语言语块研究的理论与实证进展

Advances in Theoretical and Empirical Research on Formulaic Language

主　编　王立非

副主编　陈香兰

上海外语教育出版社
SHANGHAI FOREIGN LANGUAGE EDUCATION PRESS
www.sflep.com

语言语块研究的
理论与实证进展

Advances in Theoretical and Empirical Research
on Formulaic Language

主 编 王立非
副主编 陈香兰

前　　言

　　2009 年 5 月,首届全国语言语块教学与研究学术研讨会在北京召开,中国认知语言学研究会的近 30 名常务理事以及来自英、美、新加坡、我国港澳台地区等国内外 100 多所院校的近 200 位代表参加了会议。本次会议共收到 250 篇论文摘要,内容涉及语言语块理论、语块与认知、语块与语用、语块与二语习得、语块与语料库、语块与翻译、语块与口译/同传、语块与教学、语块与商务话语等,本次会议是一次推动我国语言语块研究和教学的盛会。

　　本论文集共分 2 个部分:1. 语块研究的理论视角;2. 语块研究的实证视角,收录了部分研讨会论文,并精选出近几年国内核心期刊发表的部分论文,共 16 篇,其中理论文章 9 篇,实证文章 7 篇。这些论文从 2 个角度探讨语块的本质特征,使我们从不同维度领略了语块的复杂性和精彩性。

　　诚然,语块研究在国内刚起步,还有许多问题亟待研究,如:1. 确定语块的界限和分类,以及语块在主流语言学理论中的位置;2. 丰富和完善语块理论,从认知心理、社会文化、计算语言学等多维视角认识语块的句法接口,认识其固化性和生成性并存的复杂本质,以及不同语言中的语块特性和共性;3. 探讨汉语语块的独特性,包括汉语高度语块化的构式特点,加强英汉语块的对比,探讨语块在二语学习中的有效性,包括在对外汉语教学中的应用;4. 加强语块的实证研究,包括多方法的实验研究。正如陆俭明教授所预言的那样,“构式和语块很可能改写汉语语法的历史”。希望本文集能激发更多学者关注语块,推动语块研究在中国的发展。
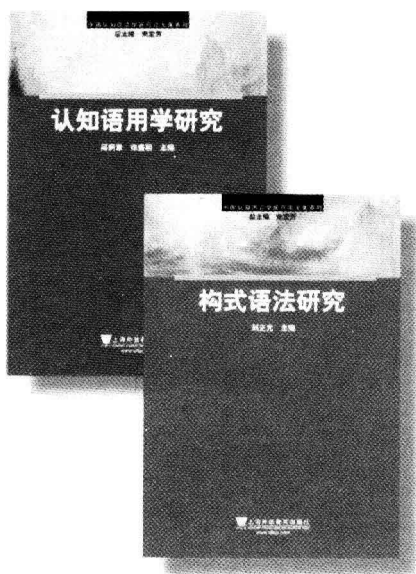
<div align="right">

编　者

2011.8.

</div>

中国认知语言学研究论文集

本论文集遴选了国内认知语言学领域发表的优秀学术论文，分为构式语法研究、认知语义学研究、认知语用学研究、隐语转喻研究以及心理空间理论和概念和成理论研究等若干文集，汇聚了认知语言学在国内的最新研究成果。

丛书包括

目　录

一、

语块研究的理论视角

Future Directions in Formulaic Language Research

Alison Wray

(Cardiff University, UK)

1. Introduction

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. My sense is that research into formulaic language is, in 2009, well on its way through the consolidation and confirmation part of its cycle. If I am correct, then the research that becomes significant in the near future will be that which careful considers what we have found out so far, and what it means for our understanding of the bigger picture of language knowledge and language learning. Probably several claims currently being taken for granted will be questioned, and new models of the nature of formulaic language may emerge. Certainly it is likely that there will be changes in

focus, as emerging patterns suggest new opportunities for investigation.

We are, then, at a potentially 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 middle part of this decade the majority of the work on formulaic language that I have read, particularly in the domain of second language learning, has been of this kind. Such work has been 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, once we have a reasonable volume of empirical evidence, it is time to develop the ideas. 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 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 particularly claim or assumption in one domain can really be true, given what we know about another domain. Do the claims about formulaic language being stored and retrieved holistically correspond with findings from psycholinguistic research into language retrieval, for instance, and if not, why not? How do the claims about the difficulties of language learning as an adult relate to claims in the pop-literature that cognitive decline in old age can be staved off by learning another language (Weil 2005, p. 225)? To what extent to do the patterns of collocation revealed in corpus linguistic research challenge the easy "formulaic/non-formulaic" categorisation used in some experiments?

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. Others would write a rather different story, and I shall refer the reader to one or two such stories so that a more rounded perspective can be gained. In section 3 I identify five 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. Section

4 considers very specifically the opportunities for research into formulaic language in China, where I see huge potential for answering certain questions that may not be so easily answered elsewhere.

2. The emergence and consolidation of formulaic language 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.

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 same tension was under evaluation in relation to first language acquisition, where it was evident that children could not be learning *only* formulaically, but also could not be learning without *any* reliance on multiword strings. A very significant study was that of Ann Peters (1983), who chronicled the language development of a child who did not follow the word-by-word learning pattern described in earlier research, such as Brown (1973). As a result of that study and research by Nelson (e.g. 1981) into styles of interaction, it became evident that there is more than one path to effective language acquisition, and that children vary along a continuum from very word-based to much more phrase-based in the unit preference. Nelson's work suggested that the path they took was at least partly determined by the way language was used by their carers.

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 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).

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 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 focused 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, forthcoming) — 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