CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

PART I

WHAT FORMULAIC SEQUENCES ARE

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

1

The Whole and the Parts

'Twelve-inches-one-foot. Three-feet-make-a-yard. Fourteen-pounds-make-a-stone. Eight-stone-a-hundred-weight'. . . . Unhearing, unquestioning, we rocked to our chanting, hammering the gold nails home. 'Twice-two-are-four. One-God-is-Love. One-Lord-is-King. One-King-is-George. One-George-is-Fifth . . .' So it was always; had been, would be for ever; we asked no questions; we didn't hear what we said; yet neither did we ever forget it.

Laurie Lee: *Cider with Rosie*. Penguin:53–4

She would go and smile and be nice and say 'So kind of you. I'm so pleased. One is so glad to know people like one's books'. All the stale old things. Rather as you put a hand into a box and took out some useful words already strung together like a necklace of beads.

Agatha Christie: *Elephants Can Remember*. Pan:12

Introduction

In a series of advertisements run on British TV early in 1993 by the breakfast cereal manufacturer Kellogg, people were asked what they thought Rice Krispies were made of, and expressed surprise at discovering that the answer was rice.[1] Somehow they had internalized this household brand name without ever analyzing it into its component parts. It was as if the name of the product had taken on a life of its own, and required no more reference back to its 'meaning' than do words of foreign origin such as *chop suey* ('mixed bits') and *spaghetti* ('little cords'). But how could this come about in the case of a name which, although oddly spelled, so transparently refers to crisp rice? In actual fact, overlooking the internal composition of names is a far more common phenomenon than we might at first think. Many personal names have 'meanings' which we simply ignore: we do not expect someone called 'Verity Baker' to be a truthful bread maker, or someone called

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

'Victor Cooper' to win barrel-making competitions.[2] Since interpreting such names in a literal way would be a distraction, it is actually very useful that we can choose the level at which we stop breaking down a chunk of language into its constituent parts. Nor is it just names that we treat in this way. We also overlook the internal composition of a great many words. Although there is a historical reason why a ladybird is so called, there is no more sense in decomposing the word than there is in falsely breaking down *carpet* into 'car' and 'pet'.

If this phenomenon were restricted to proper names and single words, it would be remarkable enough. But this is just the thin end of the wedge, for we are also able to treat entire phrases, clauses, and even lengthy passages of prose in this way. Just as with the name *Rice Krispies*, which, in effect, means both 'crisp rice' and 'common breakfast cereal of indeterminate composition', the result is often, though not always, two layers of meaning. If you break the phrase up, it means one thing, but if you treat it whole, in its accustomed way, it possesses a meaning that is something other than, or in addition to, its constituent parts. Idioms are a clear example of this. The phrase *pull someone's leg* has a literal, if rather improbable, meaning, which involves a person, the person's leg, and the action of pulling. But the phrase as a whole has the meaning 'tease', and it is difficult, in that interpretation, to work out why there is any reference to legs or pulling at all.

Words and word strings which appear to be processed without recourse to their lowest level of composition are termed *formulaic*, and they are the focus of this book. They are interesting because their widespread existence is an embarrassment for certain modern theories of linguistics, which have unashamedly pushed them aside and denied their undoubted significance. In exploring the way in which formulaicity contributes to our management of linguistic communication, we shall address such questions as: Just how common is formulaic language? What forms can it take? What is it used for? What role does it play in our production and comprehension of normal discourse? How is it to be accommodated within linguistic theory? How do first language learners acquire it? Why is it so problematic for second language learners? What happens to it when someone loses language capabilities through brain damage? And what role might it play in the general and linguistic recovery of such individuals?

In the course of the book, we shall see that research on formulaic language has lacked a clear and unified direction, and has been diverse in its methods and assumptions. Both within and across subfields such as child language, language pathology and applied linguistics, different terms have been used for the same thing, the same term for different things, and entirely different starting places have been taken for identi-

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

fying formulaic language within data. As a result, little headway has been made in spotting larger, more general patterns, and no attempt has been made before, to compare and contrast the full range of findings and to reconcile them within a single theoretical account.

The momentum of the book is towards a unified description and explanation of formulaic language and its status relative to the lexicon. Parts I and II culminate in the assertion that recognizing the role of formulaicity is fundamental to understanding the freedoms and constraints of language as a formal and functional system. Specifically, it is proposed that formulaic language is more than a static corpus of words and phrases which we have to learn in order to be fully linguistically competent. Rather, it is a dynamic response to the demands of language use, and, as such, will manifest differently as those demands vary from moment to moment and speaker to speaker. This hypothesis, developed with reference to what is known about formulaicity in the language of adult native speakers, is then tested through a comprehensive survey of findings in the published research of three other fields: first language acquisition (Part III), second language acquisition (Part IV) and aphasia (Part V). For each area of research, individual descriptive and explanatory models are developed, and these are drawn into a unified model in Part VI.

