

Norbert Schmitt
Ronald Carter
University of Nottingham

Vocabulary has been traditionally thought of as individual words. Of course, this layman's view is inadequate because vocabulary includes many units which are larger than individual orthographic words. For example, a single meaning is attached to each of the following: *give up*, *fish shop*, *freeze-dry*, and *burn the midnight oil*. Research using large corpora has shown that these *multi-word lexical units* (MWU) are ubiquitous in language use, at least in English (Moon, 1997). This is especially true of spoken discourse (McCarthy and Carter, 1997). Moreover, this same corpus research is now beginning to indicate that there is lexical patterning that exists beyond even the MWU level. Some longer strings of language recur frequently and are often connected with the functional usage of language. These longer strings have been called various names, including *lexical phrases* (Nattinger and DeCarrico, 1992), *lexical chunks* (Lewis, 1993), *lexicalized sentence stems* (Pawley and Syder, 1983), and *ready made (complex) units* (Cowie, 1992). Following Schmitt (2000), we will refer to these strings as *lexical phrases*. This paper will argue that lexical phrases are a key element of fluent language production and that they also play a part in vocabulary learning.

Lexical Phrases in Language Learning¹

本論では、言語学習におけるlexical phrasesの重要性について論じる。lexical phrasesは、習得され、頭の中に保存されている単語の連続体である。lexical phrasesは、頭の中に個々のかたまりとして存在しており、これが、流暢な言語使用を助けているということは心理言語学で証明されている。話し手は、新しい発話をする際、lexical phrasesをその発話の枠組みとして使うことができる。つまり、新しい発話の一部が既に頭の中ででき上がっているというのはとても便利なことである。また、聞き手は話し手の言葉を単語ごとに解読して意味を理解するのではなく、lexical phrasesを使って意味を理解している。lexical phrasesは、聞き手にとっても有益である。また、lexical phrasesは言語学習の上でも重要である。人は、lexical phrasesがわかれば、すぐにそれをその構成語に分解し、それらの語を習得している。lexical phrasesは、英語使用を促進させる重要な要素であるので、授業に取り入れられるべきである。導入方法については、提案がされ始めたばかりである。

The Psychological Reality of Lexical Phrases

Most of the evidence for lexical phrases comes from research into large databases of written and spoken language called *corpora*. Although these corpora contain the linguistic production of countless native-speakers, the data contained is only the manifestation of linguistic mental processing, and does not give us direct insight into the processes themselves. Of course, the fact that lexical phrases are so commonplace in the corpora surely suggests that they are in some way important in the mental processing of language. But do lexical phrases exist in the mind as single units, i.e. do they have psycholinguistic reality?

To answer this question, it might be useful to first look at individual words. Aitchison (1987a) suggests words can be stored either as wholes or composed on-line by combining stems and affixes. She reviews psycholinguistic experimental results and concludes that inflections are added on-line because they are rule-based and relatively consistent, and therefore can be added with little cognitive effort. On the other hand, prefixes and derivative suffixes are less transparent and are evidently retrieved from memory as polymorphemic wholes. This is partly because the affixation is somewhat arbitrary and

needs to be memorized on a case-by-case basis. For example, there is little reason why the noun derivation of *adjust* could not be *adjustion*, but it just happens to be *adjustment*. (Of course, there are phonological bases for some affix combinations; they are simply easier to pronounce.) One example of the psycholinguistic reasoning which leads to the conclusion that derivations are stored as wholes involves the 'bathtub effect' (Aitchison, 1987b: Chapter 11). This refers to the fact that the beginning of words tend to have the highest saliency, the ends of words the next highest, and the middle of words the least. If stems were stored separately from the affixes, one would expect the middle segment of derivations to be the most salient, not the prefix.

