, the classroom questions (mentioned earlier) *What does X mean? How do you spell it?* etc., are potentially analysable into instances of do-auxiliary use which can be generalisable to the creation of new questions. This is also an argument for not teaching idiomatic expressions that are 'non-canonical', i.e. that do not reflect current usage, such as *come what may, long time no see, once upon a time*, etc. This may also be the thinking behind Nattinger & DeCarrico's (1992: 117) argument that 'one should teach lexical phrases that contain several slots, instead of those phrases which are relatively invariant'.

To summarise: given that there are many more combinations of words than individual words, and given the learning load that is thereby implicated, the inclusion of lexical chunks into the curriculum requires rigorous and principled decisions with regard to selection and sequencing – decisions that are not always at the forefront when it comes to course planning and design.

The inclusion of lexical chunks into the curriculum requires rigorous and principled decisions with regard to selection and sequencing – decisions that are not always at the forefront when it comes to course planning and design.

4. How are chunks best learned and taught?

Having selected chunks for a pedagogical focus, how should that focus be implemented? Responses to this question have tended to fall into four groups, which may be summarised as:

1. **The phrasebook approach**
2. **The awareness-raising approach**
3. **The analytic approach**
4. **The communicative approach**

1. The phrasebook approach

Phrasebooks for travellers, of course, have always accepted the usefulness of memorizing set phrases (with or without unfilled slots) for specific situations, without the need for any explicit analysis. A phrase book approach to the learning of chunks makes similar assumptions. After all, if chunks are going to serve to lubricate fluent speech, they need to be readily and accurately retrievable from long-term memory and hence, as with all vocabulary learning for production, an element of deliberate memorization is essential.

Nattinger himself was not averse to the idea that techniques associated with (by then discredited) audiolingualism, such as pattern practice drills, might be rehabilitated for the purposes not only of committing lexical phrases to memory but of modelling their generative potential. 'Pattern practice drills could first provide a way of gaining fluency with certain fixed basic routines... The next step would be to introduce the students to controlled variation in these basic phrases with the help of simple substitution drills, which would demonstrate that the chunks learnt previously were not invariable routines, but were instead patterns with open slots.' (Nattinger & DeCarrico, 1992: 116–17).

Less behaviouristic, perhaps, is the technique of 'shadowing' whereby the learner listens to extracts of authentic talk, and 'subvocalises' at the same time. For younger learners, preselected chunks can be embedded in chants and songs – see, for example, Carolyn Graham's (1979) work on 'jazz chants'.

To summarise: the practical applications of the phrasebook approach might be:

- rote learning of formulaic expressions
- drilling
- shadowing
- jazz chants.

2. The awareness-raising approach

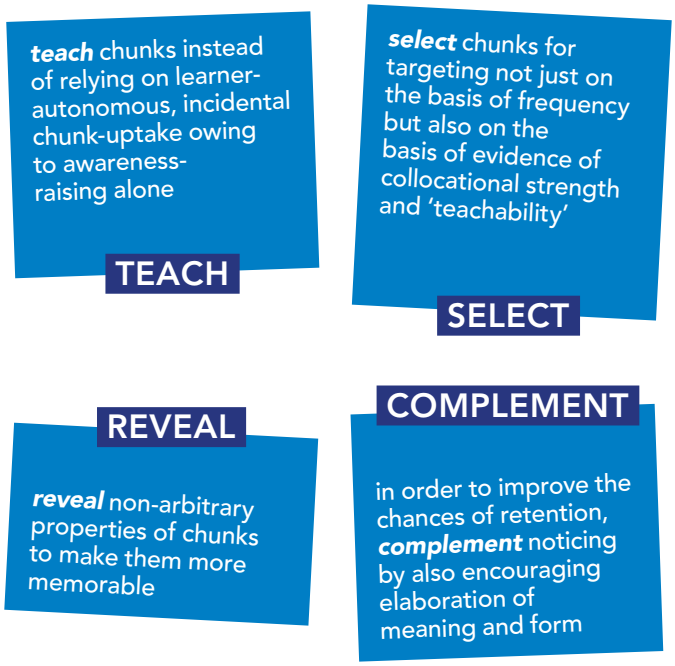
In formulating his Lexical Approach, Lewis showed no particular allegiance to any existing theory of second language learning, although he often makes reference to Krashen's (1985) Input Hypothesis and the necessity for high quantities of roughly-tuned input as a source for learning. Hence, in place of the dominant PPP [present-practise-produce] methodology of the time, he offers OHE (observe-hypothesise-experiment) – an inductive, awareness-raising approach, predicated on the learners noticing common sequences in the input. Lewis calls this process 'pedagogical chunking' (1997: 54) and its practical applications include:

- extensive reading and listening tasks, preferably using authentic material
- 'chunking' texts, i.e. identifying possible chunks, and checking these against a collocation dictionary, or by using online corpora (e.g. COCA: Davies, 2008) to check the relative frequency of word sequences
- listening to extracts of authentic speech and marking a transcript into tone units in order to identify likely chunks
- record-keeping and frequent review
- recycling chunks in learners' own texts, either spoken or written.