Setting the Scene

The Shape of Formulaicity

It is something of a joke amongst those who write for a living that it is possible to construct plausible text out of prefabricated chunks (Figure 1.1). The humour of such examples resides in our recognition that "just as we are creatures of habit in other aspects of our behavior, so apparently are we in the ways we come to use language" (Nattinger & DeCarrico 1992:1). Despite Pinker's (1994:90ff) assertion that using prefabricated chunks of language is a peripheral pursuit that tells us nothing about real language processing, there is plenty of evidence to the contrary. For, in our everyday language, "the patterning of words and phrases . . . manifests far less variability than could be predicted on the basis of grammar and lexicon alone" (Perkins 1999:55–56). There are words and phrases that we are likely to say when we see a particular friend, or find ourselves in a certain situation (Coulmas 1981). If we tell the same story, or deliver the same lecture, more than once, we will soon find that whole ideas are expressed in the same chunks of language each time (Peters 1983:80, 109). We may re-echo a form of words that we used earlier, or which someone else has just used (Pawley & Syder 2000:178). In the context of 'collocation' we find that some words seem to belong

© in this web service Cambridge University Press www.cambridge.org

To pad out a report or an essay which is short on words without having to do any original thinking simply take one phrase from each column and join them as a sentence.

I	II	III	IV
On the other hand	the realization of preset programme assignments	compels us to reanalyze thoroughly the forms	of existing administrative and financial conditions.
Similarly	the scope of staff schooling	requires the explicit formulation and definition	of further directions of development.
However, we must not forget that	permanent growth volume and the range of our activities	helps in the preparation and realization	of the system of universal participation.
The weight and meaning of these problems does not need justification as	the current structure of organizations	safeguards the involvement of a wider group in the forming	of participants' attitudes in the face of the tasks set by organizations.
Richly diversified experiences and	the new model of organizational activity	fulfils important tasks in the elaboration	of new propositions.
The concern of the organization, in particular	further development of various forces of activity	enables the creation, to some extent,	of the directions of progressive education.
Higher ideological assumptions, and also	permanent safety of our activities in information and propaganda	is causing appreciation of the scale of importance	of needs-related systems.
In this way	consultation with active members	presents an interesting verification test	of appropriate conditions of activation.
Broadly speaking,	any alternative approach to the questions	draws after itself the initiation and modernization processes	of the model's development.

Figure 1.1. Advice on using prefabricated chunks of text (origin unknown).

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE
Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

together in a phrase, while others, that should be equally good, sound odd. For instance, Biber, Conrad and Reppen (1998) report that, in a 2.7 million word corpus of academic prose, *large number* was more than five times more common than *great number* (48.3 per million versus 8.9 per million).[3]

Whether these preferred strings are actually stored and retrieved as a unit or simply constructed preferentially, it has been widely proposed that they are handled, effectively, like single "big words" (Ellis 1996:111). They are "single choices, even though they might appear to be analysable into segments" (Sinclair 1991:110). Some are fully fixed in form (e.g., *Fancy seeing you here; Nice to see you*) and can bypass the entire grammatical construction process (Bateson 1975:61). Others, termed *semi-preconstructed phrases*, such as NP_i *set* + TENSE $POSS_i$ *sights on (V)* NP_j, require the insertion of morphological detail and/or open class items, normally referential ones (giving, for instance, *The teacher had set his sights on promotion; I've set my sights on winning that cup*).

A Long-Recognized Phenomenon

Observations of unexpected levels of fixedness in language can be traced back to the mid-nineteenth-century writings of John Hughlings Jackson, whose interest was in the ability of aphasic patients fluently to utter rhymes, prayers, routine greetings and so on, even though they had no ability to construct novel utterances (see Chapter 12). Half a century later, Saussure (1916/1966) talked of "synthesizing the elements of [a] syntagm into a new unit . . . [such that] when a compound concept is expressed by a succession of very common significant units, the mind gives up analysis – it takes a short cut – and applies the concept to the whole cluster of signs, which then becomes a simple unit" (p. 177). Jespersen (1924/1976) observed that "a language would be a difficult thing to handle if its speakers had the burden imposed on them of remembering every little item separately" (p. 85). He characterized the *formula* as follows:

