

# Formulaic language: its characteristics and how it is used and acquired

Takeshi Matsuzaki

## Abstract

This paper considers the various definitions given for ready-made multi-word expressions referred to *formulaic sequences* (henceforth, FSs), examining the different constructs of FSs for the purpose of better understanding their functionality in communication. It also endeavors to clarify how we process and acquire chunks of language. Implications will be provided for the ways in which FSs complicate second language learning, particularly for adult learners in a foreign as opposed to second language environment.

## 1. Introduction

Formulaic language abounds in language use, and a number of studies have shown its pervasiveness. For example, in Foster's investigation (2001), 32.3% of the unplanned native speech analyzed was judged to consist of formulaic language. As another example, Erman and Warren (2000) categorized 58.6% of the spoken and 52.3% of the written English discourse that they examined as formulaic word strings of various kinds. In regard to the percentages, Altenberg (1990) further suggests that if the enormous set of simple lexical collocations that cannot be elegantly categorized from a formal grammatical point of view is regarded as part of formulaic language, then, possibly as much as 70% of adult native language may be formulaic. A range of corpus studies (e.g. Altenberg, 1993; Baayen & Lieber, 1991; Barkema, 1993; Kjellmer, 1984) have also demonstrated that most ordinary language

production, written or spoken, appears to be composed largely of collocational sets or frameworks, manifesting far less variability than could be possible on the basis of grammar and lexicon alone. Taken together, as Sinclair (1991) puts it, “all the evidence points to an underlying rigidity of phraseology, despite a rich superficial variation” (p. 121).

When it comes to the learning of formulaic language in an L2 (second language), formulaicity, pervasive and integral as it is, remains an area where L2 learners only very slowly approximate to or will never reach the proficiency of native speakers. Results of a number of studies (Altenberg & Granger, 2001; Bardovi-Harlig, 2009; Durrant & Schmitt, 2009; Howarth, 1996; Kaszubski, 2000; Laufer & Waldman, 2011; Li & Schmitt, 2010; Nekrasova, 2009; Qi & Ding, 2011; Siyanova & Schmitt, 2007; Siyanova & Schmitt, 2008; Yamashita & Jiang, 2010) reveal that L2 learners even at advanced levels of proficiency are unable to produce formulaic sequences in the L2 that are comparable to those used by native speakers. Pawley and Syder (1983) observe that it is often the failure to utilize natively-like formulaic sequences that ultimately distinguishes the advanced L2 learner as non-native.

According to the thorough search for past observations of formulaic language by Allison Wray, author of the seminal book “Formulaic Language and the Lexicon,” the existence of this linguistic phenomenon was recognized as early as the mid-nineteenth century (Wray, 2002). Narrowing down the scope of her search to the past half century, she found that the first discussions on the significance of formulaic language are Bolinger (1976), Fillmore (1979), and Pawley and Syder (1983).<sup>1</sup> Their critiques were then followed up by Sinclair (1991) with his well-known ‘idiom principle,’ while the exploration of the relationship between lexical phrases and functional language was commenced by Nattinger and DeCarrico (1992). Wray (2002, p. 9) also offers a list of over fifty terms denoting the phenomenon of formulaic language. The list includes some everyday words used by ordinary people (e.g., *idioms*, *formulae*, *clichés*), terms that we do not see fully addressed in current literature, such as *amalgams*, *fossilized forms*, *frozen phrases*, *gambits*,

*gestalt*, *holophrases*, and those that seem to be preferred by present linguistic specialists, which include the following:<sup>2</sup>

*chunks / constructions / collocations / conventionalized forms / fixed expressions / formulaic language / lexical phrases / lexicalized sentence stems / multi-word items (units) / non-compositional / prefabricated routines and patterns / ready-made expressions / recurring utterances / sentence builders*

In many ways, formulaic sequences accomplish the same functions as single words. Many (e.g., collocations: *tie your shoes*, *still waters*; and complex verbs: *run over*, *break it down*) have mainly a referential or ideational purpose and thus operate as content words do. Others (e.g., exclamations: *Are you serious*, *no way*; and idioms: *back to the drawing board*, *far cry from*) are particularly effective for portraying an evaluative stance. Some ensure effortless social interaction (pragmatic formulae such as *Good to see you* and *I'm really happy for you*), while others are similar to function words in that they act, for example, to unify discourse (e.g., *as a side note*, *to offer a different perspective*). Collectively they make up a substantial and vital part of one's lexicon, performing an essential role in facilitating the understanding and expression of messages that could otherwise be misinterpreted.