3. The analytic approach

While Boers and Lindstromberg (2009) agree with Lewis – that class time should be devoted to raising awareness about the role of chunks – they are sceptical as to whether learners will be able to identify chunks unaided. Moreover, their own research supports the view that directing learners' attention to the compositional features of chunks – e.g. their metaphorical origin or their phonological repetition – can 'unlock' their mnemonic potential and hence optimise their memorability (see above).

They summarise their more analytic approach in these terms:



The practical application of these principles is spelled out in more detail in their activity book, *Teaching chunks of language: from noticing to remembering* (Lindstromberg & Boers, 2008). Typical activities include:

- targeted teaching of selected chunks, e.g. those that are derived metaphorically from a particular 'source domain', such as seafaring: *give someone a wide berth*; take the helm, etc
- noticing patterns of sound repetition, as in *short and sweet*, *take a break*, etc
- use of memory-training techniques, such as mnemonics
- frequent recycling and review.

4. The communicative approach

Finally, coming from a background in communicative language teaching, with a focus on fluency in particular, Gatbonton & Segalowitz (2005) build on their earlier work in promoting 'creative automatization' by proposing an approach called ACCESS. This approach incorporates stages of controlled practice of formulaic utterances, embedded within communicative tasks. Put simply, this involves an initial stage in which short chunks of functional language are presented and practised, before learners take part in an interactive task that requires the repeated use of these chunks in order to fulfil a communicative purpose.

A well-known task type that fits this format is the 'Find someone who…' activity: learners have to survey one another according to prompts that require the use of a lexical phrase with an open slot, e.g. *Have you ever…?* This might first be highlighted and practised in advance of task performance. Gatbonton & Segalowitz (2005: 338) sum up the rationale for their version of the Lexical Approach in these terms: 'The ultimate goal of ACCESS is to promote fluency and accuracy while retaining the benefits of the communicative approach. In ACCESS, this is accomplished by promoting the automatization of essential speech segments in genuine communicative contexts.'

A similar approach was explored by Wray & Fitzpatrick (2008), in which intermediate/ advanced learners predicted conversational situations they were likely to encounter, and then worked with native speakers to script and rehearse these conversations, including the incorporation of formulaic language.

The researchers report that the subjects found:

> *the use of memorized sentences in anticipated conversations [was] a liberating experience, because it gave them exposure to an opportunity to sound native-like, promoted their fluency, reduced the panic of on-line production in stressful encounters, gave them a sense of confidence about being understood, and provided material that could be used in other contexts too (p.143).*

To summarise, the practical applications of a communicative approach to teaching chunks might include:

- the use of survey-type activities and guessing games that involve the repetition of formulaic expressions
- repeating speaking tasks, e.g. with different partners and/or to a time-limit, in order to encourage 'chunking'
- scripting, rehearsing, memorizing and performing dialogues that include formulaic expressions.

THE PHRASEBOOK APPROACH	THE AWARENESS-RAISING APPROACH	THE ANALYTIC APPROACH	THE COMMUNICATIVE APPROACH
• rote learning of formulaic expressions • drilling • shadowing • jazz chants	• extensive reading and listening tasks, preferably using authentic material • 'chunking' texts, i.e. identifying possible chunks, and checking these against a collocation dictionary, or by using online corpora (e.g. COCA: Davies, 2008) to check the relative frequency of word sequences • listening to extracts of authentic speech and marking a transcript into tone units in order to identify likely chunks • record-keeping and frequent review • recycling chunks in learners' own texts, either spoken or written	• targeted teaching of selected chunks, e.g. those that are derived metaphorically from a particular 'source domain', such as seafaring: *give someone a wide berth; take the helm*, etc • noticing patterns of sound repetition, as in *short and sweet, take a break,* etc • use of memory-training techniques, such as mnemonics • frequent recycling and review	• the use of survey-type activities and guessing games that involve the repetition of formulaic expressions • repeating speaking tasks, e.g. with different partners and/or to a time-limit, in order to encourage 'chunking' • scripting, rehearsing, memorizing and performing dialogues that include formulaic expressions

A summary of the practical applications for each of the four teaching approaches above:

Key considerations

Whatever procedures we adopt, it is worth bearing in mind that chunks are really just 'big words', and hence most – if not all – of the principles of effective vocabulary teaching apply equally to the teaching of chunks as they do to the teaching of individual words.

