Future Directions in Formulaic Language Research

Alison WRAY
Research Professor, Centre for Language and Communication Research, Cardiff University, UK
WrayA@cardiff.ac.uk

Abstract

This paper describes the development of research into formulaic language and situates recent research within the broader, cyclical, research process. Three main areas of observation and investigation in formulaic language research are presented: native speaker language (including first language acquisition), clinical language, and second language learning. Developments made so far in each area are summarised. Previous work by this author is outlined, providing a cross-disciplinary explanation of formulaic language, pointing to some practical ways forward for future research in the field, and suggesting new research questions. Second language teaching and learning is directly concerned by the need to balance holistic and analytic approaches to language use. The status of multiword strings as useful phrases to be memorised or as forms of complicated words incorporated into vocabulary learning may reflect assumptions about what is useful for learners and the ways in which teachers can help them to achieve native-like fluency, or idiomaticity. A description of The Formulaic Language Research Network is followed by a review of six claims in current research and opportunities for development. The six claims concern variation in idiomaticity, social causes of idiomaticity, idiomaticity as a consequence of processing, whose processing is being reduced?, where do collocations fit in?, and variation according to genre.

Résumé

Cet article décrit le développement des recherches portant sur les séquences pré-fabriquées (*formulaic language*) et situe les travaux récents dans un processus de recherche cyclique plus vaste. Trois domaines principaux d'observation et d'investigation sont présentés : le langage de locuteurs natifs (y compris l'acquisition de la langue maternelle), les troubles du langage et l'apprentissage de langues secondes. Les avancées dans chaque domaine sont résumées. L'auteure donne un aperçu de ses propres travaux, qui fournissent des explications transdisciplinaires concernant les séquences pré-fabriquées, proposent des perspectives concrètes pour de futures recherches et soulèvent de nouvelles questions. L'apprentissage et l'enseignement de langues secondes sont directement concernés par le besoin d'équilibrer des approches holistiques et analytiques pour apprendre à utiliser la langue cible. Le statut des énoncés multilexicaux, soit comme des séquences qui seraient intéressantes à mémoriser, soit comme des mots compliqués à intégrer dans l'apprentissage du vocabulaire, reflète peut-être des représentations quant à ce qui est utile pour les apprenants et quant aux moyens que les enseignants peuvent employer pour les aider à aboutir à une fluidité proche de celle des natifs, ou à une expression idiomatique. Une

N°4, 2013

ISSN 1760-4753

description du réseau de recherche sur les séquences pré-fabriquées (*Formulaic Language Research Network*) est suivie d'une discussion des résultats récents obtenus dans six secteurs ainsi que des perspectives de développement. Ces résultats concernent la variation de l'idiomaticité ; les causes sociales de l'idiomaticité ; l'idiomaticité en tant que conséquence du traitement ; la question à savoir si la réduction du traitement concerne à la fois la production et la compréhension ; le rôle des collocations ; et la variation liée au type de discours.

Keywords: clinical language, collocations, formulaic language, foreign language teaching, idiomaticity, idioms, second language learning

Mots clés : apprentissage des langues, didactique des langues, idiomaticité, idiomes, séquences pré-fabriquées, troubles du langage

Outline/Plan

N°4, 2013

ISSN 1760-4753

1.0 Introduction[1]

In this paper I offer some reflections on the development of research into formulaic language: how it started, what stage it has reached, and where I think it might go in the future. My intention is to encourage researchers to explore the fullest range of opportunities for making a useful contribution to work in this field. Research into a particular topic or issue often falls into patterns across time, with periods of intense innovation followed by phases of consolidation and confirmation, and then of reflection, during which the previous work is reviewed and located within broader frameworks and areas of investigation. Then the cycle can begin again. It is quite important for us as researchers to have some sense of where in the cycle things are. We establish this by examining the sort of work that is predominantly being undertaken at the moment, and considering how the questions being asked relate back to previous work. I think there are clear signs that research into formulaic language is reaching the end of the consolidation and confirmation part of its cycle. We have entered the transition to a period of questioning what we know, and how our understandings of formulaic language relate to other developments in linguistic research. Claims previously taken for granted are being questioned, and we should anticipate that new models of the nature of formulaic language will soon emerge.

We are, then, at a very exciting moment, and can anticipate some interesting developments in the research now underway and soon to begin. However, this point in the cycle is also potentially hazardous, particularly for less-experienced researchers, of whom there are a great many working on formulaic language—for it has proved a rich seam for postgraduate students. The most comfortable place in the cycle for new researchers is the consolidation phase, where it is possible—and indeed necessary—to pick up an existing theory, model or claim, and test it. During the latter half of the last decade the majority of the work on formulaic language that I read, particularly in the domain of second language learning, was of this kind.

