CHUNKS IN THE CLASSROOM: LET'S NOT GO OVERBOARD

Formulaic language
Formulaic language ('chunks') has attracted increasing attention among researchers and teachers in recent years, as the growth of large electronic corpora has made it easier to tabulate the recurrent combinations that words enter into. Such combinations include, for instance:
- fixed phrases (idiomatic or not) such as *break even*, *this morning*, *out of work*
- collocations (the preferences that some words have for particular partners) such as *blazing row* (more natural than *burning row*) or *slightly different* (more natural than *mildly different*)
- situationally-bound preferred formulae such as *Sorry to keep you waiting* (more natural than *Sorry I made you wait*)
- frames such as *If I were you, I'd …* , *Perhaps we could …* or *I thought I'd …*

   Researchers differ in their analysis and classification of formulaic language, and the storage and processing models they propose – see Wray (2002) for a clear and comprehensive survey. It is, however, generally agreed that these chunks behave more like individual words than like separately constructed sequences. *Unemployed* and *out of work*, for instance, both consist of three morphemes. If the first is handled mentally as a unit for comprehension and production, rather than being analysed into or built up from its constituents every time it is processed, it seems reasonable to suppose that its multi-word synonym may be treated similarly, even if we happen to write this with spaces between the three components.

   Languages clearly contain very large numbers of such items: one often-quoted estimate suggests that English may have hundreds of thousands. If this seems implausible, think how many common fixed expressions are built around one meaning of the noun *work*: *at work, work in progress, go to work, a day's work, man's/woman's work, take pride in one's work, part-time work, shift work, the world of work, nice work, carry out work, in the course of one's work, out of work, build on somebody's work, work permit, take work home, equal pay for equal work, the work of a moment, look for work, all my own work …* It seems possible, in fact, that languages may have preferred formulaic sequences for virtually every recurrent situation that their speakers commonly refer to.

   Language of this kind is notoriously challenging for learners. A knowledge of grammar and vocabulary alone will not indicate that *slightly different* is preferred to *mildly different*, or that *Can I look round?* is a more normal thing to say in a shop than *May I see what you have?* – such things have to be learnt as extras. Paradoxically, therefore, what looks easiest may be hardest. To construct a novel utterance like 'There's a dead rat on the top shelf behind Granny's football boots', a learner only needs to know the words and structures involved, but such knowledge will not help him or her to produce a common phrase like 'Can I look round?'– if the expression isn't known as a whole, it can't be invented. Since chunks constitute a large proportion of spoken and written text – studies put forward figures ranging between 37.5% and 80% for different genres – it seems sensible to give them a central role in our teaching, and we are often urged to do so. Four reasons are commonly advanced.

'Chunks save processing time'
The brain has vast storage capacity, and memorisation and recall are cheap in terms of mental resources. For a foreign learner, as for a native speaker, it is obviously more

efficient to retrieve *If I were you* as a unit than to go through the process of generating the sequence from scratch in accordance with the rules for unreal conditionals. Using chunks means that processing time and effort are freed up and made available for other tasks.

'You can learn grammar for free'
Children learn their mother-tongue grammar by unconsciously observing and abstracting the regularities underlying the sequences they hear. Many of these sequences are recurrent and formulaic (*Who's a good baby, then?*; *'s time for your bath*; *If your father was here now*; *One more spoonful*; *All gone*), and children's internalisation of such elements plays a central role in acquisition. It seems logical that second language learners, too, should be able to take a similar route, abstracting the grammar of a language from exposure to an adequate stock of memorised formulae. Lewis (1993) suggests for instance that, instead of learning the *will*-future as a generalised structure, students might focus on its use in a series of 'archetypical utterances', such as *I'll give you a ring, I'll be in touch, I'll see what I can do, I'll be back in a minute*.

'You can produce grammar for free'
Formulaic 'frames' bring their grammar with them. Take for example a sentence like *I thought I'd start by just giving you some typical examples of the sort of thing I want to focus on.* This consists almost entirely of frames and fixed expressions:
- *I thought I'd* + infinitive
- *start by ...ing*
- *give you* + noun phrase
- *typical example of* + noun phrase
- *the sort of thing* + (*that*)-clause
- *I want to* + infinitive
- *focus on*.

So, given a knowledge of the component frames and expressions, the sentence can be produced with minimal computation – hardly any reference to general grammatical rules is required.

'A mastery of formulaic language is desirable/necessary if learners are to approach a native-speaker command of the language'

Even students who have an advanced knowledge of English grammar and vocabulary may be far from native-speaker-like in their use of the language. What lets them down is likely to be their imperfect mastery of formulaic language, especially collocation and situationally-bound language. This seems, therefore, an obvious area for pedagogic intervention. '… formulaic sequences have been targeted in second language teaching because they seem to hold the key to native-like idiomaticity' (Wray 2000).

How good are these reasons?

Persuasive though these arguments are, they need to be looked at critically.