These include taking into account:

- the distinction between teaching for production (i.e. speaking and writing) as well as for recognition (listening and reading)
- the necessity of focusing on meaning as well as on form
- the importance of teaching vocabulary in context, rather than as isolated items
- the need for the deliberate teaching of vocabulary, rather than relying solely on incidental learning (Webb & Nation, 2017)
- the value of having learners make decisions about the items they are learning, according to the principle that 'the more one engages with a word (deeper processing), the more likely the word will be remembered for later use' (Schmitt, 2000: 121)
- the importance of learners forming associations with vocabulary items in order to establish mutually-reinforcing semantic networks as an aid to memory
- the necessity of regular review, including 'spaced repetition', i.e. reviewing previously learned material at increasingly larger intervals of time
- the value of having learners make their own decisions as to what and how they learn vocabulary, including setting their own targets and measuring their success at meeting these.

Summary: What are the implications for teachers?

The current state of play, with regard to the four questions posed above, suggests a number of implications with regard to classroom teaching. In brief these are:

- 'One of the future challenges for teachers will be to **help learners become aware of the pervasiveness of phraseology** and its potential in promoting fluency' (Granger & Meunier, 2008: 248). Fortunately, there are now a number of relatively recent titles that provide practical ideas for doing so, including Lindstromberg & Boers (2008), Dellar & Walkley (2016) and Selivan (2018).
- At the level of selecting and sequencing of chunks, teachers may need to **be more systematic and more rigorous,** taking into account not only frequency, but also such factors as utility, idiomaticity, fixedness, generalisability, and teachability.
- At beginner/elementary levels, chunk learning should **take the form of the formulaic ways that certain common speech acts are realised,** such as making requests, apologising, etc.
- Teachers should **be encouraged to exploit the texts they use** not only for their grammatical content, or their single-word vocabulary items, but also for any lexical chunks and patterns that fulfil the selection criteria listed above.
- **Familiarity with online corpus tools,** particularly those that provide collocational data, should be encouraged (see page 19, for a list of recommended websites).
- Teachers should **train learners in strategies for identifying possible chunks** in the input that they are exposed to, as well as strategies for recording and reviewing them, and for re-integrating them into their output.
- For teachers of younger learners, in particular, the **design and use of activities such as songs and chants should be promoted,** so as to maximise their chunk-learning potential.
- Teachers of specialised courses, such as English for academic purposes (EAP), should **foreground and highlight those formulaic features of different registers and genres,** including commonly used chunks that are typical.
- Testing of both spoken and written language should **include criteria that address the candidates' command of formulaic and/or idiomatic usage.**

At the same time, a balance needs to be maintained with other components of the curriculum. As Granger & Meunier (2008: 251) summarise it, 'Phraseology is a key factor in improving learners' reading and listening comprehension, alongside fluency and accuracy in production. However, its role in language learning largely remains to be explored and substantiated and it should therefore not be presented as the be-all and end-all of language teaching. Teachers have to perform a 'delicate balancing act', which consists of exposing learners to a wide range of lexical strings while at the same time ensuring that they are not overloaded with them and are also able to express key concepts and useful rules of grammar'.

Conclusion

The ubiquity and usefulness of lexical chunks is generally well accepted, but how chunks should best be integrated into existing teaching approaches is less clear: we have seen four different 'schools of thought', and it may be the case that the judicious selection of techniques from each approach may be the wisest course. And that implementing the principles of effective vocabulary teaching applies equally well to the teaching of chunks as it does to the teaching of individual words.

Scott Thornbury is a teacher educator and methodology writer. He teaches on an MA TESOL program run from The New School in New York. His two latest books (for Cambridge) are *30 Language Teaching Methods* and *101 Grammar Questions*.

To cite this paper:

Scott Thornbury (2019) *Learning language in chunks.* Part of the Cambridge Papers in ELT series. [pdf] Cambridge: Cambridge University Press

Available at cambridge.org/cambridge-papers-elt