Such work was urgently needed, and has made a very valuable contribution to our understanding of what this phenomenon is, and what a language teacher or learner can do with it. However, now we have a reasonable volume of empirical evidence, it is time to develop the research in other ways. We do this by considering and comparing the ideas that begin to emerge in the discussion sections of research papers—the questions and suggestions about what it *means* that a given theory or model was or was not supported by the data. It requires a measure of boldness, as a relatively new researcher, to take on this kind of activity and responsibility, and, in truth, it probably involves a great deal more work too. Rather than just reading existing published accounts, accepting their claims, turning one of the claims into a hypothesis and collecting data to see if it is true (all the while being fairly confident that it will be), something much more daring is needed. The researcher needs to examine the detail of the existing claims, the evidence on which

1. This is a revised and updated version of a paper published in the Journal of Foreign Languages 2009, generously permitted by the editor of the JFL, Dingfang Shu.

N°4, 2013

ISSN 1760-4753

they were built, and the assumptions underlying them, and then apply both imagination and intense scholarship to develop new insights and ideas.

The scholarship aspect is hugely important, and entails familiarity with not only the literature on formulaic language but also of the research with which it intersects: normally one or more of second language acquisition theory and practice, clinical linguistics, first language acquisition, grammatical theory, psycholinguistics and corpus linguistics. The imagination part is equally necessary though, because just knowing what is in that literature is not enough. One has to compare, contrast, make links. One has to ask, for instance, whether a particular claim or assumption in one domain can really be true, given what we know about another domain.

For instance, there is now evidence from psycholinguistic and neurolinguistic studies that some types of formulaic language are processed faster than, or differently from, non-formulaic language (Siyanova-Chanturia, Conklin, & Schmitt, 2011, Siyanova-Chanturia, Conklin & van Heuven, 2011; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). But do those findings apply to all types of formulaic language, and if not, is that simply because other types are more difficult to test, or might there be differences between types that seemed insignificant before, but now must be examined in the light of this evidence (Wray, forthcoming)? To take another example, which I pick up on briefly later, how should we proceed with the striking convergence of work on formulaic language and on Construction Grammar and other emergent grammar models? How different are these approaches? Are we adequately addressing the commonalities and contrasts and if not, what risks are there for future misunderstandings?

Such questions are difficult to answer, and the researcher bears heavy responsibility to answer them appropriately and with as much information and circumspection as possible, for it is from this work that the next set of claims, theories and models emerge, and they will be the basis for the testing and consolidation that occupies the following generation of researchers.

In section 2, I shall outline the development of research into formulaic language so far, as I see it. Inevitably, this is a reflection of my own interests and biases—particularly as a psycholinguist. In section 3 I identify six claims about formulaic language that future research needs to examine and test, and home in on some questions that researchers in the area of language learning and teaching might particularly consider.

2.0 The emergence and consolidation of formulaic language as a research topic

Research into formulaic language research was, for many years, rather fragmented, with at least three areas of observation and investigation developing rather independently: native speaker language (including first language acquisition), clinical language, and second language learning.

N°4, 2013

ISSN 1760-4753

2.1 Patterns in native language

Those investigating language patterns in native speakers have always known that some words clump together, and that our 'ordinary' language, particularly in speech, is more repetitive, clichéd or idiomatic than formal written language is. A number of observations from linguists in the first half of the 20th Century bear this out (see Wray, 2002b, pp. 7-8 for a brief account) and suggest that, at that time, a certain flexibility in the size of the lexical unit was not as contentious as it later became under the influence of Chomsky's atomic view of the lexicon. A point of interest has been whether the status of certain multiword strings 'belonging together' is anything more than simply a by-product of our tendency to want to express the same messages many times. If it were more, might at least some of the repetition be due to direct benefits to the speaker and/or hearer that accrue from using familiar forms? In linguistic theory a tension has arisen between the evident fact that we are able to understand and produce language forms that we have not previously encountered—which clearly indicates that we use rules to create and extrapolate meaning—and the suggestion that we might take the opportunity, where possible, to bypass novelty in favour of routine. Much linguistic research today is about where to draw the line between the capacity for novelty and the processing advantages of being less novel.

The necessity of balancing the holistic and analytic approaches to language management was well made in two papers by Pawley & Syder (1983a; 1983b). More recently, Pawley (2007) has provided his own overview of how formulaic language research, in the native speaker domain, has developed over the past four decades. One of the key changes in research during that period has been the advent of computational means to examine language patterns. Rather than simply asserting that a particular expression is common, it is now possible to demonstrate just how common, and to compare its distribution with that of other expressions of compatible structure or meaning. A waymark in this transition was Sinclair's (1991) book reflecting on discoveries made while working on the COBUILD corpus project. He introduced there the oft-cited distinction between the 'open choice principle' and the 'idiom principle', suggesting that in normal language use we first try to match a larger wordstring to a lexical representation in memory, before resorting to the more demanding alternative of decoding each word separately.

