



PERGAMON

Language & Communication 20 (2000) 1–28

www.elsevier.com/locate/langcom

LANGUAGE
&
COMMUNICATION

The functions of formulaic language: an integrated model

Alison Wray^{a,*}, Michael R. Perkins^b

^aSenior Research Fellow, Centre for Language and Communication Research, Cardiff University,
PO Box 94, Cardiff CF10 3XB, UK

^bDepartment of Human Communication Sciences, University of Sheffield, Sheffield, S10 2TN, UK

1. Introduction

1.1. The nature of formulaic language

‘Formulaicity’ and ‘formulaic sequence’ will be used in this paper to describe, in a neutral way, a phenomenon that encompasses various types of wordstring which appear to be stored and retrieved whole from memory. Our working definition of the *formulaic sequence* will be:

a sequence, continuous or discontinuous, of words or other meaning 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.

This includes, at the one extreme, tightly idiomatic and immutable strings, such as *by and large*, which are both semantically opaque and syntactically irregular, and, at the other, transparent and flexible ones containing slots for open class items, like *NP be-TENSE sorry to keep-TENSE you waiting* (Pawley and Syder, 1983, p. 210). Perkins (in press) defines formulaicity as follows: “manifested in strings of linguistic items where the relation of each item to the rest is relatively fixed, and where the substitutability of one item by another of the same category is relatively constrained”. If we take formulaicity to encompass, as some do, also the enormous set of ‘simple’ lexical collocations, whose patterns are both remarkable and puzzling from a formal grammatical point of view (e.g. Sinclair, 1991), then possibly as much

* Corresponding author.

as 70% of our adult native language may be formulaic (Altenberg, 1990). A range of corpus studies (e.g. Kjellmer, 1984; Baayen and Lieber, 1991; Altenberg, 1993; Barkema, 1993) have shown that the patterning of words and phrases in ordinary language manifests far less variability than could be predicted on the basis of grammar and lexicon alone, and in fact most natural language, written or spoken, appears to consist largely of collocational ‘sets’ or ‘frameworks’ (Renouf and Sinclair, 1991; Renouf, 1992). As Sinclair (1991) puts it: “all the evidence points to an underlying rigidity of phraseology, despite a rich superficial variation” (p. 121). Formulaicity contrasts with productivity, the ability to use the structural system of language (syntax, semantics, morphology and phonology) in a combinatorial way to create novel utterances and in an analytical way to understand them (Perkins, in press). We explore the relationship between these in Section 2.

A particularly noteworthy feature of formulaicity is the variability found in the forms, functions and distributions of sequences across types of language (Wray, 1999). While some adult native forms can already be heard in the output of the young child or early-stage L2 learner, other formulaic sequences, that are different from anything in the adult native language, also appear during these acquisition processes, and then disappear again. Formal classroom L2 learners use some formulaic sequences that are native-like and others that are non-native-like. Items of both types may later disappear or may persist, and, despite the apparent ease with which they are adopted during learning, it is often the failure to use native-like formulaic sequences that ultimately marks out the advanced L2 learner as non-native (Pawley and Syder, 1983). Aphasics of various kinds also make use of formulaic sequences, some corresponding with the pre-trauma norm, but others idiosyncratic in their form and meaning.

Wray’s (1999) survey of descriptions of formulaic language in the speech and writing of adult native speakers, first and second language learners and aphasics demonstrates that its forms and functions are determined, within the output of the individual, by a complex interaction of factors. The best way to account for the patterns of distribution is in terms of a dynamic model, and in this paper we propose such a model. In Section 1 we shall explore the phenomenon of formulaicity in general terms, considering both the ways in which it has been characterised and the problems that have arisen from the under-specification of fundamental contrasts. Section 2 briefly summarises the findings of Wray’s (1999) survey and identifies the nature of the challenge to a processing model, especially the need to accommodate our capacity for novel utterances. We then present our model (Section 3), which is in three parts. The first two offer an account of the individual’s strategy choices at any point in time, within the framework of interactional and processing constraints respectively. The third part describes the developmental aspect of processing, that is, the role played by formulaic language in the L1 learner across time.

1.2. Clearing the ground

A considerable number of accounts of formulaic language exist, focussing on adult natives (e.g. Becker, 1975; Bolinger, 1976; Coulmas, 1979; Cowie, 1992,

1994a,b, 1998; Moon, 1998), first language learners (e.g. Peters, 1977, 1983; Garvey, 1977; Nelson, 1981; Bates et al., 1988), second language learners (e.g. Wong Fillmore, 1976; Ellis, 1994; Weinert, 1995; Granger, 1998; Howarth, 1998) and aphasics (Hughlings Jackson, 1874/1958; Code, 1987, 1997; Van Lancker, 1987; Van Lancker and Kempler, 1987); for a review of these and many others, see Wray (1999). While in the early accounts there was a tendency to look only inwards, largely failing to recognise the existence of the phenomenon beyond the bounds of the specific area under investigation, most later reports feature some measure of cross-referencing, aimed at demonstrating how the formulaic sequences in each area reflect a more general property of language. While this wider perspective is undoubtedly valid, the eagerness to contextualise in this way seems to have rather underestimated some basic problems with the looseness of the terminology, which makes it extremely difficult to be sure when like is being compared with like (see also Howarth, 1998, p. 25). Well over 40 terms have been used to refer to one or more type or subtype of formulaic language, including those in Table 1.

What is the significance of the existence of so many terms? Is it safe to assume that the only reason for the variation in terminology is that the phenomenon and its subtypes have been found and named independently in different fields? Conversely, dare we assume that where a single term is used by commentators in more than one field, it actually refers to exactly the same phenomenon? We think not. Rather, it seems that there are genuinely deep-seated and significant differences, which have become obscured by the tolerance of terminological variation on the one hand, and, on the other, the indiscriminate appropriation of certain favoured terms across data types. The multi-faceted nature of formulaic language is evident from the variety of ways in which it has been characterised: according to its form, function, semantic, syntactic and lexical properties, and its relationship with novel (analytic) language. We shall consider each in turn.

Table 1

Terms used in the literature to describe formulaic sequences and formulaicity

Amalgams	Gambits	Preassembled speech
Automatic	Gestalt	Prefabricated routines and patterns
Chunks	Holistic	Ready-made expressions
Clichés	Holophrases	Ready-made utterances
Co-ordinate constructions	Idiomatic	Rote
Collocations	Idioms	Routine formulae
Composites	Irregular	Schemata
Conventionalized forms	Lexical(ised) phrases	Semi-preconstructed phrases that constitute single choices
FEIs ^a	Lexicalised sentence stems	Sentence builders
Fixed expressions	Multiword units	Stable and familiar expressions with specialized subsenses
Formulaic language	Non-compositional	Synthetic
Formulaic speech	Non-computational	Unanalysed chunks of speech
Formulas/formulae	Non-productive	
Fossilized forms	Petrification	
Frozen phrases	Praxons	

^a Fixed Expressions including Idioms (Moon, 1998).

1.3. Form-based taxonomies

Many have offered descriptions and/or categorizations of formulaic sequences in adult native language, including Becker (1975), Bolinger (1976), Hatch et al. (1979) Coulmas (1979, 1994), Yorio (1980), Lattey (1986), Van Lancker (1987), Moon (1992, 1998) and Howarth (1998). Although by no means the most detailed, Becker's (1975) basic six category taxonomy of adult native speaker formulas is a useful reference point:

- **Polywords**, e.g. *(the) oldest profession; to blow up; for good*.
- **Phrasal constraints**, e.g. *by sheer coincidence*.
- **Meta-messages**, e.g. *for that matter...* (message: 'I just thought of a better way of making my point'); *...that's all* (message: 'don't get flustered').
- **Sentence builders** *(person A) gave (person B) a (long) song and dance about (a topic)*¹.
- **Situational utterances**, e.g. *how can I ever repay you?*
- **Verbatim texts**, e.g. *better late than never; How ya gonna keep 'em down on the farm?* (adapted from Becker, 1975, p. 6f).

Becker's categorisation, like Nattinger and DeCarrico's (1992) considerably more detailed one which draws heavily upon it, suffers from difficulties in teasing apart form and function. Furthermore, he does not focus directly on the potential of formulaic sequences to tolerate grammatical and semantic oddity, including the complex relationship they have with metaphor, and he fails to capture the possibility of placing them on a continuum from fixed to novel (see Section 1.6).

1.4. Semantic irregularity

In many, but not all, cases, formulaic sequences have relinquished their semantic compositional meaning in favour of a holistic one (Nattinger and DeCarrico, 1992, pp. 32–33), and in this respect they coincide in part with idioms and metaphors (Yorio, 1980; Moon, 1992). It is common for a formulaic sequence to carry a metaphorical meaning, and in some cases it would be impossible for a hearer to understand it for the first time without substantial pragmatic or direct explicational context (e.g. *straight from the horse's mouth; to pull someone's leg*) (see Gibbs, 1991). In other cases, the metaphorical meaning can be retrieved more directly (e.g. *the autumn of one's life; I can read you like a book*). Semantically opaque sequences have to be idioms, else they would become unusable, while poetry shows us that transparent metaphors need not be formulaic at all (e.g. *young death sits in a cafe smiling*—e.e. cummings). This variability in the transparency of sequences makes it superficially attractive to use \pm idiom as a defining variable in characterising formulaicity. However, Howarth (1998) demonstrates the usefulness of separating out this variable from the main structure of the definition (see below). The scope of formulaic sequences is,

¹ In fact Becker has conflated two quite important formal features in his *Sentence builders* category by illustrating the sentence template pattern, which can often be literally understood, with a metaphor. Metaphors are very common in formulaic sequences but there is no primary or specific association with any form-type (see Section 1.4).

however, much wider than idioms, as even the most basic taxonomies, such as Becker's (1975) above, indicate.