## 2. Characteristics of formulaic language

Formulaic language has been studied from diverse perspectives, resulting in a variety not only of criteria or definitions to describe the phenomenon but also of terminology (Schmitt & Carter, 2004). Corpus linguistics is generally concerned with the identification and description of formulaic sequences as they are found in various kinds of corpus data (Schmitt & Carter, 2004). Criteria that are commonly used in this field of inquiry include *institutionalization*, *fixedness*, *non-compositionality*, and *frequency of occurrence*.<sup>3</sup> Psycholinguists and language acquisition specialists, in contrast, employ criteria such as

whether a word string is used more than once by a participant (suggesting that the use is not so much a single, one-time-only imitation as a manifestation of the participant's proceduralized knowledge) and whether the production is accompanied by an intact intonation contour (indicating that the sequence is stored and retrieved as a whole) (Schmitt & Carter, 2004).

Wray (2002) provides the following inclusive, umbrella definition of a formulaic sequence:

a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

This is a definition that is very extensive in its coverage, applicable to the entire spectrum of different types of word strings. These include, for example, tightly idiomatic and immutable strings (e.g., *by and large*) at one end of the spectrum, and range to transparent and flexible strings with slots for open class items such as *NP be-TENSE sorry to keep-TENSE you waiting* at the other. Broadly speaking, formulaic sequences can usefully be defined as strings of linguistic items where the relation of each item to the rest is relatively fixed, and where the substitutability of one constituent of the sequence by another of the same category is relatively constrained (Perkins, 1999). An additional, and essential, component of this definition includes “[w]ords and word strings which appear to be processed without recourse to their lowest level of composition” (Wray, 2002, p. 4).

Because there is so much diversity in the use of formulaic sequences, it is difficult to agree upon absolute criteria to define them. Taking into account the proposal by Schmitt and Carter (2004) that even though each particular example may not manifest all characteristics, it seems helpful to discuss the typical characteristics of formulaic sequences, the following sub-sections overview their distinctive features from formal, semantic, and functional perspectives.

### 2.1. *Formal mutability of formulaic sequences*

The major formal facet of formulaic manifestations is that they can be dichotomized according to whether the component words are totally invariable or only partially fixed.<sup>4</sup> Starting with frozen, immutable strings of words, one of their syntactic characteristics or irregularities is that some sequences are subject to an inflectional or transformational restriction (Verstraten, 1992). Wray and Perkins (2000) provide examples of this type, cited from other sources. For example, it is not possible to pluralize *beat around the bush* or passivize *face the music* without the strings losing their idiomatic meaning (Flavell & Flavell, 1992, p. 6). Additionally, you *slept a wink last night* or to make someone *fed up by feeding them up* are not possible variants (Irujo, 1986, p. 237). Another syntactic irregularity of the fixed FSs is that some do not even follow normal restrictions. Two examples of this kind of irregularity, again cited by Wray and Perkins, are to *come a cropper* and to *go the whole hog* (Flavell & Flavell, 1992, p. 7), in each of which case an intransitive verb is followed by a direct object. Another example of Wray and Perkin's is *by and large*, in which non-identical constituents are juxtaposed. With such syntactic oddities, however, fixedness is undoubtedly advantageous to both the speaker and the hearer. An example provided by Schmitt and Carter (2004) is *Watch Out!* Even though a sentence with more contextual information like *Watch out for the car coming behind you!* could also be generated and understood, the speaker is inclined to choose, when milliseconds count, a shorter and more customary warning that does not call for extended realtime computation by the intended recipient, and the core message of the language so chosen is also likely to be readily conveyed to the hearer, which is also the speaker's intention.