2.2 Formulaicity in language disorders

A long, quite independent tradition of observations existed regarding the retention of certain wordstrings after brain damage that had destroyed the capacity for novel sentence construction. The earliest observations date back to the 17th Century (see Benton & Joynt, 1960 for an informative review), but there was particular interest from surgeons in the 19th Century, most notably the Englishman John Hughlings Jackson (1866/1958; 1874/1958). His notes regarding the link between what he termed 'non-propositional language' and the right hemisphere complement those of his contemporaries, Broca and Wernicke, regarding the role of areas of the left hemisphere in the construction and comprehension of novel language. Subsequently, in various countries, much research was carried out in the immediate aftermath of the Crimean and First World Wars, when many young men received localised brain damage as a result of bullet wounds. Destruction of very precise, small areas of the brain in otherwise healthy people offered

N°4, 2013

ISSN 1760-4753

valuable insights into the possible relationship between specific locations and the particular function that had been lost—in many cases a linguistic one. Although it is, as it turns out, often not possible simply to attribute to a brain area the function that is lost if the area is damaged—for the brain is more sophisticated in its response to damage, and in its underlying management of processes—nevertheless a large amount of what we have come to understand about language function in the brain arises from such studies. Key discoveries were that a person could lose the areas of the brain associated with language generation yet still produce certain formulaic expressions; and damage to the right hemisphere could result in losing the ability to understand the holistic, often metaphorical or pragmatically loaded meaning of an expression, so that it was interpreted only literally, word by word. More generally, it was observed that, across many types of language disability, both developmental and acquired, formulaic language seemed to remain and to play an important role in facilitating continued communication. An excellent overview of the early observations of formulaic language in the clinical domain is given in Van Lancker (1987). Extensive discussion of what the findings might mean can be found in two chapters in Wray (2002b), while and some exploration of the latest research is undertaken in Wray (2008b).

More recently, research into how people with Alzheimer's cope in every day communicative situations has been increasingly of interest (Brewer, 2005; Bucks, Singh, Cuerden, & Wilcock, 2000; Davis, 2005, 2007; Davis, Russell-Pinson, & Smith, in preparation, 2008; Grainger, 2004; Hamilton, 1994; Orange, 2001; Sabat, Napolitano, & Fath, 2004), and formulaic language is a particular focus of study for many (Davis, 2006, 2007; Davis & Maclagan, 2010; Lindholm & Wray, 2011; Liu 2006; Moore & Davis, 2002; Wray, 2010, 2011). The assumption overall is that the characteristic repetitions observed in the language of people with Alzheimer's Disease are due to reduced memory capacity, which affects both their awareness of what they have already said, and also their ability to plan their utterances effectively.

2.3 Second language learning and idiomaticity

Language teachers and learners had separately been driven to consider the role of multiword strings in approximating nativelike knowledge and language behaviour. Phrase books have long been a favoured means by which people can 'manage' with another language (Wray, 2007). But for more methodical learners, the question has been, how useful is it to memorise phrases rather than single words and the rules that combine them? The piecemeal construction of utterances enables someone to express a greater range of different ideas. On the other hand, some have argued that one can only become idiomatic by knowing which of several possible ways to assemble words grammatically is preferred by native speakers. Over the past few decades the general fashion in language teaching in the western industrialised countries has been decidedly away from repeating and memorising, and towards attempting to communicate with whatever you can work out for yourself. An unintended consequence of this pursuit of expressive freedom is that ideas tend to be expressed unidiomatically—that is, the learner attempts to say things without any idea how a native speaker would say them.

Some materials writers concerned about how learners can be supported in achieving idiomaticity have found it useful to consider phrases and common collocations as a kind of complicated word, so that they can be incorporated within vocabulary learning—see for instance the discussion in

N°4, 2013

ISSN 1760-4753

Wray (2000a). This seems to reflect the acceptance that memorising vocabulary is an acceptable aspect of language learning in the west, while, for many teachers and learners, internalising longer strings is not.

2.4 Drawing things together

It was this juxtaposition of three different lines of independent observation—four if first language acquisition is treated separately—that drew my attention to the puzzle of formulaic language in the mid-1990s. I set about trying to solve the mystery of why young children seemed to find formulaic expressions so easy to handle, they were so resilient in language disorders, and they were so attractive to language teachers and learners that they were often used in the very first stages of courses—and yet they could be the most difficult obstacle in the later stages of second language learning (Wray, 2002b, pp. ix-x). Drawing on a wealth of individual studies, and a small number of reviews of them, particularly those of Van Lancker (1987) and Weinert (1995), I focussed my attention on looking for patterns across the different types of evidence, so as to develop an explanation of formulaic language that was coherent across these different strands. The model that resulted (Wray, 2002b; Wray & Perkins, 2000) created opportunities to ask new questions, including ones associated with the evolutionary origins of language (Wray, 1998, 2000b, 2002a; Wray & Grace, 2007) and language teaching and learning (Wray, 2000a, 2007, 2008d), and to explore the boundaries of the theory by examining a range of particular uses of language, such as machine translation (Wray, Cox, Lincoln, & Tryggvason, 2004), computer-supported communication in the disabled (Wray, 2002c), and experimental approaches to language learning (Fitzpatrick & Wray, 2006; Wray, 2004; Wray & Fitzpatrick, 2008, 2010)—see Wray (2008a) for a full exploration of these boundaries. At a practical level I have also paid some attention to how formulaic language can be defined and identified (Wray, 2002b, chap. 2, 3; 2008a, chap 8, 9; 2008c; Wray & Namba, 2003).