1.5. Syntactic irregularity

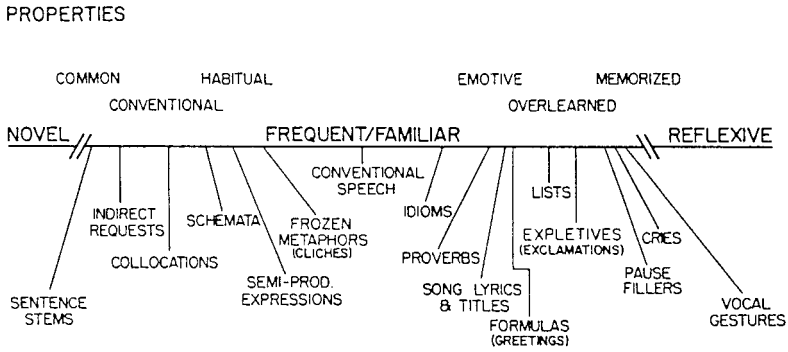
Syntactic irregularity comes in various forms. One is a restriction on the normal scope for inflexional or transformational manipulation (Verstraten, 1992). For example, it is not possible to pluralize *beat around the bush* or passivize *face the music* without the sequences losing their formulaic status (Flavell and Flavell, 1992, p. 6). Nor is it possible to say that you *slept a wink last night* or to make someone *fed up* by *feeding them up* (Irujo, 1986, p. 237). In another type of irregularity, normal restrictions may be flouted, as when a direct object appears with an intransitive verb, as in *to come a cropper*; *to go the whole hog* (Flavell and Flavell, 1992, p. 7), or non-identical constituents are co-ordinated, as with *by and large*. (Etymologically speaking, *by and large* is a nautical term for steering a boat “both before and behind...the beam” (Flavell and Flavell, 1992, p. 46), and in this context it is a co-ordination of two adverbs. However, it is rare for a user to know this and doubtful whether even one who does, actually computes it in this way).

Such oddities seem to be a natural, though not a necessary, consequence of a sequence becoming formulaic, as fossilization strands it from the normal criteria of acceptability in the mainstream language (Cowie, 1992, p. 2). This sort of irregularity is a mine-field for any account of language processing which prioritises constituent structure over message (see Wray, in preparation b). However, as with semantic irregularity, and for essentially the same reasons, it is restricted to only one part of the much larger set of formulaic sequences, the idioms, and, as such, has only limited value as a measure of the phenomenon as a whole.

1.6. Continua of formulaicity

Several have followed Bolinger (1976) in favouring a categorisation of formulaic sequences according to the extent of their fixedness. Bolinger quotes an early version of Van Lancker's (1987) continuum (Fig. 1), which teases out variation along the *analyticity-cohesion* dimension (Van Lancker, 1987, p. 55), according to the criterion of ‘amount of attention paid’. This leads to the uneasy combination of some form-based and some function-based categories. Howarth (1998) also favours a continuum description, and encounters the same necessity to identify some word combinations in terms of their function, even though his primary emphasis is form. His basic divisions are:

- *functional expressions* (sequences with a discourse role such as openers, proverbs, slogans and so on);
- *composite units* (which retain a syntactic function);
- *lexical collocations* (consisting of two open class items, such as *ulterior motive*);
- *grammatical collocations* (consisting of one open and one closed class item, such as *in advance*).



Subsets of nonpropositional speech and their common properties, presented on a hypothetical continuum from most novel to reflexive.

Fig. 1. Van Lancker's continuum (from Van Lancker, 1987, p. 56).

All of these baseline categories span a continuum from non-idiomatic to idiomatic. This approach enables him to capture the difference between *under the table* (free combination),² *under attack* (restricted collocation), *under the microscope* (figurative idiom) and *under the weather* (pure idiom), without relinquishing the flexibility of labelling some sequences as borderline (pp. 27–28). However, his somewhat conservative line (compared with, for example, Bolinger, 1976, pp. 5, 7), excludes some important criteria. In particular he does not recognise a category of 'fused' sequences. *Fusion* is a process described by Peters (1983) in the context of first language acquisition. Strings created using the grammar can subsequently become fixed, or 'fused', as a single unit if they occur frequently enough for this to be advantageous. Howarth believes that all 'non-institutionalised' phrases, that is, ones that are transparent in meaning, are products of the online grammar, however commonplace they may be (pp. 27, 38, 40).

What is important to recognise is the fundamentally *descriptive* nature of such continuum models. They focus on defining a given sequence of words as more or less formulaic, according to certain criteria, such as those mentioned above. A shade closer to an *explanatory* model would be a continuum that identified stages of the journey that a given sequence makes across time in the mouth of a given speaker: is it formulaic when first uttered? Does it remain formulaic or become less so? Can it sustain a formulaic and non-formulaic identity at once?

1.7. Focus on word frequency

There is undoubtedly some sort of relationship between frequency and formulaicity, both in the sense that *some* formulaic sequences are very frequent, and

² Not in the sense of 'drunk', which would fall into the idiom category.

that formulaic output is frequently called upon. With regard to the first, we may note that in the case of idiomatic phrases such as *can of worms* and *bite the dust*, which have a literal counterpart, the frequency of the idiomatic version in a given language sample is by far the higher (Chafe, 1968; Barkema, 1993; Nunberg et al., 1994). As for the second, Sinclair (1991) ascribes the prevalence of formulaicity in language use to “the recurrence of similar situations in human affairs...a natural tendency to economy of effort [and]...the exigencies of real-time conversation” (p. 110). Cognitive grammarians argue that the formulaic nature of the adult language system comes about via a process of ‘schematization’ “through the reinforcement and progressive entrenchment of recurring commonalities, as well as the ‘cancellation’ (non-reinforcement) of features that do not recur” (Langacker, 1991, p. 107).

Frequency counts in corpus-based studies (e.g. Sinclair, 1991; Altenberg, 1993; Butler, 1997; Stubbs, 1997) can be word-focussed³ or sequence-focussed. One common frequency measure for estimating the degree of productivity of a language sample is the lexical type-token ratio, calculated by dividing the number of different words (types) by the total number of words (tokens). This has been widely used in areas ranging from language acquisition (Templin, 1957; Hess, 1986) and language pathology (Wachal and Spreen, 1973; Manschreck et al., 1984) to literary studies (Yule, 1944; Youmans, 1991). More recently, Perkins (1994) has extended the notion of type-token ratio beyond the level of the word in an analytical procedure which is sensitive to the whole range of the formulaicity continuum.

However, it may be premature to judge frequency as a *defining* feature of formulaicity. It has yet to be established that commonness of occurrence is more than a circumstantial associate. There are certainly many formulaic sequences whose culturally-based familiarity belies their comparative rarity in real text (e.g. *That’s another fine mess you’ve gotten me into; Time for bed, said Zebedee; Here’s one I made earlier*). As Hickey (1993) notes, “we must not rule out the possibility that an utterance which does not occur repeatedly is a formula” (p. 33). In other words, “phraseological significance means something more complex and possibly less tangible than what any computer algorithm can reveal” (Howarth, 1998, p. 27).

1.8. Function-based accounts of formulaic sequences

Nattinger and DeCarrico (1992, pp. 60ff) offer a lengthy and detailed function-based description of formulaic sequences in English as a target language for L2 learners. Their major divisions are:

³ Stubbs (1997) offers the following examples of formulaic sequences containing the word *care*: “would you care for a drink?; would you care to join us?; he was too tired to care; I couldn’t care less; I don’t care!; I never cared much for...; she was past caring; who cares!; anyone who cares to listen; anything you care to name; more than I care to think; the utmost care should be taken (not) to...; great care should be taken (not) to...; treated with great care; soothe away cares; take care of yourself; tender loving care; that takes care of that!; that’s been taken care of; the cares of state; the problem will take care of itself; well cared for; with studied care; without a care in the world; worldly cares” (p. 156).

- Social interactions:
 - conversational maintenance, including:
 - summoning (e.g. *how are you; I didn't catch your name*)
 - clarifying (e.g. *what did you mean by X?*)
 - shifting turns (e.g. *could I say something here?*)
 - conversational purpose, including:
 - questioning (e.g. *do you X?*)
 - refusing (e.g. *I'm sorry but X*)
 - expressing sympathy (e.g. *I'm very sorry to hear about X*).

- Necessary topics [that is, “lexical phrases [which] mark topics about which learners are often asked” (p. 63)], including:
 - autobiography (e.g. *my name is __*)
 - time (e.g. *what time X?; a __ ago*)
 - location (e.g. *what part of the __?*)
 - weather (e.g. *it's (very) __ today*)

- Discourse devices:
 - temporal connectors (e.g. *the day/week/month/year before/after __*)
 - exemplifiers (e.g. *in other words; it's like X*)
 - summarizers (e.g. *to make a long story short; my point (here) is that X*) (Nattinger and DeCarrico, 1992, pp. 60–66).

This typology both gains and loses usefulness as a descriptive tool by having a large number of sub- and sub-sub-types. However, this is probably inherent to the nature of the phenomenon, for Aijmer's (1996) functionally-based categorisation suffers the same problem. Based on the London-Lund Corpus of Spoken English, Aijmer's study is organised under the major headings of *thanking, apologies and requests and offers*, each of which occupies an entire chapter that is multiply sub-divided. Both of these surveys demonstrate the difficulty of distilling data into a streamlined model along the functional dimension at this level of abstraction. The complexity is all the greater because of *cross-functioning* (Moon, 1992), “the phenomenon of an expression being used with a function other than and additional to its primary one” (pp. 21–22). Indeed, an expression carries extremely subtle messages to the hearer about the text and subtext, so that there may be “a retreat or sheltering behind shared values which coerces agreement and pre-empts disagreement” (Moon, 1992, p. 24). However, Nattinger and DeCarrico's (1992) approach does support both Lattey's (1986) proposal that idioms can usefully be presented to learners via a pragmatic classification, and Martins-Baltar's (1998) attempt to create a dictionary of formulaic sequences that is organised not by keyword but by illocutionary force.

1.9. *Towards an inclusive picture*

While parts of the ground have been laid for an inclusive and integrated explanatory model of formulaic language, no single account to date has offered more than part of the story. Although there are many useful descriptive accounts, the variation between them in both content and approach underlines the complexity of the phenomenon. This complexity is clearly demonstrated by Wray's (1999) survey, to which we now turn.

2. **Where formulaic language fits in**

2.1. *The distribution of forms and functions*

Wray (1999) reports an extensive survey of the literature on formulaic language in the output of adult native speakers, L1 learners, child and adult L2 learners and aphasics. She establishes the characteristics of each subtype of data, and then compares the distributions. Her findings can be represented as a set of observational statements (Fig. 2).