> [it] may be a whole sentence or a group of words, or it may be one word, or it may be only part of a word, – that is not important, but it must always be something which to the actual speech instinct is a unit which cannot be further analyzed or decomposed in the way a free combination can. (p. 88)

Bloomfield (1933) observed that "many forms lie on the border-line between bound forms and words, or between words and phrases" (p. 181). According to Firth (1937/1964), "when we speak . . . [we] use a whole sentence . . . the unit of actual speech is the holophrase" (p. 83). Firth

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

considered it central to characterizing communication within a speech community to identify and list the "usual collocations" (1957/1968:180ff). Hymes (1962/1968) proposed that "a vast portion of verbal behavior . . . consists of recurrent patterns, of linguistic routines . . . [including] the full range of utterances that acquire conventional significance for an individual, group or whole culture" (pp. 126–127). Bolinger (1976) asserted that "our language does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs" (p. 1), and Charles Fillmore (1979) argued that "a very large portion of a person's ability to get along in a language consists in the mastery of formulaic utterances" (p. 92).

However, insofar as these descriptions applied beyond the realm of the noncomponential idiom, they became increasingly marginalized as Chomsky's approach to syntactic structure gained prominence. Only with the new generation of grammatical theories, based on performance rather than competence (see later), has the idea of holistically managed chunks of language been slowly reinstated, and its implications recognized.

Terminology

Figure 1.2 lists some of the terms which can be found in the literature to describe a larger or smaller part of the set of related phenomena that we shall be examining in this book. While there is undoubtedly a certain measure of conceptual duplication, where several words are used to describe the same thing, it is also evident that some of the terms shared across different fields do not mean entirely the same thing in all instances. The label used by a given commentator may reflect anything from the careless appropriation of a nontechnical word to denote a specific meaning, to the deliberate selection of a particular technical term along with all its preexisting connotations. Overall, we must exercise some doubt about the likelihood that "while labels vary, it seems that researchers have very much the same phenomenon in mind" (Weinert 1995:182), for we shall see in Chapter 3 that this large and unwieldy set of types has been carved up and categorized in innumerable ways, all of which have something useful to say, but none of which seems fully to capture the essence of the wider whole.

Because of this plethora of terms, and the individual ways in which they are implicitly or explicitly defined,[4] we encounter a certain difficulty in wanting to refer to findings within and across research areas without appearing to impose one or another theoretical position. The survey which will unfold in the course of the book is intended to cast a fresh

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

amalgams – automatic – chunks – clichés – co-ordinate constructions – collocations – complex lexemes – composites – conventionalized forms – F[ixed] E[xpressions] including I[dioms] – fixed expressions – formulaic language – formulaic speech – formulas/formulae – fossilized forms – frozen metaphors – frozen phrases – gambits – gestalt – holistic – holophrases – idiomatic – idioms – irregular – lexical simplex – lexical(ized) phrases – lexicalized sentence stems – listemes – multiword items/units – multiword lexical phenomena – noncompositional – noncomputational – nonproductive – nonpropositional – petrifications – phrasemes – praxons – preassembled speech – precoded conventionalized routines – prefabricated routines and patterns – ready-made expressions – ready-made utterances – recurring utterances – rote – routine formulae – schemata – semipreconstructed phrases that constitute single choices – sentence builders – set phrases – stable and familiar expressions with specialized subsenses – stereotyped phrases – stereotypes – stock utterances – synthetic – unanalyzed chunks of speech – unanalyzed multiword chunks – units

Figure 1.2. Terms used to describe aspects of formulaicity.

eye over the range of accounts and data, in order to establish the larger pattern into which they all fit. What is needed, then, is a term which does not carry previous baggage, and which can be clearly defined. The neutral term *formulaic language* is too commonly used in the literature to be free of such associations. In its place, therefore, we shall use *formulaic sequence*.[5] The word *formulaic* carries with it some associations of 'unity' and of 'custom' and 'habit', while *sequence* indicates that there is more than one discernible internal unit, of whatever kind. As we shall see, there are good reasons for avoiding any implication that these internal units must be words. Our working definition of the *formulaic sequence* will be as follows:

a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.[6]

It is clear from this definition that the term aims to be as inclusive as possible, covering any kind of linguistic unit that has been considered formulaic in any research field.[7] The intention is to make reference easier, not to constrain the discussion, so, despite the features of the definition, the term will have to be used fairly loosely as a coverall. In particular, although our starting place is a recognition that there is

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

something about a formulaic sequence that makes it appear to be unitary, we shall also cover accounts which do not embrace, or do not require, holistic storage, including purely frequency-based descriptions. At times, especially in Chapters 11 and 13, we shall be focussing on the ability of morphemes and polymorphemic words to count as formulaic sequences. In such contexts, we shall need to differentiate between types of formulaic sequence and the terms *formulaic word string, formulaic word* and *morpheme* will be used.