During the time that I have worked in this area it has transformed into a major, very productive domain. The Formulaic Language Research Network (FLaRN), founded in 2002, and now an electronic forum, e-FLaRN[2], now has nearly 200 members, a healthy mix of students and academic staff, who exchange information and gather every couple of years for a conference. I have been struck at these conferences by the fact that many of the key developments in research feature the imaginative use of existing or new paradigms and technologies for answering questions about formulaic language. For instance, researchers at Nottingham University used eye-tracking technology to establish whether formulaic language is read differently from non-formulaic language (Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011, Siyanova-Chanturia, Conklin & van Heuven, 2011). A team at the University of Alberta, Canada, measured brain activity during the processing of formulaic language (Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). Besides this, there are ever more ways of interrogating corpora, and more and more corpora, from different languages, to interrogate. For one useful collection of research papers on formulaic language in learning and use, see Wood (2010).

2 For details see http://www.cardiff.ac.uk/encap/research/networks/flarn/index.html

N°4, 2013

ISSN 1760-4753

As noted earlier, a typical characteristic of formulaic language research at present is that it tests central theoretical claims. This testing is a crucial stage in the cycle. The theory, developed on the basis of observations from studies that usually were not specifically set up to demonstrate those claims, must be tested robustly using studies designed expressly to show whether or not the claims are true. However, it is easy for research in the testing phase to not be all that productive, because it simply replicates aspects of the earlier research that formed the basis of the original theory—the process is circular. For instance, the observation that common idiomatic phrases seem easier to recall than non-idiomatic ones was part of what fed into the theoretical proposal that the idiomatic ones might be easier and quicker to process. There is then only limited value in testing the theoretical proposal by running an experiment to see if common idiomatic phrases are easier to recall than non-idiomatic ones.

The way out of that circularity is to be familiar with both the theory and the reasoning that led to its development. With regard to the claim that formulaic language is stored and processed holistically, for example, it is important to realise the provenance of that claim, and that it is contingent rather than absolute. It would be dangerous to suppose that there are many wordstrings that can be identified as definitely and always formulaic for everyone and in every circumstance. Rather, many have found it more convincing to suppose that, in one way or another, there are choices about how processing is done—that is, dual routes (Sinclair, 1991; Van Lancker, 1987; Van Lancker Sidtis, 2009; Wray, 1992). Anticipating that participants in an investigation might be exercising a choice about how they process the input given to them could substantially affect how the findings are interpreted.

In similar vein, when researcher/practitioners ask ‘how can I teach my students more idiomatic L2 English?’, the answer will be severely limited unless they have at least some idea about the answers to questions such as ‘what is idiomatic English?’, ‘how robust is it relative to other kinds of language?’, ‘what are native speakers doing when they speak idiomatically?’, ‘how did they learn to be idiomatic?’ and ‘what sort of psychological and neurological processes are entailed in speaking idiomatically?’

All of this requires researchers to explore the research literature critically (Wallace & Wray, 2006), by thinking about why findings and claims sometimes contradict each other, and how apparent contradictions might fit together in a broader picture. In this way, the testing of existing claims will be infused with more fundamental challenges about the underlying assumptions upon which the claims are based (Wray, forthcoming), resulting in competing models of the nature of formulaicity, and new understandings of how we process and use language.

3.0 Six current claims and opportunities for development

The following observations seem key to our current understanding of the nature and cause of formulaicity in the language of native speakers. An expanded version, with evidence, can be found in Wray (2002b). No claim made here, or in the subsequent sections, need be seen as unquestionable, and the alert reader will spot places where an *if* construction points to an

N°4, 2013

ISSN 1760-4753

opportunity to examine the validity of the premise upon which the point is built. For each claim, I offer some observations about the implications for L2 research and teaching.

3.1 Variation in idiomaticity

Claim: Central to achieving idiomaticity, and thus sounding nativelike, is knowing which of the various different ways to express a given idea is not only grammatical but also usual for that speech community (Pawley & Syder, 1983b).

Expansion: Idiomaticity is localised, both geographically and in terms of social and demographic groups, so that one cannot truly learn *the* nativelike way to express oneself in a language. Even within a small country like Britain, native speakers moving to another geographical area will discover differences in the form and meaning of formulaic and idiomatic expressions. For example, in the south of England the expression *I'm starving* means 'I'm hungry', but in the north of England it can mean 'I'm cold'. Moving to another English-speaking country also requires adjustments for a native speaker of English. Meanwhile, parents and grandparents find they are unfamiliar with the formulaic expressions of teenagers. And within different professions there are expressions that are often viewed disparagingly as jargon by outsiders, but which are a practical means to achieve effective communication within the group (for examples see Kuiper, 2009; for a discussion of the dynamics of jargon use in social groups see Wray & Grace, 2007).

Implications for L2 research and teaching: For practical reasons, classroom teaching in the L1 context is often focussed around idiomaticity in some notional target variety. It is understood that when learners go to the L2 country, they will actually encounter substantial differences. Does this matter? Can and should anything be done about it?

3.2 Social causes of idiomaticity

Claim: Native speakers learn to be idiomatic in relation to the groups with which they most strongly identify by virtue of observing and imitating the language of the group. The motivation for imitating the group is social and deep-rooted in human survival instincts (Wray & Grace, 2007).

Expansion: In first language acquisition children adopt their carers' linguistic patterns, though later they may identify more strongly with other groups and alter their language accordingly. We want and need to belong to groups, since our individual welfare relies on cooperation from others. We signal unity with others by various means, including how we dress, walk and behave, as well as how we speak. We also identify as outsiders anyone who does not speak or behave like our group.