Fig. 2 indicates that formulaic sequences are more than simply a linguistic unit. It suggests that they are a tool that can be put to many uses. The distribution of the uses appears to depend upon several independent factors, including maturational level, language knowledge and personal interactional need. It follows that a useful model of how formulaicity operates within our general linguistic and interactional competence must accommodate both moment-by-moment and developmental changes in the individual's processing and communicational agendas (Section 3).

2.2. *The relationship between formulaic and creative language*

Even though formulaicity was recognised as an important feature of normal language by Saussure (1916/1966), Sapir (1921), Bloomfield (1933), and Firth (1937/1964, 1952–9/1968), (see Wray, 1998, in preparation b), the Chomskian tradition has subsequently tenaciously challenged the idea that it plays anything more than a peripheral role in language production and comprehension. The focus of the debate is the nature of the lexicon. Chomsky's Standard Theory viewed the role of the lexicon as to "specify just those properties that are idiosyncratic, that are not determined by linguistic rule" (Chomsky, 1972, p. 39). One entailment of this is that the lexicon contains only items which cannot be further divided in any semantically or morphologically useful way (Mackay, 1979). Despite dissent regarding this definition of the lexicon (see Harlow and Vincent, 1988; Pinker, 1998) and evidence from empirical psycholinguistics that at least frequent words are stored in polymorphemic form (Harley, 1995, p. 287f), it is central to X-bar theory and hence to Government and Binding and Minimalism, where the lexicon "lists the syntactically atomic elements" (Webelhuth, 1995, p. 32). Early criticism of generative syntactic theory for failing satisfactorily to accommodate idioms (e.g. Chafe, 1968; Weinreich, 1969;

-
- All types of speaker seem to use formulaic sequences to achieve specific interactional goals (e.g. greeting, chastising) and to sustain the interaction (e.g. backchannelling, etc).
 - Formulaic sequences of entirely fixed form seem to be in the repertoire of all types of speaker.
 - All groups, with the possible exception of aphasics, use formulaic sequences to express aspects of their individual and group identity. In the case of L2 learners this may feature the deliberate use of non-native-like forms.
 - Native speaker adults, including aphasics, have a subset of formulaic sequences that are emotional expressions, swearwords, etc. These may be available to other groups too, just less likely to appear in the data normally collected.
 - The use of formulaic sequences to hold the turn in conversation seems to be considerably more prevalent in normal adults, native and non-native, than other groups.
 - The common use of sequences that are syntactically and/or semantically opaque, including metaphorical idioms, is apparently largely restricted to normal adult native speakers.
 - Those in a sub-group that could be characterised as struggling to attain the adult native speaker norm (L1 and L2 learners and aphasics) all make use of 'fused' formulas, which feature non-target language forms or meanings, created by the immature, interlanguage or disordered grammar and, in the latter two cases, open to fossilisation.
 - The 'strugglers' are also much more likely to rely on immediate imitation and repetition and, with the exception of non-improving aphasics, also use some sequences that later disappear again.
 - Child, and some adult, learners appear to use formulaic sequences as input for an analysis of the language, as a way of augmenting their grammatical and lexical knowledge. Children appear able to differentiate between those formulaic sequences that can usefully be turned to this purpose, and those which would be misleading.
 - When in a naturalistic environment, the speed and success of learning appears to correlate with the learner's social integration with the native speaker group.
 - Classroom-taught learners tend to over-generate, producing grammatical, but unidiomatic language (spoken and written), and seem less sensitive than native speakers to a word's collocational associates.
 - Individual variation in the quantity of formulaic language used appears to be a characteristic of the sub-group who have learned naturalistically.
 - Naturalistic learners (L1 and L2) are also the most likely to use formulaic sequences that they are either unable to (learners and aphasics), or simply have not ever needed to (all users), fully analyse.
 - Those who gained their knowledge of the language during childhood are most likely to produce formulaic sequences in an underenunciated way, and to sound fully idiomatic when using them. Adult L2 learners may, however, also underenunciate some sequences in the early stages.
 - Children seem to be more likely than adults to use formulaic sequences for private practice, though some adult learners may specifically select this as a strategy.
 - The use of formulaic sequences as a random dummy carrier of unrelated messages appears to be restricted to aphasics though children may use them for articulatory practice.
 - In general, fewer formulaic sequences are found in writing than speech, with non-natives tending to restrict themselves to a small selection which are over-used.
 - Both natives and non-natives employ formulaic sequences in their writing as a stylistic device, particularly to indicate the discourse structure.
-

Fig. 2. Summary of forms and functions of formulaic sequences (Wray, 1999, pp. 227–8).

Fraser, 1970; Makkai, 1972; Bolinger, 1976; Coulmas, 1979; Gazdar et al., 1985) has led to the admission of irreducible formulas to the lexicon. Thus, the formulaic sequences which are syntactically irregular or semantically opaque achieve the status of "big words" (Ellis, 1996, p. 111). However, formulaic sequences which are of a regular construction are excluded, for the lexicon cannot contain any items with a regular internal structure. Rather, all sequences of words, and indeed of morphemes, which *can* be assembled by rule, *must* be assembled by rule. A problematic corollary is that all grammatical sequences are equally valid and equally likely to occur. This is untenable in the face of our evident preference for some grammatical expressions of an idea over others (Coulmas, 1979, p. 239). For example, *I'm really glad you could come* could be rephrased as *I'm in a very glad state as a result of your coming* or *Your coming has brought me real gladness*, but native speakers would tend to find the latter two versions less 'natural' despite their grammaticality (Pawley and Syder, 1983). The solution to this problem is to allow for any string of words to be stored in the lexicon. As its composition is effectively overlooked once it is treated like a single

unit, it is easy to see how archaisms can survive language change and how a metaphor can become detached from its original literal meaning.

In recognising a central role for formulaic sequences, and allowing for their presence in the lexicon, it is not necessary to deny our capability for creativity, only to relegate it from the position of sole strategy (Wray, 1992, pp. 17f). Such a model of dual processing is one way of accommodating the holistic and analytic features of language. Meanwhile, other models of grammatical processing have found alternative ways of making formulaicity intrinsic. In recent years a range of cognitive–functional approaches have emerged (for an overview see Tomasello, 1998) which see grammar not as an autonomous system or module, but rather as largely deriving from more general cognitive processing mechanisms (e.g. Langacker, 1987, 1991; Bates and MacWhinney, 1989) and/or the communicative functions which it subserves (e.g. Hopper, 1988, 1992; Halliday, 1994; Chafe, 1994; Givón, 1995). A further recent trend in linguistics is to incorporate into the lexicon much of what was once thought to be grammatical (e.g. Bresnan, 1982; Pollard and Sag, 1994; Chomsky, 1995) and this progressive lexicalisation of grammar finds its ultimate expression in the approach known as Construction Grammar (Fillmore et al., 1988; Goldberg, 1995) in which formulaicity is a central principle.

In the model presented below, neither creativity nor formulaicity is seen as exclusive. The balance between these two capabilities has been discussed by Wray (1992), Locke (1993) and Perkins (in press), concurring with Sinclair's (1991) proposal that “[analyticity] could be imagined as a ... process which goes on in principle all the time, but whose results are only intermittently called for” (p. 114). Specifically, our view is that the best deal in communicative language processing is achieved by the establishment of a suitable balance between creative and holistic processes. The advantage of the creative system is the freedom to produce or decode the unexpected. The advantage of the holistic system is economy of effort when dealing with the expected (Wray, 1992, p. 19). Either system alone would be restrictive:

Without the rule-based system, language would be limited in repertoire, clichéd, and, whilst suitable for certain types of interaction, lacking imagination and novelty. In contrast, with only a rule-based system, language would sound pedantic, unidiomatic and pedestrian. It would require full access to all of the language faculties at all times, and there would be no ‘short cuts’. It would be a much more accurate reflection of what Chomsky terms competence, but *not* a reflection of *communicative competence* (Campbell and Wales, 1970; Hymes, 1972). (Wray, 1998, pp. 64f).

3. Formulaicity and the individual

The model proposed here is an attempt to account for the uses to which the individual puts formulaic language, and, specifically, what determines the choice, for that person, of a holistic or an analytic processing strategy at any given moment. Our starting place is the identification of two fundamental determiners, namely, the

priorities of social interaction and the constraints of memory on our processing capabilities. Both of these have been discussed in the literature over many years. Here, however, we shall take this idea much further, both by offering an explanation of the provenance of these strategies and also by demonstrating that they are two sides of the same coin.

In order fully to appreciate the usefulness to the speaker and hearer of formulaic sequences, we need first to acknowledge the usefulness of the alternative, that is, analytic language, because the model is built upon a fundamental relationship between the two. There is no question but that we are capable of using our knowledge of grammar creatively, and this has been the thrust of the Chomskian tradition for the last 40 years (see Section 2). However, it is important to recognise the nature of this ‘creativity’, because, despite what has often been claimed, not all of it is anathema to the restricted scope of formulaic language. An utterance like *Mr Brown will be sorry to have kept you waiting* is novel in the sense that it is tailor-made to a particular situation, but it is not novel in the sense that it conforms to a paradigm of sequences that can be characterised by the formulaic sequence⁴ *NP be-TENSE sorry to keep-TENSE you waiting* (Pawley and Syder, 1983, p. 210). In contrast, there is a different order of novelty in a sequence like: *bring on your fireworks, which are a mixed splendor of piston and of pistil; very well provided an instant may be fixed so that it will not rub, like any other pastel* (e.e. cummings). Such poetic constructions express complex and novel ideas through exploiting our knowledge of what the grammar and lexicon can (and cannot) do. This distinction between types of creativity is equally important whether you view it as representative of two places on a continuum (see Section 1.6) or as applications of two quite different types of language knowledge (Wray, 1992). It is important because we know that we react differently to the two types. The former—the tailoring of utterances to our everyday communicational needs—we cannot exist without, and in order to help us cope with the unexpected we employ a range of pragmatic interpretational strategies to access the most likely meaning of the incomprehensible (Grice, 1975; Sperber and Wilson, 1995). But ‘true’ novelty, which plays with the boundaries of the grammar and lexicon, is poetic precisely because it pays little heed to convention—it is the exercise of pure analysis, and consequently is difficult, challenging, and optional. The breaking of convention even extends to the analysis and reworking of formulaic sequences themselves, as in *Fred was hit, the dust was bit* in a song by C. O’Beirne. Formulaic language cannot create this sort of poetic novelty (though it can sustain it, for lines of poetry are often memorised and reproduced formulaically), but in the majority of our linguistic interaction we do not need it. The use of a ‘purely’ analytic strategy is a

⁴ Lawrence Bouton (1998) points out that some expressions which have no words in common at all still have to be seen as formulaic at some level. In response to the question *Did you enjoy the party?* a person might answer: *Is the Pope a catholic? Does a one-legged duck swim round in circles? Does the sun rise in the morning?* etc. The formula is a semantic–pragmatic one, which allows the use of any question to which the answer is both obvious and is the same as the answer to the original question. The link between this sort of formula and the paradigms like *NP be-TENSE sorry to keep-TENSE you waiting* is clearly illustrated by expressions like: *He’s one brick short of a load, She’s one sandwich short of a picnic, He’s one shelf short of a bookcase*, etc. to which one’s full creativity can be applied, within the confines of the frame.

peripheral activity, and while we do indeed need an on-line grammar to deal with novelty, it does not need to constitute a major element of normal language processing (compare Widdowson, 1989, p. 135).