Therefore, the evidence seems to show that many words are stored as wholes. If this is true, the same should be possible for multi-word strings. In an early influential paper on lexical phrases, Pawley and Syder (1983) argue this is exactly the case. They do so by highlighting the 'puzzle of native-speaker fluency'. The puzzle is that native-speakers have cognitive limitations in how quickly they can process language but are also able to produce language seemingly beyond these limitations. Pawley and Syder look at the psycholinguistic literature and conclude that native speakers have been shown to be unable to process a clause of more than 8-10 words at a time. When speaking, they will speed up and become fluent during these clauses, but will then slow down or even pause at the end of these clauses, presumably to formulate the next clause. Speakers seldom pause in the middle of a clause. Together, this evidence suggests that speakers are unable to compose more than about 8-10 words at a time.

Native-speakers can, however, fluently say multi-clause utterances. Consider the following examples:

- 1) You don't want to believe everything you hear.
- 2) It just goes to show, you can't be too careful.
- 3) You can lead a horse to water, but you can't make him drink.

They have increasingly more words, and Example 3 is clearly beyond the limit of 8-10 words. Yet native speakers can say them all without hesitation. Pawley and Syder suggest that these examples can be fluently produced because they are actually lexical phrases—prefabricated phrases which are stored as single wholes and are, as such, instantly available for use without the cognitive load of having to assemble them on-line as one speaks. For this to work, only sequences of words which are frequent and familiar occurrences in language can be recognized as somehow being 'cohesive' and stored as wholes. Less common combinations or strings will continue to be generated via syntactic rules. If this were not true, our minds would become cluttered with countless word strings seldom, if ever, used. The thing that

makes lexical phrases so practical is that they are frequently used to express some functional use of language, thus warranting the memory space necessary to store them.

Production and Reception—Lexical Phrases Enable Fluency

If we accept that lexical phrases exist in the mind, then what insights does this give us into the underlying psycholinguistic processes of language production? One place to start is by looking into the strengths and limitations of the mind itself. One resource the mind seems to have plenty of is long-term memory capacity. However, resources in relatively short supply are working memory and processing speed. This limits the amount of information that the mind can process at any one time. The result is that the mind can store vast amounts of knowledge in long-term memory, but it is able to process only small amounts of it in real time, such as when one is speaking (Crick, 1979). Pawley and Syder (1983) argue that the mind makes use of a relatively abundant resource (long-term memory) to compensate for a relative lack in another (processing capacity). It does this with language by storing a number of frequently needed word strings (lexical phrases) as individual whole units which can be easily called up and used without the need to compose them on-line through word selection and grammatical sequencing. In this way there is less demand on cognitive capacity because the lexical phrases are already 'ready to go' and require little or no additional processing.

By stringing lexical phrases together, native speakers are able to produce stretches of fluent language. Because the use of lexical phrases lessens the cognitive load, they are able to concentrate on the content and organization of what they want to say. In contrast to this, second language learners may often have to pause between every 2 or 3 words because of the need to process language as well as thoughts.

Lexical phrases do not always have to be used in production as invariable wholes. They can also be used as a ready-made scaffold upon which to build language. An example of this is the lexical phrase

_____ (person) thinks nothing of _____ing (verb)
leading to

1. Diane thinks nothing of hiking 20 miles.
2. He thinks nothing of teaching six classes a day.
3. Mike thinks absolutely nothing of flying ten hours to go to a conference.

With the basic pattern already available from a lexical phrase, it would seem relatively easy to add on modifiers and clauses to customize the pattern to the situation at hand. Once again, we can see that the use of lexical phrases aid fluency, even when the language being produced is a creative adapta-

tion of an existing fixed pattern.