Implications for L2 research and teaching: L2 learners are also L1 speakers. What assumptions, strategies and expectations about the role and use of language are transferred from a learner's L1 experience to their L2, and can such transfers be exploited? Who is the L2 learner identifying with? Does the L2 learner perceive language as one of many tools for achieving necessary

N°4, 2013

ISSN 1760-4753

functions in the real world, or as an object of study? What difference might this make to the process of learning?

3.3 Idiomaticity as a consequence of processing

Claim: The reason why native speakers express ideas in the same way as each other is that they store complex frames that encode the core meaning, and only have to slot the appropriate referential vocabulary in, to create the specific message they require. This reduces the likelihood of their using an alternative construction with the same meaning.

Expansion: My own approach to explaining formulaic language is that it is a consequence of only breaking down input when necessary (Needs Only Analysis), in order to minimise processing. The effect of this is largely the same as in emergent grammar models, including Construction Grammar (Goldberg, 2006). In Construction Grammar, language is seen as composed out of units that can range in size and complexity from morphemes up to complete sentences. The model resists assuming that the potential for small, simple units like morphemes undermines the usefulness of the larger, more complex units. Rather, these larger units have their own status as building blocks in the language. The difference between the two approaches is that Construction Grammar is anchored in the principle of patterns emerging out of frequency of use. While Needs Only Analysis does not deny that frequency plays a role, it does not view that role as primary.

Implications for L2 research and teaching: Both approaches define language as made up of chunks of different sizes. They therefore both also imply that teachers might reasonably experiment with supporting learners in working with larger as well as smaller chunks. That is, learners may benefit from paying attention to complete phrase, clause and sentence constructions, and even memorising them, as a means of modelling other sentences in the future (e.g. Ding, 2007).

3.4 Whose processing is being reduced?

Claim: Formulaic language is able to reduce not only the amount of processing entailed in production but also in comprehension (Wray, 2002b, chap. 5).

Expansion: The retrieval of larger, preformed, chunks of language is generally accepted to reduce the amount of processing during production. However, if language is a means for achieving interactional goals with others, it would be pointless for speakers to reduce their processing effort unless the hearer gained at least equal benefit. That is, if speakers' reduced processing was at the expense of *more* processing by hearers, then speakers would be trading their own convenience against the risk that the hearer did not decode their message successfully. The more processing involved, and the more separate units within the form, the greater the opportunity for the hearer to interpret the input differently than intended. Assuming that the speaker's primary aim is to deliver a message, it would make more sense to focus on reducing the *hearer's* processing, even if it cost the speaker more effort. The best kind of input, from the speaker's point of view, would be a complete message for which the meaning could be accessed on one go—a formulaic

N°4, 2013

ISSN 1760-4753

sequence. The well-acknowledged tendency of speakers to accommodate to hearers (Giles & Coupland, 1991) indicates that speakers are both sensitive to, and inclined to emulate, the speech patterns of those around them. A plausible explanation for why we pick up the formulaic expressions of others is that it is the way for us to get what we want.

Implications for L2 research and teaching: L2 learners are usually on the back foot—they do not feel powerful in relation to native speakers around them, and may not naturally have the confidence simply to imitate the patterns they hear. Yet doing so may be the fast track to both succeeding in interaction and becoming idiomatic. The road, though, is bumpy, because it may entail using structures before you are entirely sure what they mean. Research shows that young L2-learning children are bolder in this regard than adults (Wong Fillmore, 1976, 1979; Wray, 2008d). How can learners be encouraged to take the kinds of risks that lead to idiomatic language knowledge?

Meanwhile, it is part of the natural inclination of native speakers to adjust their own language towards that which they perceive the L2 learner to understand. Although in the short term this strategy can assist learners, it also deprives them of the most idiomatic kind of input, since that is the material first excluded, being least penetrable (Wray & Grace, 2007). Will it be helpful or harmful to a learner deliberately to memorise some quantity of idiomatic language in the L2, so as to give native speakers confidence that they can get their messages across? Although there are short term risks when you signal greater competence than you really have, the long term benefits could be a much greater opportunity to observe and imitate the most idiomatic input from native speakers.

A third consideration is that non-native speakers will presumably also accommodate towards each other. Care, therefore, must be taken in using too much unsupervised groupwork. It is not guaranteed that the less able will learn from the more able, because the prevailing dynamic determining accommodation is not necessarily going to be who is better at the language. It could be personality, urgency of message, or something else. How can the best outcomes from groupwork be achieved?

3.5 Where do collocations fit in?

Claim: Collocation research presents particularly interesting challenges and opportunities in relation to formulaic language, because the associations between collocates are often loose and variable.

Expansion: There is a busy agenda of research into collocation, drawing on the huge resources of major corpora and the software that interrogates them. One might say that there are at least three different types of researcher in this domain. Firstly, there are those whose primary interest is the statistical patterns of word distribution in language, and who may or may not interpret those patterns in relation to the uses of language. Secondly there are those who seek evidence from corpora to test theoretical models of what language is like. Thirdly, there is the very broad group of researchers who exploit corpus tools to examine particular words or sets of words.