In our model, the use of formulaic language is viewed as central to processing, but not to the exclusion of the full break-down and build-up of utterances from scratch as and when required. It is all a question of emphasis. Accounts that place analyticity in the central role still have to accord formulaicity some place, in order to account, at the very least, for idioms and interactional routines. In this model, the focus is shifted, so that formulaicity characterises the normal approach to processing, with analyticity on hand to pick up any difficulties, such as can be caused by a speaker's thick accent or non-native grammar, background noise, dysfluency, poetry, word games, and so on.⁵ In short, this model holds that our baseline strategy in everyday language processing, both production and comprehension, "relies not on the *potential for the unexpected* in a given utterance but upon *the statistical likelihood of the expected*" (Wray, 1992, p. 19, original emphasis). Our grammatical capabilities are on hand for emergencies, rather in the way that an engineer is on standby at a factory, while the less knowledgeable but competent operators routinely work the machines (Wray, 1992, p. 10). The way this works is described in Section 3.2. First, however, we consider the socio-interactional triggers to the use of formulaic sequences.

3.1. *Formulaicity as a tool for social–interaction*

Table 2 demonstrates that three central functions of social interaction can be achieved using formulaic language. In saying this, it is not necessary to hold that these functions cannot also be achieved using novel structures created from scratch at the time of uttering, but it is clear from the examples that it is formulaic sequences that we associate with these functions, and that they appear to achieve them very well. We shall return later to the question of *why* this is the case. The three functions relate to the speaker's manipulation of his/her world and the expression of his/her individual and group identity. The first reflects the fact that we are unable fully to meet our own physical, emotional and cognitive needs, and therefore, at times have to engage others to achieve outcomes that are beyond our own power. We use commands, requests, bargain structures, etc., as the carriers of these directives, and a range of markers (such as politeness markers) to frame them in a way that will maximise the likelihood of the required event coming about. Just as meeting our physical needs is essential for our survival, asserting our individual identity and our membership of the group (our differentness and our sameness—categories 2 and 3 in Table 2) is central to ensuring that we neither become subsumed within, nor are excluded from, the social networks which we feed off emotionally, and which directly

⁵ Wray (1992, pp. 80–95) points out that the majority of psycholinguistic experiments take the form of word games or other tasks that expressly encourage an analytic approach to processing, so that we should be wary of using the findings from such experiments to make sweeping statements about the way processing takes place during interaction. See also Paradis (1997).

Table 2
Formulaic sequences as devices of social interaction

Function	Effects	Type	Examples
Manipulation of others	Satisfying physical, emotional and cognitive needs	<ul style="list-style-type: none"> • Commands • Requests • Politeness markers • Bargains, etc. 	<ul style="list-style-type: none"> • Keep off the grass; hand it over • Could you repeat that please? • I wonder if you'd mind... • I'll give you __ for it
Asserting separate identity	(a) Being taken seriously	<ul style="list-style-type: none"> • Story-telling • Turn claimers and holders, etc. 	<ul style="list-style-type: none"> • You're never going to believe this, but... • Yes, but the thing is... ; Thank you very much (<i>in response to invitation to speak</i>); The first thing that you have to realize, of course, in addressing this issue is...
	(b) Separating from the crowd	<ul style="list-style-type: none"> • Personal turns of phrase. 	<ul style="list-style-type: none"> • I wanna tell you a story (<i>Max Bygraves</i>); You know what I mean, Harry (<i>Frank Bruno</i>)
Asserting group identity	(a) Overall membership	<ul style="list-style-type: none"> • 'In' phrases • Group chants • Institutionalized forms of words, etc. • Ritual 	<ul style="list-style-type: none"> • Praise the Lord!; as the actress said to the bishop • We are the champions • Happy birthday; dearly beloved, we are gathered here today... • Our Father, which art in Heaven...
	(b) Place in hierarchy (affirming and adjusting)	<ul style="list-style-type: none"> • Threats • Quotation • Forms of address • Hedges, etc. 	<ul style="list-style-type: none"> • I wouldn't do that if I were you • "I wouldn't want to belong to any club that would have me as a member" (<i>Groucho Marx</i>) • Your Highness • Well I'm not sure (as a polite denial or refusal)

contribute to our success in finding a reproductive partner and rearing offspring. Thus it is not surprising to find that these three functions (though particularly the first and the third) characterise the messages contained in the holistic noise–gesture communication of primates (Reiss, 1989). Dunbar (1996) also highlights the centrality of social interaction in both primate and human society. Wray (1998, in preparation b) proposes that this parallel is indicative of human language having developed out of a holistic protolanguage which was a phonetically-mediated larger and more subtle version of an ancient primate system. If so, human protolanguage had no words and no rules, just utterances associated with specific meanings, that achieved goals very similar to those described in Table 2.

In Section 2 we saw that different types of speakers (children, non-natives, aphasics, adult natives) favour different types of formulaic language and use it for different things. Table 2 indicates why this should be. The child has a different agenda from the adult, regarding its need to get people to do things and to express belonging. A priority for the young child is getting its carers to provide food and comfort, at a level unnecessary for an adult. The adult has more subtle needs and also has to frame them differently because of the hierarchical relationships that exist between communicating adults. A two-year old child can use *that's mine* to claim her own and other people's possessions, to request food, and demand space on a chair, but an adult needs individual hedged formulas like *Oh, is that yours?*, *I don't suppose there's any more is there?* and *Room for a little one?* to achieve the equivalent effects. The words are different because the worlds of the child and adult are different, and different things work. The adult L2 learner is in a different position again. Often under-informed about the culturally appropriate way to express social relationships and request actions from others, the drive to achieve the three socio-interactional goals identified in Table 2 will nevertheless direct the learner towards assembling a particular set of formulaic sequences (Wray, in press b). Just as the literature attests (Wray, 1999), in the absence of sufficient knowledge of the language, these may turn out to be a mix of target language sequences used appropriately and inappropriately, and interlanguage ones coined to fill the gaps. Meanwhile, other phrases, considered useful by a teacher and dutifully studied by the learner, may well not become fully internalised unless and until they actually have a function for the speaker in interaction (Wray, in press b, in preparation a). Finally, the aphasic has yet another agenda of priorities. If physically disabled, one major change will be the need to harness a range of formulaic sequences that request others to perform actions previously achieved independently. Also important may be the re-establishment of a strong sense of personal status and individual identity, at a time when many interlocutors may fail to see the person behind the symptoms. Thus, we should expect that aphasics will employ formulaic language to achieve these functions. This is indeed the case. Where an aphasic has a restricted range of available sequences, a greater functional load is placed upon the ones which can be accessed, so that they convey a number of meanings not normally associated with them (Van Lancker, 1987). All of this offers an explanation for why an aphasic can produce some fluent strings whilst experiencing difficulties in producing novel utterances. Wray (1992, pp. 78ff) has suggested that formal tests of aphasic language may miss the formulaic language, because the nature of testing tends to cause a focus on analytic production and comprehension (compare footnote 5 on page 13).

3.2. Formulaicity as a short-cut in processing

The second major function for formulaic language is that represented in Table 3. It seems that we use prefabricated sequences as a way of minimising the effects of a mismatch between our potential linguistic capabilities and our actual short term memory capacity. As Becker (1975) points out, it makes little sense to produce from scratch those word strings which we use many times, and we appear to use formulaic

sequences to reduce the amount of new processing to only that which *has* to be new. Recent research (Raichle, 1998; McCrone, 1999) shows that once the brain is familiar with a linguistic task, it is able to by-pass the processing route that was used to learn it. As the top row of Table 3 illustrates, strings of words stored and retrieved together will become associated with agreed meanings, which may be entirely transparent (e.g. *I have known_ for_ years in my capacity as_*) or rather more indirect (e.g. *put the kettle on, will you?* meaning *please make me a hot drink*). Words may collocate to form phrases which could, in principle, mean several different things, but which are only normally interpreted in one agreed way (e.g. *bullet point*); in some cases these may even become clichés (e.g. *the current economic climate*). Thus, the benefits of prefabricated language in reducing processing effort can account for why an individual or indeed a whole speech community comes to prefer certain collocations and expressions of an idea over other equally permissible ones (Pawley and Syder, 1983).