Moving on to flexible formulaic sequences, the constituents of a flexible formulaic sequence are a varied number of prefabricated 'frames' and 'slots' for flexibility of use to be filled by applicable words or strings of words (Nattinger & Decarrico, 1992), although the slots typically have semantic constraints. For example, when we would like

to convey the idea that some action or accomplishment is irregular, unbelievable or extraordinary, we are able to utilize statements such as *He stood in disbelief, as the magician sawed the woman in half* or *They watched in disbelief, as the woman dove from the 100 meter cliff into the ocean*. The fundamental composition of these two sentences is the frame ‘\_\_\_ in disbelief, as \_\_\_’, and the second slot of the frame offers the possibility of expressing something unexpected in a wide variety of contexts. This scaffold can be an aid to fluent language, because some of the language is already pre-assembled and can be called on in diverse situations. Bear in mind, however, that the second slot must, in normal circumstances, convey the idea of something unusual, unbelievable or unexpected, precisely because that is the reason or purpose for using this formulaic sequence. Thus, a sentence like *She listened in disbelief, as the radio announcer read the advertisements* renders the whole peculiar because the reading of the advertisements by an announcer is an everyday occurrence. The fact that it is theoretically possible that the context will make the sentence acceptable to the listener does not detract from the point being made about the core characteristic of this frame. The semantic limitations of such pre-assembled frames appear to leave them with sufficient flexibility and adaptability within a wide range of contexts as to make them widely used in discourse.<sup>5</sup>

Looking at the formal attributes of formulaic sequences from a bottom-up perspective, it is certainly fair to observe that certain words (especially adjectives and verbs) rather than a string of words are constrained by particular syntactic structures. An example introduced by Schmitt and Carter (2004) is the adjective *rife*. As this is a predicate adjective (with a negative connotation), a typical sentence structure in which this word is embedded is *SOMETHING UNDESIRABLE is/are rife in LOCATION/TIME*. Thus, while from a formulaic sequence perspective the collocation of the frame (i.e., *is/are rife in*) and the two semantically constrained slots (*SOMETHING UNDESIRABLE* and *LOCATION/TIME*) together constitute one formulaic sequence, it is also possible to point to the structural rules of the single word *rife*. Which way to interpret the phenomenon is dependent on how one observes

and analyzes the language. Yet, from the language processing perspective, the more holistic approach seems far more beneficial to the language user/learner (see below).

## **2.2. *Semantic transparency of formulaic sequences***

Some types of formulaic sequences are semantically distinct. Prime examples are idioms, proverbs, sayings, and phrasal verbs. While consisting of multiple orthographic words, these sequences evidently operate as single units. The fact that these multi-word units express a single meaning makes them stand out. In the case of metaphoric word strings, component words have relinquished their respective semantic meanings (in some cases syntactic rules too) in favor of the collective, holistic meaning assumed in combination with the rest of the string (Moon, 1992; Nattinger & DeCarrico, 1992, chapter 2; Yorio, 1980). In other words, the meaning cannot be derived from the sum of meanings of the component words. These types of word strings are collectively referred to as non-compositional formulaic sequences. It would be impossible for a hearer to understand these for the first time without substantial pragmatic or explanatory context (e.g., *kick the bucket*; *hot potato*), although there are some cases where the metaphorical meaning can be derived with less guessing (e.g., *from the cradle to the grave*; *hit the nail on the head*).

Thus far, characteristics of formulaic sequences have been described in terms of surface form and meaning. However, classifications depending solely on formal and semantic aspects are sometimes not completely clear. For one thing, most proverbs are semantically incomprehensible, and would be classified as idioms on this basis (e.g., *A bird in the hand is worth two in the bush*), so what is the difference between proverbs and idioms? One useful way of differentiating the two is their conditions of use, or pragmatic functions. The next sub-section deals with this functional facet of formulaic language.

## **2.3. *Pragmatic functions of formulaic sequences***

Formulaic sequences are often tied to particular conditions of use,

or pragmatic functions. To answer the question raised in the previous sub-section concerning the difference between formal and semantic aspects of FSs involving idioms and proverbs, idioms are typically used to express a concept (e.g., *play it by ear* = adjust one's actions to fit the situation), while proverbs are usually about some commonly believed truth and thus used as advice (e.g., *Two wrongs don't make a right* = an admonition to not seek revenge).