N°4, 2013

ISSN 1760-4753

Formulaic language research, coming from the other direction—asking how the meanings and functions of messages tend to command particular associations of words—engages with corpus research in interesting ways. On the one hand, corpus linguistics challenges formulaic language research to clarify the relationship between formulaic language and collocations. While multiword strings with a clearly defined non-literal meaning can be fairly easily envisaged as holistic in memory, some of the more literal associations between two ordinary words that often occur together are more difficult to imagine that way. This is because the attraction between these words is not sufficient to exclude other pairings. For example, *happy event*, a typical British English way to refer to a forthcoming birth, is transparent in meaning other than in relation to its accustomed association, and neither *happy* nor *event* is precluded from pairing with many other words. Mutual information (MI) values, often used as a means for identifying genuine collocations, may not be useful in such cases, since neither word strongly predicts the other.

Meanwhile, formulaic language researchers challenge corpus linguists to justify their choices about which associations of words to focus on. Central to this issue is the role of frequency, which usually plays a central role in collocation research, but which is not always viewed as a primary determiner of formulaicity. That is, while there is undoubtedly some kind of relationship between frequency and formulaicity (Wray, 2002b, pp. 25, 31) it does not necessarily entail that the most frequent associations are the most formulaic. A subset of research in corpus linguistics homes in on what can be, to others, rather uninteresting configurations, particularly where they involve frequent function words.

Implications for L2 research and teaching: Materials writers are inevitably interested in discovering which kinds of pattern should be featured in teaching, and it may seem obvious that one should begin with the most frequently occurring associations. However, this is not necessarily quite true, for at least two reasons. One is that the small core of frequent material will not necessarily furnish one with enough language to cope with certain situations—there is a long tail of less frequent language that covers a wide number of situations and topics important for linguistic competence. The second reason is that frequency is a product of two separate mechanisms: how often something needs to be said and how often, when it is said, it is said in a given way. A word or wordstring could be very infrequent in the raw sense, yet be the predominant way of expressing that idea—definitely needed when that idea is expressed. The term *duck billed platypus* is rather infrequent wordstring in English (there are 9 occurrences in the 100m word BNC and 3 in the 385m Corpus of Contemporary American English). However, when you want to name that particular Australian marsupial, no other word will really do.

3.6 Variation according to genre

Claim: Patterns of formulaic language, like those of vocabulary more generally and also of grammar, will vary according to genre and medium (e.g. Kuiper, 2009).

Expansion: If our linguistic choices are determined by an attempt to deliver messages to hearers and readers as effectively as possible, we must expect that those choices will vary according to whom we are addressing, why, and in what communicative context. Part of the mastery of genre

N°4, 2013

ISSN 1760-4753

is fine-tuning the selection of formulations, either to be directly more effective (e.g. using shorter sentences in one genre than another) or else to be perceived as appropriate according to culture and custom. Some of the features of language characteristic of writing rather than speech are a reflection of the capacity for writer and reader to review the text without the constraints of limited short term memory that help determine the choices in speech (Wray, 2008a, chap. 4, 5), while others may simply be cultural norms. Genre commands many subtle dynamics, and native speakers rarely learn full proficiency in all of them, particularly those associated with aspects of performance, such as creative writing, oral story telling, etc.

Implications for L2 research and teaching: Learners encounter and ultimately may need to operate within a number of different genres. Exposure to one genre will not automatically provide information about others. One cannot expect a learner to know how to have a chat with someone, if they have only read discursive essays. Research evidence reveals that learners of even an advanced proficiency can fall down on subtle aspects of genre (e.g. Wiktorsson, 2003).

4.0 Conclusion

As noted at the beginning of this paper, no two people's overviews of research will be the same, and certainly no two people will have quite the same vision for the future. This, then, has been just one perspective. However, I certainly share with many others the optimism that formulaic language will occupy a more and more central place in research across linguistics, provided we can make explicit why its role should be seriously considered in studies of lexis, grammar, collocation, processing and interaction (Wray, 2008a, chap. 6, 7).

References

Benton, A. L., & Joynt, R. J. (1960). Early descriptions of aphasia. *Archives of Neurology, 3*, 109-126/ 205-222.

Brewer, J. P. (2005). Carousel conversation: aspects of family roles and topic shift in Alzheimer's talk. In B. H. Davis (Ed.), *Alzheimer talk, text and context: enhancing communication* (pp. 87-101). Basingstoke: Palgrave Macmillan.

Bucks, R. S., Singh, S., Cuerden, J. M., & Wilcock, G. K. (2000). Analysis of spontaneous, conversational speech in dementia of Alzheimer type: evaluation of an objective technique for analysing lexical performance. *Aphasiology, 14* (1), 71-91.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 29* (1), 72-89.

Davis, B. H. (2005). Introduction: some commonalities. In B. H. Davis (Ed.), *Alzheimer talk, text and context: enhancing communication* (pp. xi-xxi). Basingstoke: Palgrave Macmillan.

Davis, B. H. (2006). *Culturally Competent Materials on Communication and Dementia*. Charlotte: University of North Carolina, http://www.english.uncc.edu/bdavis/AlzAssnYear1report.pdf.