Thus far, we have discussed the effect of lexical phrases on the productive side of spoken discourse. However, the use of lexical phrases also eases the processing load of the listener. Because lexical phrases can be recognized as individual wholes, this spares the listener some of the processing effort of parsing the language stream into individual words, recognizing each of those individual words, and extracting meaning from the composite of the individual word meanings and the related grammatical structuring. So language which includes recognizable phrases is easier to understand. This can be related to the cooperative principles described by Grice (1975) which state that the speaker needs to use language which is relevant, clear, and appropriately concise. The use of lexical phrases in language helps to realize these cooperative principles. Certainly, lexical phrases are so commonplace in language that they are expected in any discourse. In addition, the use of lexical phrases tends to make language clearer. This is because most lexical phrases are the linguistic expressions of functional language use. An example of this is *Have you heard the one about ...*, which is used almost solely as the typical introduction to telling a joke or humorous story. This use is ritualized and can be anticipated whenever a speaker is preparing to tell a joke. Since it can be anticipated, its meaning is instantly clear. Also, since it is a ritualized, nothing else needs to be said in introducing the joke. Thus lexical phrases are typically the most economical way of performing functional language (sometimes called *speech acts*), which is a major reason why the formulaic use of lexical phrases is so tightly related to functional language use. This also means that lexical phrases which realize functional language use are maximally relevant to the functional situations in which they are used. Taken together, the use of preformulated sequences seems to have real advantages for both the speaker and listener. Thus it is not at all surprising that McCarthy and Carter (1997) found widespread evidence of these sequences in their analysis of the CANCODE2 spoken corpus. Common examples are *the thing is ___* (meaning 'the problem/point is ___'), and *I see* (I understand).

Learning Vocabulary by the Segmentation of Lexical Phrases

We have seen how lexical phrases aid in language use, but it has also been argued that they play a part in vocabulary acquisition. To illustrate how this might work, we need to first introduce the idea of *item learning* versus *system learning*. In language acquisition, learning seems to take place in two ways in phonology, morphology, and other linguistic aspects:

1) Item learning: learning individual units, such as

the words *sled* and *walked*

2) System learning: learning the system or 'rules', such as *sl=s+l* and *walked=walk+ed*

Lexical phrases clearly fall into the category of item learning, because their key feature is that they are wholes. Grammar, on the other hand, falls into the category of system learning. But these two types of learning are not mutually exclusive; rather they feed into one another. Thus once a lexical phrase is known, it can be analyzed and segmented into its constituent words. In this way, unanalyzed phrases can be analyzed to provide additional vocabulary. Hakuta (1974) was the first to suggest that phrases could be analyzed into words plus grammar. Wong-Fillmore (1976: 640) also believes that L2 children use many prefabricated phrases which "evolve directly into creative language". Peters (1983) presents the argument in its most considered form, proposing that learning vocabulary from lexical phrases is a 3-part process. First, lexical phrases are learned which are frozen wholes with no variation possible. At this point they are unanalyzed and are single lexemes. Common examples are idioms e.g. *kick the bucket*; *burn the midnight oil*, and proverbs e.g. *An apple a day keeps the doctor away*; *A stitch in time saves nine*. Also included are some expressions which are tightly related to a functional use e.g. *Ladies and Gentlemen ...* which is a typical opening address in a formal situation.

Second, a language learner may realize that some variation is possible in certain lexical phrases, and that they contain open slots. For example, after having heard the phrase *How are you today?* several times, it may be acquired as a lexical phrase with the meaning of 'a greeting'. However the learner may later notice the phrases *How are you this evening?* or *How are you this fine morning?* At that point, the learner may realize that the underlying structure is actually *How are you _____?*, where the slot can be filled with most time references. The learner is then able to realize that what fits in the slot is a separate lexical unit from the rest of the phrase, which opens the door to learning that lexical unit. Thus phrases can be segmentalized into smaller lexical units, oftentimes individual words. Lexical phrases at this stage are partly fixed and partly creative.