Selecting a Theoretical Reference Point

"Linguists seem to underestimate the great capacity of the human mind to remember things while overestimating the extent to which humans process information by complex processes of calculation rather than by simply using prefabricated units from memory" (Lamb 1998:169). It will be proposed in this book that although we have tremendous capacity for grammatical processing, this is not our only, nor even our preferred, way of coping with language input and output. In particular, it will be argued that much of our entirely regular input and output is not processed analytically, even though it could be. Clearly, in order to explore this idea, it is necessary to engage with at least one established model of grammatical processing. Several recent models intrinsically accommodate some or all aspects of formulaicity, including Cognitive Grammar (e.g., Langacker 1987, 1991), Construction Grammar (e.g., Fillmore, Kay & O'Connor 1988; Michaelis & Lambrecht 1996; Tomasello & Brooks 1999), the Emergent Lexicon (Bybee 1998), Lexical-Functional Grammar (Bresnan 1982a, 1982b), the Cardiff Grammar, a version of Systemic Functional Grammar (e.g., Tucker 1998), and Pattern Grammar (Hunston & Francis 2000). However, to adopt one of these for our current purposes would be premature, since their very tolerance of formulaicity means that they will not challenge the core assumptions about its nature which this book seeks to tease out and examine. Rather, we need a model which directly opposes those assumptions, so that every claim about what formulaicity is and why it exists has to be fully justified.

The theoretical positions least sympathetic to formulaicity as a principle feature of language structure are the ones which propose a single grammatically based processing system. Within those, it is Chomsky's (1965) claim, that we have a greater understanding of language structure than we could possibly construe only from the observation of input, which remains the most difficult to defeat. Therefore, the argument of this book will be directed against the traditional generative account of syntax, in which language structure is founded on abstract universal and local rules. The Chomskian position offers the clearest contrast to the

© in this web service Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-77309-6 - Formulaic Language and the Lexicon
Alison Wray
Excerpt
More information

whole notion of circumstantial associations between words, is least tolerant of internally complex units, and holds itself separate from performance and pragmatics, the two axes of the model of formulaicity developed here. There is also a particular advantage in pitching what will be a model of part-analytic, part-holistic processing against a purely analytic one. It invites us to construe the analytic grammatical system and the holistic formulaic one as essentially separate. Now this may well not be desirable in the end, but it will make much clearer the path of the argument.

It would be short-sighted simply to ignore the alternative theories, however, especially since they may offer plausible solutions to the problem of how formulaicity is to be accommodated within a productive knowledge of grammar.[8] So we shall return to the question of theoretical models of grammatical structure and processing in Chapter 14, when we shall be able to assess more directly the demands that the existence of formulaicity makes on explanatory adequacy.

Formulaicity and Our Capacity for Novelty

The reason why formulaicity has been somewhat overlooked in the last few decades is that, from the standard perspective of how linguistic systems must be designed, it does not sit easily with our capacity for novel expression. Novelty in language, or rather the potential for it, has lain at the centre of modern linguistic theory for several decades: "an essential property of language is that it provides the means for expressing indefinitely many thoughts and for reacting appropriately in an indefinite range of new situations" (Chomsky 1965:6).

Chomsky's observations about our inherent capacities to generate and to understand sentences that we have never encountered before are fundamental and entirely valid. But the significance of this capability has been considerably overstated, relative to our actual use of language on a minute-by-minute basis:

> native speakers do *not* exercise the creative potential of syntactic rules to anything like their full extent, and . . . indeed, if they did so they would not be accepted as exhibiting nativelike control of the language. The fact is that only a small proportion of the total set of grammatical sentences are nativelike in form – in the sense of being readily acceptable to native informants as ordinary, natural forms of expression, in contrast to expressions that are grammatical but are judged to be 'unidiomatic', 'odd', or 'foreignisms'. (Pawley & Syder 1983:193)

In order to understand the significance of this fundamental misalignment of positions, we need to consider just what 'novelty' is in this context, and how it can be reconciled with the repetitiveness of much of our everyday language.

© in this web service Cambridge University Press www.cambridge.org