N°4, 2013

ISSN 1760-4753

Davis, B. H. (2007). *Culturally Competent Materials on Communication and Dementia: Year Two*. Charlotte: University of North Carolina, http://www.english.uncc.edu/bdavis/AlzGrtRepYr2-DAVISDS.pdf.

Davis, B.H. & Maclagan, M. (2010). Pauses, fillers, placeholders and formulaicity in Alzheimer's discourse: gluing relationship as impairment increases. In N. Amiridze, B.H. Davis, & M. Maclagan (eds). *Fillers, pauses and placeholders* (pp.189-215). Amsterdam: John Benjamins.

Davis, B. H., Russell-Pinson, L., & Smith, M. (2008). Collaboration in ESP: towards a multilingual, multicultural nursing assistant workforce. ESP across cultures 5: 7-22.

Ding, Y. (2007). Text memorization and imitation: the practices of successful Chinese learners of English. *System, 35*, 271-280.

Fitzpatrick, T., & Wray, A. (2006). Breaking up is not so hard to do: individual differences in L2 memorization. *Canadian Modern Language Review, 63* (1), 35-57.

Giles, H., & Coupland, N. (1991). *Language: contexts and consequences*. Milton Keynes: Open University Press.

Grainger, K. (2004). Communication and the institutionalized elderly. In J. F. Nussbaum & J. Coupland (Eds.), *Handbook of communication and aging research* (pp. 479-497). Mahwah, NJ: Lawrence Erlbaum.

Hamilton, H. E. (1994). *Conversations with an Alzheimer's patient*. Cambridge: Cambridge University Press.

Hughlings Jackson, J. (1866/1958). Notes on the physiology and pathology of language. In J. Taylor (Ed.), *Selected writings of John Hughlings Jackson, vol.2* (pp. 121-128). London: Staples Press.

Hughlings Jackson, J. (1874/1958). On the nature of the duality of the brain. In J. Taylor (Ed.), *Selected writings of John Hughlings Jackson, vol.2* (pp. 129-145). London: Staples Press.

Kuiper, K. (2009). *Formulaic genres*. London: Palgrave MacMillan.

Lindholm, C., & Wray, A. (2011). Proverbs and formulaic sequences in the language of elderly people with dementia. *Dementia: The International Journal of Social Research and Practice.* 10 (4): 603-623.

Liu, H. (2006). *Investigation of Situated Discourse Abilities and Disabilities for Chinese Alzheimer's Patients in Comparison with Normal Aging. ??????????????????????*. Beijing Foreign Studies University, Beijing.

Moore, L. A., & Davis, B. H. (2002). Quilting narrative: using repetition techniques to help elderly communicators. *Geriatric Nursing, 23* (5), 262-266.

Orange, J. B. (2001). Family caregivers, communication and Alzheimer's Disease. In M. L. Hummert & J. F. Nussbaum (Eds.), *Aging, communication and health* (pp. 225-248). Mahwah, NJ: Lawrence Elbaum.

Pawley, A. (2007). Developments in the study of formulaic language since 1970: a personal view. In P. Skandera (Ed.), *Phraseology and Culture in English* (pp. 3-34). Berlin: Mouton de Gruyter.

Pawley, A., & Syder, F. H. (1983a). Natural selection in syntax: notes on adaptive variation and change in vernacular and literary grammar. *Journal of Pragmatics, 7*, 551-579.

Pawley, A., & Syder, F. H. (1983b). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication.* (pp. 191-226). New York: Longman.

N°4, 2013

ISSN 1760-4753

Sabat, S. R., Napolitano, L., & Fath, H. (2004). Barriers to the construction of a valued social identity: a case study of Alzheimer's disease. *American Journal of Alzheimer's Disease and Other Dementias, 19* (3), 177-185.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research* 27 (2): 251-272.

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 37 (3), 776-784.

Tremblay, A., & Baayen, H. R. (2010). Holistic processing of regular four-word sequences: a behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D.Wood (Ed.), *Perspectives on formulaic language in acquisition and communication* (pp.151-173). London & New York: Continuum.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (Forthcoming). Processing advantages of lexical bundles: evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61 (2): 569-613.

Van Lancker, D. (1987). Nonpropositional speech: neurolinguistic studies. In A. W. Ellis (Ed.), *Progress in the psychology of language, vol 3* (pp. 49-118). Hillsdale, NJ: Lawrence Erlbaum.

Van Lancker Sidtis, D. (2009). Formulaic and novel language in a 'dual process' model of language competence: evidence from surveys, speech samples, and schemata. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.), *Formulaic Language, Volume 2: Acquisition, loss, psychological reality, and functional explanations* (pp. 445-470). Amsterdam: John Benjamins.

Wallace, M., & Wray, A. (2006). *Critical reading and writing for postgraduates*. London: Sage.

Weinert, R. (1995). The role of formulaic language in second language acquisition. *Applied Linguistics, 16* (2), 180-205.

Wiktorsson, M. (2003). *Learning idiomaticity: a corpus-based study of idiomatic expressions in learners' written production*. Lund Studies in English, vol 105. Lund: Lund University.