Third, this segmentation process can continue until all of the component words are recognized as individual units by use of syntactic analysis. When this happens, every word in the lexical phrase is potentially available for learning. This does not mean that the segmentation process has to continue to this point; in fact it can stop at any stage. There are some lexical phrases which the learner may never start to analyze, and which may be retained only as unanalyzed wholes. Likewise, learners may or may not realize that certain lexical phrases contain vari-

ability and slots. When the variability is realized, it is possible that only the slots are analyzed; the rest of the pattern may remain unanalyzed. Still, it seems safe to assume many, if not most, of the lexical phrases a learner knows will eventually become fully analyzed, and Peters (1983) suggests that much of a learner's vocabulary is learned in this way. This is especially true because learners are likely to eventually know numerous lexical phrases, seeing how they are easy to learn, efficient to use, and cover a wide variety of lexical content.

This segmentation process can lead to more than lexical knowledge however. Segmentation also requires grammatical information, which focuses attention on syntax as well as lexis. Ellis (1997) argues that grammar can be learned through the implicit recognition of the patterns in strings of language, some of which are bound to include lexical phrases. In this line of reasoning, innate grammar would not consist of an inborn understanding of grammatical rules, but rather a facility for recognizing the systematic patterns in language input. A perceptive ability to recognize such patterning does seem to be a sufficient condition for at least some types of grammar acquisition; a model developed by Kiss (1973) demonstrated that simply calculating which words occur sequentially eventually provides

enough input to distinguish their word-class.

Some Implications of Lexical Phrases

A main implication of lexical phrases is multiple storage in the mental lexicon. A large number of lexical phrases are likely to be fully analyzed, even though they are retained in longer-term memory because of their utility. Thus it is possible that the production of a frequent sequence of words can stem from the retrieval of a lexical phrase, or from the syntactic generation of the string from individual words. (It would seem likely that the lexical phrase approach will be used when possible due to the lower cognitive load.)

This means that if a learner produces a sequence of words which contains an error, the source of the error might be a weakness in lexical or grammatical construction, or it might be that a lexical phrase has been acquired in a faulty manner. If the language error is due to a faulty lexical phrase, then any amount of grammar-based correction would seem unlikely to remedy the error. What would be required is a re-learning of the correct form of the lexical phrase.

One should not assume that this paper downplays the importance of grammar in language use or in language teaching. The point is that language ability requires not only the ability to produce language

Figure 1. Suggestions for ways of teaching awkward and frequently occurring words such as *just*.

Keyword: just

Match these remarks and responses:

1. Would you like a cup of coffee?
 2. Are you ready? It's time we were off.
 3. It looks as if the train is going to be late.
 4. Were you late last night?
 5. Everybody is worried about the situation.
 6. They've changed their mind again.
 7. It's almost 9 o'clock. It's time we started.
 8. Have you got Helen's phone number there?
- a. That's just what we don't need.
 - b. Oh, it's not just me, then.
 - c. No, we got there just in time.
 - d. I think so. Just a moment—I'll have a look.
 - e. Not just now, thanks.
 - f. Don't worry. I think everything is just about ready.
 - g. That's just what I expected.
 - h. Right, I'll just get my coat.

Notice all the responses include the word *just*. It is very difficult to translate *just*, but it is used in a lot of fixed expressions. Can you think of a similar word in your own language? Learn the responses so you can use them yourself. Make sure you know the equivalents in your language.

Sometimes *just* is used to make a problem or mistake seem less important or serious:

It just slipped my mind.
I just couldn't get there any earlier.
I'm just not going to get upset about it.

Pres. Perfect: *I've just passed my exam.* (*just* = very recently)

Pres. Cont: *I'm just making some tea.* (*just* = emphasizing exactly now)
 was going to: *I was just going to ring you.* (*just* = very soon after now)

Can I just ask/tell you/say that ... (If you know that an interruption will be quick)

I was talking to her just now. (*just now* = a short time ago)

I couldn't tell you just then. (*just then* = at that particular time)

Source: Lewis 1997. ©Language Teaching Publications

through syntactic generation (via grammatical competence), but also requires the ability to use lexical phrases. This is especially true if learners hope to gain the pragmatic fluency which comes from knowing the right lexical phrase for the right situation. Ultimately, language learners need both abilities to use language well.