The ways in which recurring situations in the social world require particular language from people are often illustrated in terms of the functions that are fulfilled by that language (Schmitt & Carter, 2004). For example, speech acts such as apologizing, making requests, giving directions, and complaining typically have conventionalized language forms attached to them (e.g., *I'm (very) sorry to hear about \_\_\_\_* to express sympathy and *I'd be happy/glad to \_\_\_\_* to comply with a request) (Nattinger & DeCarrico, 1992, pp. 62-63). Another typical function performed by formulaic sequences is that of organizing the discourse. Logical connectors are abundant in discourse, both spoken and written (e.g., *Having said that, Specifically, On the contrary, Speaking of which, Such being the case*). Yet another common function served by formulaic sequences is maintenance of social interaction. We participate in casual and light conversation just to pass the time of day or for amusement, so engaging in such communication is unlikely to involve serious attempts to exchange information or to manipulate someone into doing something. The content per se is not as important as the existence of some communication, superficial though it may be. To handle such a situation, we rely on a set of conventionalized social phrases that are non-threatening in any way and support the flow of the conversation. Examples include comments about the weather (*Beautiful day, isn't it?*), agreeing with the interlocutor (*You're right*), providing backchannels and positive feedback to another speaker (*Uh-huh; That's great*). Kecskes (2003) points out that such sequences serve as a social lubrication as well as an active co-constructing device for interpersonal communication. One feature all these examples have in common is that members of a speech community know these expressions, and this

makes it possible for them to serve as a quick and reliable vehicle for the desired function.

In relation to this functional facet of formulaic sequences, Wray and Perkins (2000) provide an iconoclastic account of how they serve us in language use (for details, see pp. 13–19). According to Wray and Perkins, there exist two fundamental determiners of a person's preference for a formulaic, holistic expression over an analytic, generative expression (see Section 3 for details) at any given moment: these are the socio-interactional priorities and the constraints on our processing capabilities. With regard to this dichotomy, Wray and Perkins propose that the functions of FSs as devices of social interaction are (1) "manipulation of others," (2) "asserting separate identity," and (3) "asserting group identity," and the functions they serve as compensatory devices for memory limitations are (1) "processing shortcuts," (2) "time-buyers," and (3) "manipulation of information." A sage observation offered by Wray and Perkins here is that "these two [seemingly unrelated purposes for formulaic language] are in actual fact two sides of the same coin" (p. 17). They explain:

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. . . . 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. 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-interactive functions ... In both cases, it is the speaker who benefits from using formulaic sequences. (p. 18)

### 3. Formulaic nature of our language knowledge and processing

Given the widespread use of formulaic sequences in discourse, a number of scholars have argued that proficient language users must have an extensive knowledge and command of these sequences. Pawley and Syder (1983, p. 213), for instance, suggest that the number of "sentence-length expressions familiar to the average, mature English speaker probably amounts, at least, to several hundreds of thousands." In a similar vein, Jackendoff (1995) postulates, based on a small corpus study of spoken language in a TV quiz show, that the significance of formulaic sequences may be equal to, if not greater than, the lexicon of single words. While Schmitt and Carter (2004) point out that these assertions are not supported by enough empirical work, there is some evidence to suggest that formulaic sequences are generally processed as unitary wholes and, as a corollary, stored in memory as such, even if this is not the case for every instance.<sup>6</sup> For example, Kuiper (1996, 2004) and his colleagues (Kuiper & Haggio, 1984) demonstrate that smooth talkers (auctioneers, sportscasters) rely heavily on formulaic language as a means of fluently conveying large amounts of information under severe time pressure.

Before the advent of computerized corpus studies, our great capacity to remember and use prefabricated units was underestimated on the one hand, and the extent to which we (can) process language by complex processes of calculation was overestimated on the other (Lamb, 1999, p.169). Until then, multi-word units to enable fast processing were acknowledged but often relegated as a peripheral phenomenon that plays only a minor role in language (Wray, 2002). With more and more such studies, however, corpus linguistics has revealed the pervasiveness of formulaicity, in its widest sense, in corpora (reviewed by Wray, 2002,

chapter 2), and now, “[t]he real issue is whether it is, or isn’t, possible to account for real language data without invoking prefabrication” (Wray, 2002, p. 12). As a consequence, the Chomskyan view that the language of normal adult native speakers is processed piecemeal in output production and input comprehension has been under severe attack. There is no doubt that we are capable of grammatical processing, but it has been made clear that such processing is not our only, nor even our preferred, way of handling language production and comprehension.<sup>7</sup> On the contrary, much of our input and output is processed holistically, albeit analyzable, and manifests far less variability than could be predicted on the basis of grammar.<sup>8</sup> On the issue of nativelike selection and fluency, Pawley and Syder (1983) claim:

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

The corpus linguist John Sinclair was one of the first researchers to introduce the distinction between holistic processing and analytic processing, with his ‘idiom principle’ and ‘open choice principle’ (Sinclair, 1991). The idiom principle posits that “a language user has available to him a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments” (Sinclair, 1991, p. 110). This principle brings about the selection of two or more words together, on the basis of previous and frequent co-occurrence. The open choice principle, conversely, states that “syntax is there to specify the slots into which memorised items – normally single words – can be inserted” (Warren, 2005, p. 36). That is, the open choice principle results in the selection of single words, and gives

interlocutors the same kind of creative freedom as the Chomskyan account. As for the operation of these principles, Sinclair (1991) proposes:

the first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle, and quickly back again. Lexical choices which are unexpected in their environment will presumably occasion a switch. (p. 114)

To put it another way, our baseline strategy in normal language processing, whether in production or 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). Importantly, the operation of holistic processing (according to the idiom principle) is not restricted to only, say, those non-compositional multi-word strings such as idioms, which cannot be generated or comprehended with the operation of analytic processing (according to the open choice principle), but can also deal with linguistic manifestations for which analytic processing would have rendered exactly the same outcomes (Wray, 1992, 2002).

#### 4. Acquisition of formulaic sequences

As seen in the previous section, adults' knowledge of L1 is considered to be largely holistic. When it comes to the acquisition of formulaic sequences, however, the amount of research into this phenomenon has been fairly modest (Schmitt & Carter, 2004; Weinert, 1995; Wray, 2002). Nevertheless, it has been suggested that acquisition of each holistic sequence does not appear to take place at a single point in time. Rather, it appears that the mastery of each particular formulaic sequence is realized in a gradual, rather than all-or-nothing, manner. For instance, L1 acquirers seem to construct the phonological mappings of a formulaic sequence starting from the whole sequence and then analyzing it into components, but with some elements still incompletely

cognized, particularly in the case of unstressed phonemic constituents; later on the gaps in the initial stages of the rendering of the sequence will be fulfilled (Peters, 1977; Schmitt & Carter, 2004; Wray, 2002, Chapter 6). For another example, transparent sequences such as *my point (here) is that* \_\_\_\_ are perhaps even generated automatically or somewhat unconsciously in the first instance through knowledge of the component words and syntactic knowledge, and the newly constructed sequence in this manner is stored as a single multi-word unit in holistic language knowledge. It is proposed that it is in these ways that humans acquire formulaic sequences in their L1 over time.<sup>9</sup>

In the case of L2 learning, navigating the route of acquisition of formulaic sequences is far more complicated, because of the wide diversity of conditions for learning. “There may well be an underlying systematicity to the acquisition and use of L2 formulaic language, but there is simply not enough focused research at present to say very much with conviction” (Schmitt & Carter, 2004, p. 13). One certainty is the incompleteness of the ultimate learning outcome, lexically as well as grammatically (Abrahamsson & Hyltenstam, 2009; see also the Introduction). Such being the case, instead of seeking further to identify and describe the underlying route—if there is one—for formulaic language development in the L2, the focus here is placed on why learning formulaic language in the L2, especially for adults, is so consistently difficult.

There is discussion of this issue in Wray (2000) and Wray and Perkins (2000). According to Wray and Perkins, children, as opposed to adults, operate within a “socio-interactional bubble ... both protected from, and largely impervious to, any need to interact with anyone other than its carers” (Wray & Perkins, 2000, p. 22), and “by being protected from the intellectual and emotional stress of interacting in the world beyond the bubble,” they “can apply analytical processes to derive grammatical and lexical information from formulaic sequences” (Wray, 2000, p. 481). Adults, in contrast, have to handle the whole variety of socio-interactional demands in communication, and thus, on the basis that children can extract underlying linguistic information from formulaic sequences, “it would be unwise to assume that ... adults can too”

(Wray, 2000, p. 481; also compare VanPatten, 1990).