Table 3
Formulaic sequences as compensatory devices for memory limitations

Function	Effects	Type	Examples
Processing short-cuts	Increased production speed and/or fluency	<ul style="list-style-type: none"> • Standard phrases (with or without gaps) • Standard ideational labels with agreed meanings 	<ul style="list-style-type: none"> • Put the kettle on, will you?; I have known __ for __ years in my capacity as __ • Personal computer; bullet point; the current economic climate
Time-buyers	<ul style="list-style-type: none"> • Vehicles for fluency, rhythm and emphasis • Planning time without losing the turn 	<ul style="list-style-type: none"> • Standard phrases with simple meanings • Fillers • Turn-holders • Discourse shape markers • Repetitions of preceding input 	<ul style="list-style-type: none"> • Make a decision; draw a conclusion; a sea change; at the end of the day (<i>in the sense of 'really'</i>); one way and another • If the truth be told; if you want my opinion; if you like • And another thing; and let me just say.... • There are three points I want to make. Firstly... Secondly... Thirdly/Lastly... • (A: What's the capital of Peru?) B: What's the capital of Peru? (Lima isn't it?)
Manipulation of information	Gaining and retaining access to information otherwise unlikely to be remembered	<ul style="list-style-type: none"> • Mnemonics • Lengthy texts one is required to learn • Rehearsal 	<ul style="list-style-type: none"> • Thirty days hath September...; Richard of York gave battle in vain • Shall I compare thee to a summer's day? • <i>Rehearsing a telephone number while looking for a pen</i>

Prefabricated time-buying sequences (the second row of Table 3) promote fluency and protect the speaker's turn during planning. The pace and rhythm of an utterance can be fine-tuned by using word-strings which are semantically equivalent to single words (e.g. *make a decision: decide; draw a conclusion: conclude/realise*). Other time-buyers are fillers, turn-holders, discourse shape markers and repetitions, including the ephemeral verbatim repetitions of a previous turn that Tannen (1989, p. 45) observes in conversation and Bygate (1988) reports for L2 learners. The promotion of fluency is important for being heard out, and thus for being taken seriously as an individual (see Section 3.1). The bottom row in Table 3 represents a somewhat different set, which relates to the use of language to *extend* memory. Mnemonics and lengthy texts deliberately memorised or repeated offer us a means of retrieving information that we might otherwise find it difficult to recall.

Wray (1998, in preparation a) views the processing-related functions of formulaic sequences as secondary, and of much later origin in evolutionary terms. They relate to the nature of the grammar, which was late-appearing, and represent an attempt to square an over-powerful analytic system that developed for reasons other than communication, with the inadequate working memory which is its forum.

Again, it is easy to see how differences in the processing constraints of different types of speaker can contribute to an explanation of the variation outlined in Section 2. Children appear to be constrained in language processing by a limited short term memory capacity (Bates et al., 1988, pp. 122f; Elman, 1993; see also Cromer, 1991). Adult L2 learners, especially in the early stages, may have a somewhat cumbersome arrangement of language knowledge (e.g. R. Ellis, 1994, pp. 388ff; N. Ellis, 1996). Both types of learner can alleviate their problems by relying on formulaic sequences where possible, as a way of bypassing the difficulties of processing. This permits the analytic processing to be focussed where it is most needed. Adult native speakers, too, have processing limits, evident if they attempt to share their speech production with other concomitant tasks, such as listening to the radio or TV (Jaffe, 1978, p. 55), performing a difficult manoeuvre on the road, or when distracted by an extraneous event. There are even studies which show that aphasic-like language can be induced in normal adults when put under stress (e.g. Kilborn 1991; Miyake et al., 1994; Blackwell and Bates, 1995). That dysfluency can also occur when we are simply trying to formulate and express a complex idea, or, as a hearer, trying to follow one, suggests that the short-cuts afforded by formulaicity are no luxury, but are crucial to managing an over-demanding on-line processing system. The need for input to contain some measure of formulaicity is central to the integration of our model, and it is to this that we now turn.

3.3. *The common purpose of formulaicity*

Sections 3.1 and 3.2 outline two apparently unrelated purposes for formulaic language. On the one hand it is a means of ensuring the physical and social survival of the individual through communication, and on the other it is a way of avoiding processing overload. However, these two are in actual fact two sides of the same coin. On the one hand, the driving force behind the processing short-cuts is ensuring

that the speaker's production is fluent and that information is available when required: formulaic language by-passes, partially or entirely, depending on the form, the generative system. The driving force behind the socio-interactional formulas is ensuring that the speaker gets what he/she wants and is perceived as an individual within the group. Significantly, formulaic language is better suited to this than novel language is, because a hearer is more likely to understand a message if it is in a form he/she has heard before, and which he/she can process without recourse to full analytic decoding.⁶ For example, army commands, which have to be obeyed quickly, maximise their chance of being understood by being formulaic. Thus, we see that, just as the processing short-cuts are a means of ensuring that the *speaker* achieves successful production, so the socio-interactional formulae are a means of ensuring that the *hearer* achieves successful comprehension (Fig. 3). This, however, is not some kind of altruism on the speaker's part. The hearer's success is entirely in the interests of the speaker because it is the speaker's way of achieving the socio-interactional functions identified in Table 2. In both cases, it is the speaker who benefits from using formulaic sequences.

In Fig. 3, then, we see represented the production and comprehension benefits (to the speaker) of using formulaic language. The point of intersection is the set of formulaic sequences which are used to structure discourse, for these aid both production and comprehension simultaneously.

The model thus far, then, accounts for the moment-by-moment strategy choices of the individual, by characterising a consistent tension between a number of priorities:

- the need to decode successfully input that is in a novel (i.e. non-formulaic or otherwise unexpected) format, including not only dysfluent, heavily accented,

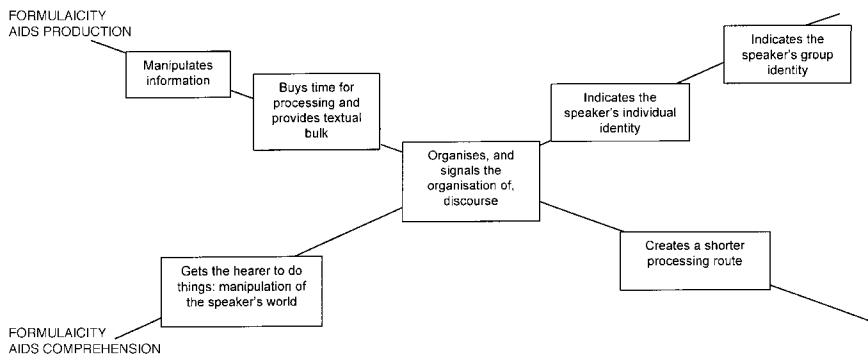


Fig. 3. The roles of formulaic language in benefiting the speaker.

⁶ This means that the speaker's success depends on his/her ability to correctly anticipate the hearer's knowledge of the language. Thus, rather than speakers all going their own way, constructing their own set of preferred collocations and sequences, the uniformity of the language within a given speech community is kept in check by the speakers' self-interest in matching their output to what the hearers will understand.

ungrammatical and poetic language but also straightforward novel propositions, especially out of context.

- the need to process quickly and accurately, without overloading the system.
- the need to ensure the message (whether primarily manipulative, social or informational) has maximum chance of being successfully understood.

In normal interaction, the default setting is formulaicity, both for production and for comprehension. This enables the individual to focus his/her analytic faculties away from the linguistic ‘packaging’ and onto the production and evaluation of propositions, the updating of contextual information and the making of predictions about what is going to happen next (Wray, 1992). Focus can switch to an utterance itself if there is any irregularity or breakdown in comprehension or production. This is often marked by dysfluency in the speaker and/or by a hearer failing at first to decode it, until suitable attention is brought to bear upon it.

What the model has not offered up to this point is any specific account of the developmental dimension, beyond the observation that the child’s interactional agenda changes over time. What is needed is an explanation for how, if formulaic language is so generally successful in meeting the interactional needs of the individual, the analytic grammar gets a chance to develop at all. In the next section, we offer such an explanation.

3.4. Formulaicity as a feature of development

The developmental aspect of the model presented here enables us to account for the rather curious appearance, disappearance, and reappearance of formulaic sequences during language acquisition (Perkins, in press). The acquisitional process begins with a heavy reliance on formulaic language. Yet, despite the fact that the target adult language is also highly formulaic, the child’s use and recognition of formulas actually wanes over several years (Perkins, 1983; Tomasello, 1992), with an adult-like appreciation of idiom, for example, not fully established until adolescence (e.g. Nippold and Martin, 1989; Gibbs, 1991).

We build here upon Locke’s (1993, 1995) account of early language acquisition. Locke reconciles the gestalt and analytic strategies in the child’s language processing, described at length by Peters (1977, 1983), Nelson (1981), Lieven et al. (1992) and others, and summarised by Wray (1999) by viewing them as evidence for the existence of separate but complementary neural mechanisms which he calls ‘specialization in social cognition’ (SSC) and ‘grammatical analysis module’ (GAM). The SSC, which operates initially in response to social stimulation, identifies features of the environment which are seen as socially important and thus motivates the identification, selection and storage of units meaningful within the linguistic environment (phase 1, Fig. 4). Any internal structure that such units may possess is irrelevant to the SSC which in any case is presumably unable to process it. The GAM, which begins to operate between 20 and 30 months (phase 2), depending on the prior success of the SSC in acquiring a requisite number and variety of linguistic items, is analogous to Chomsky’s innate Universal Grammar and is responsible for

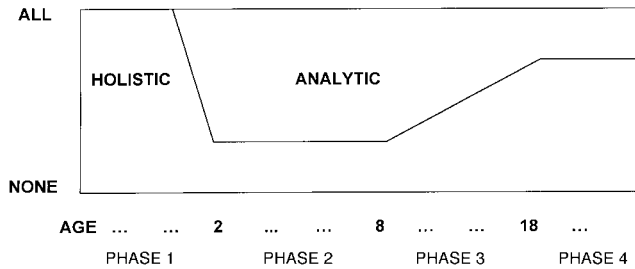


Fig. 4. Relative proportions of holistic and analytic involvement in language processing from birth to adulthood (schematic representation).

identifying the constituent structure of the items acquired via the SSC. The SSC is located by Locke in the right cerebral hemisphere and the GAM in the left.⁷

Locke (1997) briefly considers the development of communication after the transition from what we term phase 1–2, that is, after the SSC has acquired a sufficiently large store of formulas to trigger the activation of the GAM. He proposes a further phase of ‘integration and elaboration’ in which the SSC and GAM interact and which is concomitant with the expansion of the lexicon and the automatization of syntactic and phonological processing. Since he provides no details of this phase, however, we will sketch out a possible scenario (phase 3) consistent with the literature on the subject. It seems likely that at some point the wholesale generation of utterances from scratch using the grammar will become no longer efficient or economical. If the same, or similar, groups of elements are being continually encountered and/or produced it will make good economical sense to store them as separate items (compare ‘fusion’, Peters, 1983). This places an obvious burden on memory, though not an excessive one (Bolinger, 1976). Thus, in phase 3, the SSC will frequently override the GAM in the case of frequently occurring items with a resultant gain in automaticity. The storage site for such items is presumably the right hemisphere, where Locke locates the SSC.⁸

The GAM does not simply subserve the SSC, however, but still plays an active role in identifying commonalities among formulas and setting up numerous (though less numerous than the total number of individual formulaic sequences) semi-productive ‘frames’. The durability of a given formula or frame will depend on its frequency of use and its pragmatic and sociocultural salience. Thus the equilibrium of the SSC–GAM amalgam will be constantly changing. A fully equilibrated system

⁷ The separation may well not be as discrete as this suggests. In a report of research on children with early hemisphere damage, Eisele and Aram (1995) note “the results of these studies suggest that neural systems mediating the production of first words may, during early language development, depend to a greater extent on the normal functioning of both the left and right hemisphere” (p. 674). However, see also the general contextualisation provided by Bates et al. (1988, p. 63).

