L2 Learner-Made Formulaic Expressions and Constructions

Timothy Hall[1]

Teachers College, Columbia University

ABSTRACT

The emergence of lexico-grammatical productivity has been a central issue in the field of Second Language Acquisition (SLA). Ellis (2002) proposed that formulaic chunks of language are one resource for the learner to develop such productivity. This exploratory study sought to determine whether formulaic language chunks were observed in the oral production of three adult beginner learners of English as a Second Language (ESL) over a nine-week period in a community language program. It also attempted to determine whether there was a developmental relationship between formulae and productive forms called *constructions*. Tasks used to elicit the data included picture description tasks and semi-structured interviews. Results showed that formulae were minimally present in the learner output and that constructions and formulae of similar structure coexisted, but that a developmental relationship between formulae and constructions was not clearly evident. The discussion proposes the existence of a pre-formulaic stage account for the data, and submits that the relationship between formulae and productive constructions requires more intensive study.

INTRODUCTION

The role of formulaic expressions in second language (L2) development has attracted renewed interest in the field of SLA in measure with the increasing influence of Emergentist and associative theories of language development. In an Emergentist framework, *formulaic expressions*, also known as formulaic sequences, language chunks, or phraseological units, are deemed to be one facet of language that contributes to the emergence of lexico-grammatical productivity. Lexico-grammatical productivity describes a language structure's tendency to collocate or be combined with lexical and grammatical morphemes in a systematic way that expresses an intended range of meanings. For example, native speakers of English know to append the grammatical morphemes /s/ or /z/ to a noun to make it plural. The plural morpheme is highly productive because of the huge range of nouns that it can inflect. An example of a highly productive lexical collocation would be the modifier *very* because it collocates with many adjectives. The scope of productivity can be determined by the number of different linguistic forms or types that a feature can collocate with.

According to associative accounts of language use (Ellis, 2006; Ellis & Schmidt, 1998), rule-like productivity arises from the same cognitive processes that create form-meaning

---

[1] Timothy Hall is currently a doctoral student at Teachers College, Columbia University studying the relationship between formulaic expressions and second language acquisition. Correspondence should be sent to P.O. Box 497, Milford, NJ 08848. E-mail: tmh2111@columbia.edu

connections in lexical items. Speakers develop associative links between morphemes from statistical regularities in language input and possibly from output, and generalize an abstract "rule" of language use on that basis. Thus, in an associative view, the distinction between lexicon and grammar is minimized.

Formulaic expressions (hereafter "FEs" or "formulae") are argued to provide a resource for the learner to distinguish linguistic regularities from linguistic variation. FEs are multi-morphemic linguistic units, most often multi-word units, that show a high degree of formal invariance and are retrieved from memory as a whole. Examples of formulae include: (i) frequent discourse organizing forms such as *in other words*, and *first of all*, (ii) full idioms such as *hit the nail on the head*, and *it takes one to know one*, and (iii) a range of fixed and semi-fixed language features that serve a multitude of communicative functions. The learner is argued to acquire a store of multi-morphemic units and then use them for bases of comparison with other linguistic forms already stored in memory and in novel input. Subsequently, cognitive processes abstract rules from regularities in forms and meanings. The current understanding has acknowledged that FEs are implicated in second language learning (e.g., Hakuta, 1974; Myles, Hooper & Mitchell, 1998; Myles, Mitchell, & Hooper 1999; Wong-Fillmore, 1976), but there lacks specificity as to how this occurs. Collectively, the literature has advanced conflicting evidence about the precise interaction of FEs, language learning mechanisms and processes, age and learning context factors, and the ways these all conspire to yield lexico-grammatical productivity.

The remainder of this section will describe criteria for identifying formulae and contrasting them from productive structures called constructions. It will then survey the empirical and theoretical arguments that address the developmental question: Do formulae bootstrap adult second language learners into productive grammar? The section will show that there is a dearth of decisive evidence on this point, a fact that motivates the present study.