Finally, one relevant theoretical account on language acquisition to this paper is the development of pattern-based models concerning the acquisition of language, which suggest that the human capacity for language learning stems from the ability to isolate structures from a given response, instead of being under the control of instinctive determinants and constraints that supposedly predetermine which aspects of a given language may or may not be acquired at a given moment in the learning process (see Ellis, 1996, 2002). This theory proposes that we acquire the character or letter orders that are acceptable in a language (e.g., the consonant cluster *sp* can be word-initial in English, but *hg* cannot) simply by continually viewing *sp* at the beginning of words, but not *hg*. This learning is implicit, and may not be relative to conscious metalinguistic accounts of acquisition. Of course, learners may ultimately reach the point where they can conclude that there exists a 'rule' for this specific consonant clustering; however, the rule has been inductively constructed from pattern-based acquisition or experience, rather than learned through the exercise of an innate language rule acquisition capacity.

This pattern-based learning also pertains to more extensive linguistic units. For example, we gain insight into which words collocate together and which do not (e.g., *blonde hair*, *\*blonde paint*; *auburn hair* but only for women, not men). Many of these associations essentially stem from pattern recognition, as there is frequently no semantic reasoning that conveys which pairings are acceptable and which are unacceptable (*\*blonde paint* makes perfect logical sense). Collocations are not likely to be learned explicitly either, since they are not typically taught, and even if they are, only probable instances are exemplified, not inappropriate sequences. Longer formulaic series, which are also based on patterns rather than rules, seem to follow rather suitably with such sequence-based models of acquisition as well.

## 5. Summary and concluding remarks

This paper has sought to provide a sketch of the main characteristics of formulaic language. Formulaic language is, as demonstrated, a multi-faceted phenomenon entailing complexities at all formal, semantic, and pragmatic levels. Major points are summarized in Table 1. The pragmatic functional side, following Wray and Perkins' (2000) model, offers a particularly useful way to conceptualize the entire phenomenon of formulaic expressions.

**Table 1** *Characteristics of Formulaic Sequences*

Aspect	Dichotomy	Additional notes
Formal	Invariable	Syntactically constrained; some peculiar syntagmatic pairings
	Flexible	Frames with slots (often semantically constrained)
Semantic	Compositional	
	Non-compositional	Metaphorical
Functional	Socio-interactional	Main functions: 1) manipulation of others; 2) assertion of separate identity; 3) assertion of group identity
	Memory compensatory	

This paper has also looked at Sinclair's account of how we process chunks of language. Of course, we can and do create and understand novel language, which has been the thrust of the Chomskyan tradition for the last 50 years. To acknowledge a central role for formulaic sequences that are processed holistically and stored as such in the memory system is not to exclude our capability to handle novelty and creativity, "only to relegate it from the position of sole strategy" (Wray, 1992, p. 17). Analytic processing, then, "could be imagined as a ... process which goes on in principle all the time, but whose results are only intermittently called for" (Sinclair, 1991, p. 114). This dual processing/knowledge model is proposed as one of the most reasonable ways of accommodating and accounting for both the holistic and analytic

features of language.

Finally, this paper has addressed how we acquire formulaic language, especially the learning difficulty that adult L2 learners face in acquiring such language. Essentially, there exist two tremendous challenges for these learners. First, if they are to maximize their potential for L2 learning during realtime communication, they will have to somehow overcome the processing demands resulting from the lack of the socio-interactive bubble. Second, if the pattern-based account of language learning holds true, then, its implications are significant: regardless of what is to be learned, a formulaic sequence or a grammatical rule, a pattern must be extracted. It does not require a great deal of imagination to understand how potentially challenging a task that might be for an L2 learner, especially an adult and especially in a foreign language context: Such a learner simply does not enjoy sufficient enough encounters with the given language to derive any intrinsic pattern from them, or even if they did, they lack opportunities to strengthen the knowledge so learned.

**Acknowledgements:** I am very grateful to Hideyuki Takashima and Kevin Mark for their substantial and insightful advice in the writing of this paper.