- Storage may be cheap in terms of mental resources, but putting material into store is extremely time-consuming. Learning quantities of formulaic sequences may exact a high price in exchange for the time eventually saved.
- The question of whether classroom learners are able to generalise from formulaic sequences without explicit instruction has scarcely been investigated. It seems likely that (as with first-language learning), a vast amount of exposure would be necessary for adult learners to derive all types of grammatical structure efficiently from lexis by the analysis of holistically-learned chunks; and this amount of exposure is not available in instructional situations. As Granger (1998) puts it 'It would … be a foolhardy gamble to believe that it is enough to expose L2 learners to prefabs and the grammar will take care of itself'.
- Much of the language we produce is formulaic, certainly; but the rest has to be assembled in accordance with the grammatical patterns of the language, many of which are too abstract to be easily generated by making small adjustments to memorised expressions or frames. If these patterns are not known, communication beyond the phrasebook level is not possible – as Scott Thornbury once memorably put it, language becomes 'all chunks but no pineapple'. Grammar hasn't gone away because we have rediscovered lexis.
- Most importantly, the notion that foreign learners should aspire to a 'native-speaker command' of phraseology, or anything similar, requires very careful examination.

The native-speaker target

Discussion of the acquisition of formulaic language often assumes something approaching a native-speaker target:

> It appears that the ability to manipulate such clusters is a sign of true native speaker competence and is a useful indicator of degrees of proficiency across the boundary between non-native and native competence. (Howarth 1998a).
> It is impossible to perform at a level acceptable to native users, in writing or in speech, without controlling an appropriate range of multiword units. (Cowie 1992)

Such sweeping pronouncements are, however, of little value in the absence of clear quantified definitions (which we do not have) of such notions as 'a level acceptable to native users' and 'an appropriate range of multiword units.' No doubt certain lexical chunks need to be mastered for certain kinds of pragmatic competence; but we need to know which chunks, for what purposes. Certainly, a mastery of relevant formulaic and other language is necessary for effective professional or academic work, as ESP and EAP teachers are well aware.

> Both undergraduates and postgraduates serve a kind of apprenticeship in their chosen discipline, gradually familiarising themselves not only with the knowledge and skills of their field, but also with the language of that field, so that they become capable of expressing their ideas in the form that is expected. As they do this, their use of formulaic sequences enables them, for example, to express technical ideas economically, to signal stages in their discourse and to display the necessary level of formality. The absence of such features may result in a student's writing being judged as inadequate. (Jones and Haywood 2004)

Assimilating the necessary formulaic inventory of a particular professional group is not, however, the same thing as acquiring a generalised native-speaker-like command of multi-word lexical expressions. The first is necessary and achievable, the second is neither, and to require such a command of non-native students is unrealistic and damaging. The size of the formulaic lexicon makes it totally impracticable to take native-speaker phraseological competence, or anything approaching it, as a realistic target for second-language learners. (Memorising 10 formulaic items a day, a learner would take nearly 30 years to achieve a native-speaker command of. say, 100,000 formulaic items.)

Consciousness-raising and strategies

One response to the practical impossibility of teaching native-speaker-like formulaic competence is to recommend equipping learners with a conscious awareness of the learning task they face, as suggested by Howarth (1998b), or with strategies which will 'enable them to acquire the knowledge needed to use formulaic sequences accurately and appropriately in their own work' (Jones and Haywood 2004).

It is of course helpful to advise students to pay attention to and memorise instances of formulaic language (to the extent that they do not already do so). However, since formulaic expressions have to be learnt individually, like other kinds of lexis, it is not immediately clear how the enormous learning problem can be addressed, and native-speaker competence approached, by either consciousness-raising or the deployment of ill-defined strategies. Transferring the problem from the teacher to the learner in this way does little to solve it.

Realism and prioritising

Given these problems, our only realistic course, as more pedagogically oriented writers such as Willis (1990) or Lewis (1993) point out, is to accept our limitations and to prioritise. Most non-native speakers must therefore settle for the acquisition of a variety characterised by a relatively restricted inventory of high-priority formulaic sequences, a correspondingly high proportion of non-formulaic grammatically generated material, and an imperfect mastery of collocational and selectional restrictions. This may seem disappointing, but there is nothing we can do about it – languages are difficult and cannot generally be learnt perfectly. Failure to recognise this may lead teachers to neglect important aspects of language teaching, in order to devote excessive time to a hopeless attempt to teach a comprehensive command of formulaic language – like someone trying to empty the sea with a teaspoon.

References

Cowie, A. 1992. 'Multiword Lexical Units and Communicative Language Teaching' in *Vocabulary and Applied Linguistics*, Arnaud, P. and Béjoint, H. (eds.). London, MacMillan.

Granger S. 1998. 'Prefabricated Patterns in Advanced EFL Writing; Collocations and Formulae' in Cowie. A. (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford, Oxford University Press.

Howarth, P. 1998a. 'Phraseology and Second Language Proficiency' *Applied Linguistics* 19/1.

Howarth, P. 1998b. 'The Phraseology of Learners' Academic Writing' in Cowie. A. P. (ed.) *Phraseology: Theory, Analysis and Applications*. Oxford, Oxford University Press.

Jones, M. and Haywood, S. 2004 'Facilitating the Acquisition of Formulaic Sequences' in Schmitt, N. (ed.) *Formulaic Sequences*. Amsterdam, John Benjamins.

Lewis, M. 1993. *The Lexical Approach*. Hove, Language Teaching Publications.

Willis, D. 1990. *The Lexical Syllabus*. London, Collins.

Wray, A. 2000. 'Formulaic Sequences in Second Language Teaching: Principles and Practice' *Applied Linguistics* 21/4.

Wray, A, 2002. *Formulaic Language and the Lexicon*. Cambridge, Cambridge University Press.