⁸ This is almost certainly an oversimplification. The consensus in recent research on neurolinguistics is that “normal language processing seems to require the integration of the functions of both the right and left hemispheres” (Blumstein 1988, p. 231). See also Wray (1992) and footnote 7.

may not appear until late childhood or adolescence (phase 4). Evidence for this comes from studies of idiom comprehension (for a review see Perkins, 1999) and a number of production phenomena in developmental language disorders (Locke, 1994).

We assume, then, the early stages of the process (phases 1 and 2) to be accounted for by Locke's (1993, 1997) theory, whereby a specialization in social cognition, influenced by a theory of mind, identifies, selects and stores a sufficient and requisite number of salient formulaic linguistic items to activate a specifically language-oriented analytical mechanism which, through identifying commonalities among the stored formulas, begins the process of creating a generative grammar for the language of the child's environment somewhere between 20 and 30 months of age. The process continues until roughly the age of 8, and this stage of development is marked by a preference for analytic over formulaic language processing. During the subsequent period (phase 3), which lasts until adulthood, the organization of the language system becomes progressively more formulaic and the primary role of the analytical mechanism becomes that of constantly readjusting the formulaic continuum by deciding whether a given item is unique, or else shares sufficient properties in common with other items to justify subsequent collapsing and re-storage as a single, partly productive formulaic frame. During this phase, language production increasingly becomes a top-down process of formula blending as opposed to a bottom-up process of combining single lexical items in accordance with the specification of the grammar. At the same time, language comprehension becomes increasingly a top-down pragmatically driven process of formulaic 'macro-processing', with the bottom-up grammatical 'microprocessing' mechanism only being used as a default in cases where macroprocessing fails to yield a sufficiently relevant⁹ interpretation (compare Wray, 1992) (this notwithstanding the fact that both processing modes are automatically and simultaneously available). In both production and comprehension, however, there is considerable scope for variation in the balance between holistic and analytic processing as a result of individual sociocognitive and sociocultural differences.

This developmental account is further strengthened by a significant observation regarding the nature of the child's micro-environment, as determined by its more general pattern of mental and physical development. As phase 1 in Fig. 4 suggests, reflecting Locke's model, the child's earliest goal is one of social integration and the meeting of its physical needs. This requires the accumulation of a set of formulaic sequences that successfully achieve that end (compare Halliday, 1974). What is less obvious, however, is the extent to which the child's communicative needs then remain essentially static, at precisely the time when the GAM is dominant. This is what the model predicts must happen, but, in consideration of the wealth of worldly experiences which the child gains during this period, it is easy to overlook the fact that for a substantial period of time, the child is largely cushioned from the need to develop an additional interactional repertoire, because it has an extremely limited set of social roles (Gross, 1996, p. 525). During the time when it is most involved in learning to label, and is becoming adept at slotting open class items into formulaic

⁹ In the sense of Sperber and Wilson (1995).

frames, it exists within a socio-interactive bubble. It is a well-recognised feature of parenthood that, in the early years, the child–adult and adult–child routines are the same, wherever you are. The child’s focus is on its physical needs and its relationship with its carers, so the verbal interactions are largely constant whether at home, in a restaurant, at the zoo or a royal garden party. Certainly, referential utterances will differ, but a young child is both protected from, and largely impervious to, any need to interact with anyone other than its carers, let alone being party to any knowledge of how to do so appropriately.¹⁰ Indeed, a child so rarely encounters a genuinely new interactional environment (usually one in which the carers are not able to act as an intermediary), that it often has to be specifically taught the formulaic sequences that go with it, such as *Thank you for having me* when leaving a friend’s house (compare Ely and Gleason, 1995). Where new environments are encountered, and a new social role has to be adopted, such as the first days at nursery or school or the first time the child buys something in a shop, the effect can be traumatic, with the child literally lost for words. However, it is not only children who tend to accumulate interactional formulaic sequences on a needs only basis, and it is not uncommon for an adult to be similarly floored for what to say when first encountering a bereaved friend, a job interview panel or a visiting dignitary. What all of this suggests is that the child is afforded the luxury of developing the analytic grammar by being protected, during these vital years, from the need to accumulate the wide range of formulaic sequences that it will ultimately need in order to function as a normal social adult. Meanwhile, the development of the child’s analytic knowledge of language is itself fuelled by its general cognitive development (Elman, 1993; Locke, 1993), and in particular the prioritisation of learning about the world.

4. Making sense of the differences

The model presented in this paper aims to explain the basis of observed variation between the forms and functions of formulaic sequences in different types of linguistic data. It does so by identifying a complex dynamic according to which the individual’s choice of a formulaic sequence at any one time is determined by several factors, including:

- his/her overall knowledge of the language and/or stage of cognitive development
- his/her purpose in speaking, including the intended effect on the hearer
- the complexity and novelty of the idea
- the interactional and discourse context
- competition from concomitant activities

¹⁰ Children’s make-believe games indicate that they do pick up a certain amount of ‘appropriate’ talk from observation, but their failure automatically to know what to say, when first called upon to interact in a situation they may have been a passive observer in many times, indicates that this does not always happen.

The model offers itself to critical evaluation on many fronts and will no doubt evolve in the face of challenges from many kinds of data, both of the type discussed here and others. In particular, the occurrence of formulaic language in the speech of Alzheimer's patients and people with autism raises some interesting issues. As regards the latter, a standard characterisation of autism is the absence of social-interactive skills (e.g. Prizant, 1983, p. 296), which we might expect to lead to highly *unformulaic* language. However, Prizant suggests that "the [formulaic] language patterns of autistic persons ... may reflect an inability to segment others' utterances and realise their internal structure" (p. 303), in which case, the formulaicity is not socio-interactionally motivated but rather is a 'Hobson's choice' solution to processing constraints.

A further challenge comes from second language acquisition studies, where there are some indications (e.g. Hanania and Gradman, 1977; Schmidt, 1983; Rehbein, 1987; Bolander, 1989), that adults, both classroom taught and learning naturalistically, use formulaic sequences in the early stages, but find them a major sticking point in the advanced stages (Pawley and Syder, 1983). This is reminiscent of the pattern described for children in Section 3.4, yet if the similarity is more than coincidental, it is clearly problematic for our proposal that, in children, the pattern is motivated by social and cognitive development. At this stage it is not easy to be sure if this is a genuine problem for the model, but there are some indications that it is not. First, adult learners, both naturalistic and classroom-based, also need to accumulate a small set of survival phrases that achieve basic socio-interactive functions. Like children, once they have these, they move into a period of relative interactive stability. This is most true of classroom learners, but evidence from studies of naturalistic learners indicates that, although they are certainly not socially buffered—that is, they can potentially find themselves in any situation that a native speaker would—they may actually buffer themselves in at least two ways. One is by using their non-nativeness as an excuse to withdraw from, or under-communicate in, certain situations (Ellis, 1994, pp. 82ff) something which they can signal by using a fused non-native formula (cf. Rampton, 1987). The other is by specifically reducing their physical and social needs to match their existing repertoire, either permanently (e.g. Rehbein, 1987) or until they are ready to cope (e.g. Schmidt, 1983). As to their later achievements, Pawley and Syder (1983) point out that few non-native speakers ever fully accumulate the native repertoire of formulaic sequences. Rather, they tend to *over-generate*, producing grammatical utterances that are simply not idiomatic. Thus, it seems that adult second language learners can get caught in phase 2 or 3. We can explain this, perhaps, in terms of the fact that, without the child's liberty to break the social norms of adult interaction, adult second language speakers tend to find themselves plunged rather sooner, in terms of the acquisition of the language, into new interactive situations that are beyond their grammatical competence and linguistic experience. The result is the need to find suitable utterances and, in lieu of sufficient exposure to the native-like ones, the tendency to make a best guess. In this case it may not be so much a question that a certain unidiomatic sequence becomes fused and is used for ever more by that speaker, as that a wider range of options seems familiar and plausible, than would to a native speaker. After all, one of the

effects of having a preferred way of saying something is that it not only promotes the frequency of that sequence but also reduces to virtually zero the frequency of other equally grammatical alternatives (Langacker, 1991). In short, the pattern of formulaic utterances in second language learners may be more supportive to the model than it first seems.

The final judgment of the resilience of this model must, as always, lie with the weight of evidence over time. What we have presented here is, we believe, the best integrated account to date of the pattern of formulaic sequences across data types, combining clear, testable premises and predictions with a framework that is flexible enough to accommodate individual and group variation across the whole range of communicational environments.