#### Notes

- 1 For other publications on formulaic language prior to Bolinger, see Wray (2002, pp. 7–8).
- 2 Wray (2000) cautions against the assumption that researchers are dealing with very much the same phenomenon with various terms (Wray, 2002). As observed by Wray and Perkins (2000), “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” (p. 3).
- 3 The need for caution in using frequency to identify a formulaic sequence has been recognized (Hickey, 1993; Wray & Perkins, 2000). While there is

no doubt that there is some sort of connectedness between a sequence being frequent in a corpus and the conventionalized status accorded to it by a given speech community, this connectedness may actually be incidental: “[i]t has yet to be established that commonness of occurrence is more than a circumstantial associate” (Wray & Perkins, 2000, p. 7).

- 4 Another most obvious formal characteristics of FSs is, perhaps, their varying lengths. Very short sequences can be composed of two words (e.g., *Come on!*). FSs can be very long too, as in lengthy proverbs (e.g., *The grass is always greener on the other side of the fence*).
- 5 Semantic constraints on slots of formulaic frames can be broadly discussed in terms of ‘semantic prosody,’ a notion introduced by Sinclair (1991, 2004). According to the idea of semantic prosody, certain seemingly neutral words can be perceived with positive or negative associations through frequent occurrences with particular collocations. A prime example is the phrasal verb *set in*, which has a negative prosody (e.g., *A plague is going to set in*). Another note on flexible FSs is that their semantic constraints are difficult to identify using current concordancing packages. Schmitt and Carter (2004) point out that modern concordancers are good at identifying contiguous sequences in corpora, but convenient software to automatically identify flexible formulaic sequences has yet to be developed.
- 6 Actually, Schmitt and Carter (2004) comment that these claims may not even require empirical studies to substantiate them, as the most obvious evidence lies in semantically-opaque, non-compositional formulaic sequences (see Section 2) where their aggregated meaning cannot be derived from knowledge of the component words, because the only way to know the meaning of the idiom is to have learned it as a whole unit.
- 7 On our generative capability, Wray (2002, p. 12) also points out that “in most cases ‘novelty’ is much less a question of doing things with grammar than juxtaposing new ideas in commonplace grammatical frames,” and thus “[m]ost of our language ... is novel in a rather uninteresting way.”
- 8 Wray and Perkins (2000) further chastise the Chomskyan view on two grounds. First, they point out that the Chomskyan view holds that “all sequences of words ... which *can* be assembled by rule, *must* be assembled by rule” (p. 10). Such a view is not agreeable to corpus linguists and scholars studying formulaic language. Second, and more problematically, Wray and Perkins criticize the corollary of that view: “all grammatical sequences are equally valid and equally likely to occur” (p. 10).

- 9 For a comprehensive account of the acquisition of formulaic language in the L1, see Wray (2002).

### References

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59(2), 249-306. <http://doi.org/10.1111/j.1467-9922.2009.00507.x>
- Altenberg, B. (1990). Speech as linear composition. In G. Caie, K. Haastrup, A. L. Jakobsen, J. E. Nielsen, J. Sevaldsen, H. Specht, & A. Zettersten, *Proceedings from the Fourth Nordic Conference for English Studies, Vol. 1. Department of English* (pp. 133-143). University of Copenhagen.
- Altenberg, B. (1993). Recurrent verb-complement constructions in the London-Lund Corpus. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi.
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173-195. <http://doi.org/10.1093/applin/22.2.173>
- Baayen, H., & Lieber, R. (1991). Productivity and English derivation: A corpus-based study. *Linguistics*, 29, 801-843.
- Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, 59(4), 755-795. <http://doi.org/10.1111/j.1467-9922.2009.00525.x>
- Barkema, H. (1993). Idiomaticity in English NPs. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, 1, 1-14.
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL*, 47(2), 157-177.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18(01), 91-126. <http://doi.org/10.1017/S0272263100014698>
- Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition.