Wong Fillmore, L. (1976). *The second time around: cognitive and social strategies in second language acquisition.*, Stanford University, Palo Alto, CA.

Wong Fillmore, L. (1979). Individual differences in second language acquisition. In F. C. J., D. Kempler & S.-Y. W. Wang (Eds.), *Individual differences in language ability and language behavior.* (pp. 203-228). New York: Academic Press.

Wood, D. (Ed.). (2010). *Perspectives on formulaic language in acquisition and communication*. London & New York: Continuum.

Wray, A. (1992). *The focusing hypothesis: the theory of left-lateralised language reexamined*. Amsterdam: John Benjamins.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & Communication, 18*, 47-67.

Wray, A. (2000a). Formulaic sequences in second language teaching: principle and practice. *Applied Linguistics, 21* (4), 463-489.

Wray, A. (2000b). Holistic utterances in protolanguage: the link from primates to humans. In C. Knight, M. Studdert-Kennedy & J. Hurford (Eds.), *The Evolutionary Emergence of*

N°4, 2013

ISSN 1760-4753

Language: Social function and the origins of linguistic form (pp. 285-302). New York: Cambridge University Press.

Wray, A. (2002a). Dual processing in protolanguage: competence without performance. In A. Wray (Ed.), *The transition to language* (pp. 113-137). Oxford: Oxford University Press.

Wray, A. (2002b). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2002c). Formulaic language in computer-supported communication: theory meets reality. *Language Awareness, 11* (2), 114-131.

Wray, A. (2004). 'Here's one I prepared earlier': formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic sequences: acquisition, processing and use* (pp. 249-268). Amsterdam: John Benjamins.

Wray, A. (2007). Set phrases in second language acquisition. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Norrick (Eds.), *Phraseology: an international handbook of contemporary research* (Vol. 2, pp. 870-881). Berlin: Walter De Gruyter.

Wray, A. (2008a). *Formulaic language: pushing the boundaries*. Oxford: Oxford University Press.

Wray, A. (2008b). Formulaic sequences and language disorders. In M. Ball, M. Perkins, N. Müller & S. Howard (Eds.), *Handbook of Clinical Linguistics* (pp. 184-197). Oxford: Blackwell.

Wray, A. (2008c). Identifying formulaic language: persistent challenges and new opportunities. In R. Corrigan, E. Moravcsik, M. Ouali & K. Wheatley (Eds.), *tbc*. Amsterdam: John Benjamins.

Wray, A. (2008d). The puzzle of language learning: from child's play to 'linguaphobia'. *Language Teaching, 41* (2), 255-273.

Wray, A. (2010). We've had a wonderful, wonderful thing: formulaic interaction when an expert has dementia. *Dementia: The International Journal of Social Research and Practice* 9 (4): 517-34.

Wray, A. (2011). Formulaic language as a barrier to effective communication with people with Alzheimer's Disease. *Canadian Modern Languages Review* 67(4): 429-458.

Wray, A. (forthcoming). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*

Wray, A., Cox, S., Lincoln, M., & Tryggvason, J. (2004). A formulaic approach to translation at the post office: reading the signs. *Language and Communication, 24*, 59-75.

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in language learning and teaching* (pp. 123-148). Amsterdam: John Benjamins.

Wray, A., & Fitzpatrick, T. (2010). Pushing speakers to the extreme: the artificial use of prefabricated material in conversation. *Innovation in Language Learning and Teaching* 4 (1): 37-51

Wray, A., & Grace, G. W. (2007). The consequences of talking to strangers: sociocultural influences on the lexical unit *Lingua, 117* (3), 543-578.

Wray, A., & Namba, K. (2003). Formulaic language in a Japanese-English bilingual child: a practical approach to data analysis. *Japan Journal for Multilingualism and Multiculturalism, 9* (1), 24-51.

N°4, 2013

ISSN 1760-4753

Wray, A., & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, *20* (1), 1-28.

Biographical information

Alison Wray read linguistics with German and Hindi at the University of York, UK, and wrote her PhD on left hemisphere lateralised language. After a three year research post in the Department of Music at York, she took a post as a lecturer at the College of Ripon and York St John in York. In 1996 she became Assistant Director of the Wales Applied Language Research Unit in Swansea. In 1999 she was made a senior research fellow at Cardiff University, and subsequently worked there as a senior lecturer and Reader, becoming a Professor in 2005 and a Research Professor in 2007.

Notice biographique

Alison Wray a étudié la linguistique ainsi que l'allemand et le hindi à l'Université de York (RU), où elle a soutenu une thèse concernant la latéralisation du langage dans l'hémisphère gauche. Après un poste de recherche de trois ans dans le Département de musique à York, elle a accepté un poste comme *Lecturer* au *College of Ripon and York St. John* à York. En 1996 elle est devenue directrice adjointe de l'unité de recherche de linguistique appliquée du Pays de Galles à Swansea. En 1999 elle a été désignée *Senior Research Fellow* à l'Université de Cardiff, où elle a travaillé comme *Senior Lecturer* et *Reader*, avant de devenir *Professor* en 2005 et *Research Professor* en 2007.

N°4, 2013

ISSN 1760-4753