References

- Aijmer, K., 1996. *Conversational Routines in English*. Longman, London and New York.
- Altenberg, B., 1990. Speech as linear composition. In: Caie, G. Haastrup, K., Jakobsen, A.L., Nielsen, J.E., Sevaldsen, J., Specht, H. and Zettersten, A. (Eds.), *Proceedings from the Fourth Nordic Conference for English Studies*, Vol. 1. Department of English, University of Copenhagen, pp. 133–143.
- Altenberg, B., 1993. Recurrent verb–complement constructions in the London–Lund Corpus. In: Aarts, J., de Haan, P., Oostdijk, N. (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Rodopi, Amsterdam, pp. 227–245.
- Baayen, H., Lieber, R., 1991. Productivity and English derivation: a corpus-based study. *Linguistics* 29, 801–843.
- Barkema, H., 1993. Idiomaticity in English NPs. In: Aarts, J., de Hahn, P., Oostdijk, N. (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Rodopi, Amsterdam, pp. 257–278.
- Bates, E., Bretherton, I., Snyder, L., 1988. *From First Words To Grammar*. Cambridge University Press, Cambridge.
- Bates, E., MacWhinney, B., 1989. Functionalism and the competition model. In: MacWhinney, B., Bates, E. (Eds.), *The Cross-Linguistic Study of Sentence Processing*. Cambridge University Press, Cambridge, pp. 3–73.
- Becker, J., 1975. *The phrasal lexicon*. Bolt Beranek and Newman Report No. 3081, AI Report No. 28.
- Blackwell, A., Bates, E., 1995. Inducing agrammatic profiles in normals: evidence for the selective vulnerability of morphology under cognitive resource limitation. *Journal of Cognitive Neuroscience* 7, 228–257.
- Bloomfield, L., 1933. *Language*. Allen and Unwin, London.
- Blumstein, S.E., 1988. Neurolinguistics: an overview of language–brain relations in aphasia. In: Newmeyer, F.J. (Ed.), *Linguistics: The Cambridge Survey*, Vol. 3: *Language: Psychological and Biological Aspects*. Cambridge University Press, Cambridge, pp. 210–236.
- Bolander, M., 1989. Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In: Hyltenstam, K., Obler, L.K. (Eds.), *Bilingualism Across The Lifespan*. Cambridge University Press, Cambridge, pp. 73–86.
- Bolinger, D., 1976. Meaning and memory. *Forum Linguisticum* 1, 1–14.
- Bouton, L., 1998. Formulaic implicatures as conversational routines. Paper presented at the 6th International Pragmatics Conference, Reims, France, July 1998.
- Bresnan, J. (Ed.), 1982. *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge, MA.
- Butler, C.S., 1997. Repeated word combinations in spoken and written text: some implications for functional grammar. In: Butler, C.S., Connolly, J.H., Gatward, R.A., Vismans, R.M. (Eds.), *A Fund of Ideas: Recent Developments In Functional Grammar*. IFOTT, University of Amsterdam, Amsterdam, pp. 60–77.

- Bygate, M., 1988. Units of oral expression and language learning. *Applied Linguistics* 9 (1), 59–82.
- Campbell, R., Wales, R., 1970. The study of language acquisition. In: Lyons, J. (Ed.), *New Horizons in Linguistics* 1. Penguin, Harmondsworth, pp. 242–260.
- Chafe, W.L., 1968. Idiomaticity as an anomaly in the Chomskyan paradigm. *Foundations of Language* 4, 109–127.
- Chafe, W.L., 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. University of Chicago Press, Chicago.
- Chomsky, N., 1972. *Language and Mind*, 2nd Edition. Harcourt Brace Jovanovitch, New York.
- Chomsky, N., 1995. *The Minimalist Program*, MIT Press, Cambridge, MA.
- Code, C., 1987. *Language, Aphasia, and the Right Hemisphere*. John Wiley, Chichester.
- Code, C., 1997. Can the right hemisphere speak? *Brain and Language* 57, 38–59.
- Coulmas, F., 1979. On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3, 239–266.
- Coulmas, F., 1994. Formulaic language. In: Asher, R.E. (Ed.), *Encyclopedia of Language and Linguistics*. Pergamon, Oxford, pp. 1292–1293.
- Cowie, A.P., 1992. Multiword lexical units and communicative language teaching. In: Arnaud, P.J.L., Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. Macmillan, Basingstoke, pp. 1–12.
- Cowie, A.P., 1994a. Applied linguistics: lexicology. In: Asher, R.E. (Ed.), *Encyclopedia of Language and Linguistics*. Pergamon, Oxford, pp. 177–180.
- Cowie, A.P., 1994b. Phraseology. In: Asher, R.E. (Ed.), *Encyclopedia of Language and Linguistics*. Pergamon, Oxford, pp. 3168–3171.
- Cowie, A.P. (Ed.), 1998. *Phraseology*. Clarendon Press, Oxford.
- Cromer, R.F., 1991. *Language and Thought in Normal and Handicapped Children*. Blackwell, Oxford.
- Dunbar, R., 1996. *Grooming, Gossip and the Evolution of Language*. Faber, London.
- Eisele, J., Aram, D.M., 1995. Lexical and grammatical development in children with early hemisphere damage: a cross-sectional view from birth to adolescence. In: Fletcher, P., MacWhinney, B. (Eds.), *The Handbook of Child Language*. Blackwell, Oxford, pp. 664–689.
- Ellis, N.C., 1996. Sequencing in SLA: phonological memory, chunking and points of order. *Studies in Second Language Acquisition* 18, 91–126.
- Ellis, R., 1994. *The Study of Second Language Acquisition*. Oxford University Press, Oxford.
- Elman, J.L., 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99.
- Ely, R., Gleason, J.B., 1995. Socialization across contexts. In: Fletcher, P., MacWhinney, B. (Eds.), *The Handbook of Child Language*. Blackwell, Oxford, pp. 251–270.
- Fillmore, C.J., Kay, P., O'Connor, M.C., 1988. Regularity and idiomaticity in grammatical constructions. *Language* 64, 501–538.
- Firth, J.R., 1937/1964. *The Tongues of Men and Speech*. Oxford University Press, London.
- Firth, J.R., 1952–9/1968. In: Palmer, F.R. (Ed.), *Selected Papers of J.R. Firth 1952–59*. Longman, Harlow.
- Flavell, L., Flavell, R., 1992. *Dictionary of Idioms*. Kyle Cathie, London.
- Fraser, B., 1970. Idioms within a transformational grammar. *Foundations of Language* 6, 22–42.
- Garvey, C., 1977. Play with language and speech. In: Ervin-Tripp, S., Mitchell-Kernan, C. (Eds.), *Child Discourse*. Academic Press, New York, pp. 27–47.
- Gazdar, G., Klein, E., Pullum, G., Sag, I., 1985. *Generalized Phrase Structure Grammar*. Blackwell, Oxford.
- Gibbs, R.W., 1991. Semantic analyzability in children's understanding of idioms. *Journal of Speech & Hearing Research* 34, 613–620.
- Givón, T., 1995. *Functionalism and Grammar*. Academic Press, New York.
- Goldberg, A.E., 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL.
- Granger, S., 1998. Prefabricated patterns in advanced EFL writing: collocations and formulae. In: Cowie, A.P. (Ed.), *Phraseology*. Clarendon Press, Oxford, pp. 145–160.
- Grice, H.P., 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), *Syntax and Semantics*, Vol. 3. Academic Press, New York, pp. 41–58.
- Gross, R., 1996. *Psychology: The Science of Mind and Behaviour*, 3rd Edition. Hodder and Stoughton, London.

- Halliday, M.A.K., 1974. A sociosemiotic perspective on language development. *Bulletin of the School of Oriental and African Studies* 37 (1). Reprinted in: Bloom, L. (Ed.), *Readings in Language Development*. Wiley, New York, pp. 256–277 (1978).
- Halliday, M.A.K., 1994. *An Introduction to Functional Grammar*, 2nd Edition. Edward Arnold, London.
- Hanania, E.A.S., Gradman, H.L., 1977. Acquisition of English structures: a case study of an adult native speaker of Arabic in an English-speaking environment. *Language Learning* 27 (1), 75–91.
- Harley, T.A., 1995. *The Psychology of Language: from Data to Therapy*, Psychology Press, Hove.
- Harlow, S., Vincent, N., 1988. Generative linguistics: an overview. In: Newmeyer, F.J. (Ed.), *Linguistics: The Cambridge Survey*. Vol. I. *Linguistic Theory: Foundations*. Cambridge University Press, Cambridge, pp. 1–17.
- Hatch, E., Peck, S., Wagner-Gough, J., 1979. A look at process in child second-language acquisition. In: Ochs, E., Schieffelin, B.B. (Eds.), *Developmental Pragmatics*. Academic Press, New York, pp. 269–278.
- Hess, C.W., Sefton, K.M., Landry, R.G., 1986. Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research* 29, 129–134.
- Hickey, T., 1993. Identifying formulas in first language acquisition. *Journal of Child Language* 20, 27–41.
- Hopper, P.J., 1988. Emergent grammar and the a priori grammar postulate. In: Tannen, D. (Ed.), *Linguistics in Context: Connecting Observation and Understanding*. Lectures from the 1985 LSA/TESOL and NEH Institutes. Ablex, Norwood, NJ, pp. 117–134.
- Hopper, P.J., 1992. Discourse: emergence of grammar. In: Bright, W. (Ed.), *International Encyclopedia of Linguistics*. Oxford University Press, Oxford, pp. 364–367.
- Howarth, P., 1998. Phraseology and second language proficiency. *Applied Linguistics* 19 (1), 24–44.
- Hughlings Jackson, J., 1874/1958. On the nature of the duality of the brain. In: Taylor, J. (Ed.), *Selected Writings of John Hughlings Jackson*, Vol. 2. Staples Press, London, pp. 129–145.
- Hymes, D.H., 1972. On communicative competence. In: Pride, J.B., Holmes, J. (Eds.), *Sociolinguistics*. Penguin, Harmondsworth, pp. 269–293.
- Irujo, S., 1986. A piece of cake: learning and teaching idioms. *ELT Journal* 40 (3), 236–242.
- Jaffe, J., 1978. Parliamentary procedure and the brain. In: Siegman, A.W., Feldstein, S. (Eds.), *Nonverbal Behavior & Communication*. Erlbaum, Hillsdale, NJ, pp. 55–66.
- Kilborn, K., 1991. Selective impairment of grammatical morphology due to induced stress in normal listeners: implications for aphasia. *Brain and Language* 41, 275–288.
- Kjellmer, G., 1984. Some thoughts on collocational distinctiveness. In: Aarts, J., Meijs, W. (Eds.), *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research*. Rodopi, Amsterdam, pp. 163–171.
- Langacker, R.W., 1987. *Foundations of Cognitive Grammar*, Vol. 1., Stanford University Press, Stanford.
- Langacker, R.W., 1991. *Foundations of Cognitive Grammar*, Vol. 2., Stanford University Press, Stanford.
- Lattay, E., 1986. Pragmatic classification of idioms as an aid for the language learner. *IRAL* 24 (3), 217–233.
- Lieven, E.V., Pine, J.M., Barnes, H.D., 1992. Individual differences in early vocabulary development: redefining the referential–expressive distinction. *Journal of Child Language* 19, 287–310.
- Locke, J.L., 1993. *The Child's Path to Spoken Language*. Harvard University Press, Cambridge, MA.
- Locke, J.L., 1994. Gradual development of developmental language disorders. *Journal of Speech and Hearing Research* 37, 608–616.
- Locke, J.L., 1995. Development of the capacity for spoken language. In: Fletcher, P., MacWhinney, B. (Eds.), *The Handbook of Child Language*. Blackwell, Oxford, pp. 278–302.
- Locke, J.L., 1997. A theory of neurolinguistic development. *Brain and Language* 58, 265–326.
- McCrone, J., 1999. States of mind. *New Scientist* 2178 (20 March), 30–33.
- Mackay, D.G., 1979. Lexical insertion, inflection, and derivation: creative processes in word production. *Journal of Psycholinguistic Research* 8, 477–498.
- Makkai, A., 1972. *Idiom Structure in English*. Mouton, The Hague.
- Manschreck, T.C., Maher, B.A., Hoover, T.M., Ames, D., 1984. The type–token ratio in schizophrenic disorders: clinical and research value. *Psychological Medicine* 14, 151–157.
- Martins-Baltar, M., 1998. Critères du caractère usuel d'un énoncé: l'approche motivationnelle du dictionnaire 'Dicomotus'. Paper presented at the 6th International Pragmatics Conference, Reims, France, 19–24 July 1998.