- Studies in Second Language Acquisition*, 24(02), 143–188. <http://doi.org/10.1017/S0272263102002024>
- Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 29–62.
- Fillmore, C. (1979). On fluency. In C. J. Fillmore, D. Kempler, & W. S-Y. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 85–101). New York: Academic Press.
- Flavell, L., & Flavell, R. (1992). *Dictionary of Idioms*. London: Kyle Cathic.
- Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching Pedagogic Tasks* (pp. 75–93). Harlow: Longman.
- Hickey, T. (1993). Identifying formulas in first language acquisition. *Journal of Child Language*, 20(01), 27–41. <http://doi.org/10.1017/S0305000900009107>
- Howarth, P. A. (1996). *Phraseology in English Academic Writing: Some Implications for Language Learning and Dictionary Making*. Tübingen: Niemeyer.
- Irujo, S. (1986). A piece of cake: Learning and teaching idioms. *ELT Journal*, 40(3), 236–242. <http://doi.org/10.1093/elt/40.3.236>
- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk, R. Schreuder, & R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives* (pp. 133–166). Hillsdale NJ: Erlbaum.
- Kaszubski, P. (2000). Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach. *Unpublished PhD Dissertation, Adam Mickiewicz University, Poznań*.
- Kecskés, I. (2003). *Situation-bound Utterances in L1 and L2*. Berlin: Mouton de Gruyter.
- Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. In J. Aarts & W. Meijs (Eds.), *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research* (pp. 163–171). Amsterdam: Rodopi.
- Kuiper, K. (1996). *Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters*. Mahwah NJ: Lawrence Erlbaum.
- Kuiper, K. (2004). Formulaic performance in conventionalised varieties of speech. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 37–54). Amsterdam: John Benjamins Publishing Company.
- Kuiper, K., & Haggio, D. (1984). Livestock auctions, oral poetry, and ordinary

- language. *Language in Society*, 13, 205-234.
- Lamb, S. M. (1999). *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: John Benjamins Publishing Company.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672. <http://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication*. London: Continuum International Publishing Group.
- Moon, R. (1992). Textual aspects of fixed expression in learners' dictionaries. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 12-27). Basingstoke: Macmillan.
- Nattinger, J. R., & Decarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nekrasova, T. M. (2009). English L1 and L2 speakers' knowledge of lexical bundles. *Language Learning*, 59(3), 647-686. <http://doi.org/10.1111/j.1467-9922.2009.00520.x>
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191-225). London: Longman.
- Perkins, M. R. (1999). Productivity and formulaicity in language development. In C. Schelleter, C. Letts, & M. Garman (Eds.), *Issues in Normal and Disordered Child Language: From Phonology to Narrative* (pp. 51-67). 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. <http://doi.org/10.2307/413177>
- Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL learners. *System*, 39(2), 164-174. <http://doi.org/10.1016/j.system.2011.02.003>
- Schmitt, N., & Carter, R. (2004). Formulaic sequence in action: An introduction. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 1-22). Amsterdam: John Benjamins Publishing Company.
- Sinclair, J. (1991). *Corpus Concordance and Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

- Siyanova, A., & Schmitt, N. (2007). Native and nonnative use of multi-word vs. one-word verbs. *IRAL*, 45(2), 119-139. <http://doi.org/10.1515/IRAL.2007.005>
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429-458. <http://doi.org/10.3138/cmlr.64.3.429>
- VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition*, 12(03), 287-301. <http://doi.org/10.1017/S0272263100009177>
- Verstraten, L. (1992). Fixed phrases in monolingual learners' dictionaries. In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and Applied Linguistics* (pp. 28-40). Basingstoke: Macmillan.
- Warren, B. C. (2005). A model of idiomaticity. *Nordic Journal of English Studies*, 4(1), 35-54.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2), 180-205. <http://doi.org/10.1093/applin/16.2.180>
- Wray, A. (1992). *The Focusing Hypothesis: The Theory of Left Hemisphere Lateralised Language Re-examined*. Amsterdam: John Benjamins Publishing Company.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21(4), 463-489. <http://doi.org/10.1093/applin/21.4.463>
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28. [http://doi.org/10.1016/S0271-5309\(99\)00015-4](http://doi.org/10.1016/S0271-5309(99)00015-4)
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44(4), 647-668. <http://doi.org/10.5054/tq.2010.235998>
- Yorio, C. A. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly*, 14(4), 433-442. <http://doi.org/10.2307/3586232>