The importance of lexical phrases would suggest that we need to include instruction on them in our language teaching. As this is a new area, it is not yet clear how we can best achieve this. We do not even have a list of the most frequent lexical phrases in English as of yet, although this gap is now being addressed by Dave and Jane Willis. Michael Lewis (1997) has done some preliminary work in the area of teaching lexical phrases, advocating a focus on inducing patterns from language input and the return of a limited amount of pattern drilling. An example of this approach is illustrated in Figure 1. Lewis's proposals are generating a great deal of interest and seem intuitively attractive, but at this point it must be said that there is little empirical evidence one way or the other as to their actual effectiveness.

Conclusion

Corpus research is making it clear that the patterning resulting from lexical phrases is a major component of language. As such, lexical phrases are likely to become an increasingly important topic in Applied Linguistic circles, simply because lexical phrases are a key element in how language is used. In light of their essential nature, we need to come to a better understanding of their behavior and develop innovative ways of incorporating lexical phrase instruction into the language syllabus.

Note

- 1 This is a revised version of a presentation given at the 1998 PASE (Polish Association for Studies of English) conference in Szczyrk, Poland.
- 2 The CANCODE (Cambridge and Nottingham Corpus of Discourse English) is a 5 million word corpus of unscripted spoken discourse compiled at the University of Nottingham in conjunction with Cambridge University Press.

References

- Aitchison, J. (1987a). Reproductive furniture and extinguished professors. In Steele and Threadgold (eds.) *Language Topics. Essays in Honour of Michael Halliday*, Vol. 2. Amsterdam: John Benjamins.
- Aitchison, J. (1987b). *Words in the Mind*. Oxford: Blackwell.
- Cowie, A.P. (1992). Multiword lexical units and communicative language teaching. In Arnaud, P. and Béjoint, H. (eds.) *Vocabulary and Applied Linguistics*. London: Macmillan.
- Crick, F.H.C. (1979). Thinking about the brain. *Scientific American* No. 9.

- Ellis, N.C. (1997). Vocabulary acquisition: Word structure, collocation, word-class, and meaning. In Schmitt, N. And McCarthy, M. (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Grice, H.P. (1975). Logic and conversation. In Cole, P. and Morgan, J.L. (eds.) *Syntax and Semantics Vol. 3: Speech Acts*. New York: Academic Press.
- Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning* 24.
- Kiss, G.R. (1973). Grammatical word classes: A learning process and its simulation. In Bower, G.H (ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*. Vol. 7. New York: Academic Press.
- Lewis, M. (1993). *The Lexical Approach*. Hove, England: LTP.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Hove, England: LTP.
- McCarthy, M. and Carter, R. (1997). Written and spoken vocabulary. In Schmitt, N. And McCarthy, M. (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Moon, R. (1997). Vocabulary connections: Multi-word items in English. In Schmitt, N. And McCarthy, M. (eds.) *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Nattinger, J.R. and DeCarrico, J.S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Pawley, A. and Syder, F.H. (1983). Two puzzles for linguistic theory: Native like selection and native like fluency. In Richards, J. and Schmidt, R. (eds.) *Language and Communication*. London: Longman.
- Peters, A. (1983). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Wong-Fillmore, L. (1976). *The Second Time Around: Cognitive and Social Strategies in Second Language Acquisition*. Stanford University: Unpublished doctoral dissertation.

Norbert Schmitt lectures at the University of Nottingham, UK in vocabulary studies, language testing, and TESOL methodology. He recently published *Vocabulary in Language Teaching* with Cambridge University Press. He likes traditional Irish music and often dreams of renewing his private pilot's license.

norbert.schmitt@nottingham.ac.uk



Ron Carter is professor of Modern English Language at the University of Nottingham, UK. His latest book on vocabulary is *Vocabulary: Applied Linguistic Perspectives* (2nd ed., Routledge). It would take a lot of money to get him in a plane flown by Norbert Schmitt.

ronald.carter@nottingham.ac.uk