- Miyake, A., Carpenter, P.A., Just, M.A., 1994. A capacity approach to syntactic comprehension disorders: making normal adults perform like aphasic patients. *Cognitive Neuropsychology* 11, 671–717.
- Moon, R., 1992. Textual aspects of fixed expression in learners' dictionaries. In: Arnaud, P.J.L., Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. Macmillan, Basingstoke, pp. 12–27.
- Moon, R., 1998. *Fixed Expressions and Idioms in English*. Clarendon Press, Oxford.
- Nattinger, J.R., DeCarrico, J.S., 1992. *Lexical Phrases and Language Teaching*. Oxford University Press, Oxford.
- Nelson, K., 1981. Individual differences in language development: implications for development and language. *Developmental Psychology* 17 (2), 170–187.
- Nippold, M.A., Martin, S.T., 1989. Idiom interpretation in isolation versus context: a developmental study with adolescents. *Journal of Speech and Hearing Research* 32, 59–66.
- Nunberg, G., Sag, I.A., Wasow, T., 1994. Idioms. *Language* 70, 491–538.
- Paradis, M., 1997. The cognitive neuropsychology of bilingualism. In: DeGroot, A.M.B., Kroll, J.F. (Eds.), *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Erlbaum, Mahwah, NJ, pp. 331–354.
- Pawley, A., Syder, F.H., 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In: Richards, J.C., Schmidt, R.W. (Eds.), *Language and Communication*. Longman, New York, pp. 191–226.
- Perkins, M.R., 1983. *Modal Expressions in English*. Frances Pinter, London.
- Perkins, M.R., 1994. Repetitiveness in language disorders: a new analytical procedure. *Clinical Linguistics and Phonetics* 8 (4), 321–336.
- Perkins, M.R., in press. Productivity and formulaicity in language development. In: Schelletter, C., Letts, C., Garman, M. (Eds.), *Issues in Normal and Disordered Child Language: From Phonology to Narrative*. Special issue of *The New Bulmershe Papers*, University of Reading.
- Peters, A.M., 1977. Language learning strategies: does the whole equal the sum of the parts? *Language* 53 (3), 560–573.
- Peters, A.M., 1983. *Units of Language Acquisition*. Cambridge University Press, Cambridge.
- Pinker, S., 1998. Words and rules. *Lingua* 106, 219–242.
- Pollard, C. and Sag, I., 1994. *Head-driven Phrase Structure Grammar*. Report No. CSLI-88-132. Center for the Study of Language and Information, University of Chicago, Chicago, IL.
- Prizant, B.M., 1983. Language acquisition and communicative behavior in autism: toward an understanding of the 'whole' of it. *Journal of Speech and Hearing Disorders* 48, 286–296.
- Raichle, M.E., 1998. The neural correlates of consciousness: an analysis of cognitive skill learning. *Philosophical Transactions of the Royal Society of London, Series B* 353, 1889–1901.
- Rampton, B., 1987. Stylistic variability and not speaking 'normal' English: some post-Labovian approaches and their implications for the study of interlanguage. In: Ellis, R. (Ed.), *Second Language Acquisition in Context*. Prentice Hall, Englewood Cliffs, NJ, pp. 47–58.
- Rehbein, J., 1987. Multiple formulae: Aspects of Turkish migrant workers' German in intercultural communication. In: Knapp, K., Enninger, W., Knapp-Potthoff, A. (Eds.), *Analysing Intercultural Communication*. Mouton, Berlin, pp. 215–248.
- Reiss, N., 1989. Speech act taxonomy, chimpanzee communication, and the evolutionary basis of language. In: Wind, J., Pulleybank, E.G., De Grolier, E., Bichakjian, B.H. (Eds.), *Studies in Language Origins*, Vol. 1. John Benjamins, Amsterdam, pp. 283–304.
- Renouf, A., 1992. What do you think of that? A pilot study of the phrasology of the core words of English. In: Leitner, G. (Ed.), *New Directions in English Language Corpora: Methodology, Results, Software Developments*. Mouton de Gruyter, Berlin, pp. 301–317.
- Renouf, A., Sinclair, J., 1991. Collocational frameworks in English. In: Aijmer, K., Altenberg, B. (Eds.), *English Corpus Linguistics: Studies in the Honour of Jan Svartvik*. Longman, London, pp. 128–143.
- Sapir, E., 1921. *Language*. Harcourt, Brace, Jovanovitch, New York.
- Saussure, F., 1916/1966. *Course in General Linguistics*. McGraw-Hill, New York.
- Schmidt, R.W., 1983. Interaction, acculturation, and the acquisition of communicative competence: a case study of an adult. In: Wolfson, N., Judd, E. (Eds.), *Sociolinguistics and Language Acquisition*. Newbury House, Rowley, MA, pp. 137–174.
- Sinclair, J., 1991. *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.

- Sperber, D., Wilson, D., 1995. *Relevance: Communication and Cognition*, 2nd Ed. Blackwell, Oxford.
- Stubbs, M., 1997. Eine Sprache idiomatisch sprechen: Computer, Korpora, kommunikative Kompetenz und Kultur. In: Mattheier, K.J. (Ed.), *Norm und Variation*. Peter Lang, Frankfurt am Main pp. 151–167.
- Tannen, D., 1989. *Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse*. Cambridge University Press, Cambridge.
- Templin, M.C., 1957. *Certain language skills in children: their development and interrelationships*. Greenwood Press, Westport, VA.
- Tomasello, M., 1992. *First Verbs: A Case Study of Early Grammatical Development*. Cambridge University Press, Cambridge.
- Tomasello, M. (Ed.), 1998. *The New Psychology of Language*. Lawrence Erlbaum, Mahwah, NJ.
- Van Lancker, D.R., 1987. Nonpropositional speech: neurolinguistic studies. In: Ellis, A.W. (Ed.), *Progress in the Psychology of Language*, Vol. 3. Lawrence Erlbaum, Hillsdale, NJ, pp. 49–118.
- Van Lancker, D.R., Kempler, D., 1987. Comprehension of familiar phrases by left- but not by right-hemisphere damaged patients. *Brain and Language* 32, 265–277.
- Verstraten, L., 1992. Fixed phrases in monolingual learners' dictionaries. In: Arnaud, P.J.L., Béjoint, H. (Eds.), *Vocabulary and Applied Linguistics*. Macmillan, Basingstoke, pp. 28–40.
- Wachal, R.S., Spreen, O., 1973. Some measures of lexical diversity in aphasic and normal language performance. *Language and Speech* 16, 169–181.
- Webelhuth, G., 1995. X-bar theory and case theory. In: Webelhuth, G. (Ed.), *Government and Binding Theory and the Minimalist Program*. Blackwell, Cambridge, MA, pp. 15–95.
- Weinert, R., 1995. The role of formulaic language in second language acquisition: a review. *Applied Linguistics* 16 (2), 180–205.
- Weinreich, U., 1969. Problems in the analysis of idioms. In: Puhvel, J. (Ed.), *Substance and Structure of Language*. University of California Press, Los Angeles, CA, pp. 23–82.
- Widdowson, H.G., 1989. Knowledge of language and ability for use. *Applied Linguistics* 10 (2), 128–137.
- Wong Fillmore, L., 1976. The second time around: cognitive and social strategies in second language acquisition. Unpublished PhD thesis, Stanford University.
- Wray, A., 1992. *The Focusing Hypothesis: The Theory of Left Hemisphere Lateralised Language Re-Examined*. John Benjamins, Amsterdam.
- Wray, A., 1998. Protolanguage as a holistic system for social interaction. *Language & Communication* 18, 47–67.
- Wray, A., 1999. Formulaic language in learners and native speakers. *Language Teaching* 32 (4), 213–231.
- Wray, A., in press a. Holistic utterances in protolanguage: the link from primates to humans. In: Knight, C., Studdert-Kennedy, M., Hurford, J. (Eds.), *The Evolutionary Emergence of Language*. Cambridge University Press, Stanford, CA.
- Wray, A., in press b. Formulaic sequences in second language teaching. *Applied Linguistics*.
- Wray, A., in preparation a. Understanding the learner's 'task' in Task Based Learning: the significance of formulaic language in preparation.
- Wray, A., in preparation b. Formulaically speaking, Cambridge University Press, Stanford, CA.
- Yorio, C.A., 1980. Conventionalized language forms and the development of communicative competence. *TESOL Quarterly* 14 (4), 433–442.
- Youmans, G., 1991. A new tool for discourse analysis: the vocabulary–management profile. *Language* 67, 763–789.
- Yule, G.U., 1944. *The statistical study of literary vocabulary*. University Press, Cambridge, MA.