Identifying Formulae

The framework in which this study was conducted acknowledges a distinction between fixed languages called *formulae*, semi-fixed patterns of language called *low scope patterns* (LSP) and productive lexico-grammatical structures commonly referred to as *constructions*. However, constructionist literature makes no definitional distinction between these three terms; all are subsumed under the term "constructions". The definition of constructions, as defined in the constructionist literature (Ellis, 2006; Goldberg, 2003), is extremely broad and seems to pose formulae as a subset feature. Table 1 presents a simplified view of the contrast between constructions and formulae.

TABLE 1
Distinctive Features of Constructions, Low Scope Patterns, and Formulae

	Mono-morphemic	*Multi-morphemic*	*Fixed constituents*	*Variable constituents*
Structure				
Constructions	x	x	x	x
Low Scope Patterns		x	x	x
Formulae		x	x	

According to Goldberg (2003), the term "construction" covers a range of linguistic phenomena that entails both mono- and multi-morphemic structures that can be fully fixed in form, partially fixed in form such as low scope patterns (LSP), or fully productive and rule-like. Constructions are, "[...] stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns" (Goldberg, 2003, p. 219). In other words, a construction is a mapping of semantic and pragmatic meaning to a phonological form correctly marked for grammar, collocation, and situational appropriateness. A construction could be constituted by a single lexical morpheme like "/dɔg/ = dog" or it could be a single grammatical morpheme like the plural -*s*. It could also be a full idiom with specified constituents like "/kɪkðə'bʌkɪt/ = kick the bucket (die)". A construction could also be constituted by a low-scope pattern with certain lexico-grammatical restrictions and slots such as *"Good __ (morning / afternoon)."* Finally, a construction can be maximally variable, entailing several unspecified morphemes within abstract frames, such as passives or ditransitive structures.

Inferring from this definition, FEs and LSPs are a subset of constructions and require further differentiation. LSPs have been deemed to show limited variation of internal constituents whereas FEs show none. In addition to fixed formal properties, the SLA literature on formulae emphasizes a processing component: formulae are also characterized by holistic retrieval from memory. In early explorations of the formula phenomenon, Krashen and Scarcella (1978), and later Towell (1987), identify language sequences that are processed and produced in one of three ways by the learner: (i) Utterances or phrases used by the learner that are wholly memorized without knowledge of internal structure. These most often correspond to FEs in the literature. (ii) Utterances that are partly creative, partly memorized, for example, sentence frames with open slots such as *"Good _____ (morning / afternoon /evening).* These most often correspond to LSPs. Both phenomena were defined as distinct from (iii) fully productive, rule-based speech. An interesting observation is that these early attempts to categorize language structures into three types actually mirror later conceptualizations of a three-stage developmental path that shall be investigated in the present study.

More recent conceptualizations, specifically of FEs, have proposed combining a statement of cognitive processes, stipulations on invariance of formal and lexico-grammatical features, and observations of idiomatic semantico-pragmatic function (Van Lancker Sidtis, 2003; Weinert, 1995; Wray, 2000, 2008). Processed as a whole, formulaic strings perform numerous communicative functions. They can (a) convey denotative meaning either as idioms or highly frequent lexical collocations with fixed order (e.g., *kick the bucket = die; salt and pepper,* not

pepper and salt); (b) convey structure information (e.g., *first of all, and another thing*); (c) perform speech acts (e.g., *get outta here* or *get out* = "I don't believe you"); (d) provide more evidence (e.g., *break it to me gently* = "tell me the bad news"); (e) serve as topic opening or topic-closing moves (e.g., *so whatsup with you? well that's enough of that*); (f) be used in honorifics or social conventions of face (e.g., *Your Majesty, I am deeply honored*); or (g) convey affective or phatic content (e.g., *what's up with him* asks what is wrong with a person, but often expresses some derision towards that person). While it has been suggested that a broad array of identifying criteria be used (Wray, 2000), the most central features of FEs attested to in the SLA literature (e.g., Myles *et al.*, 1998; Myles *et al.*, 1999; Nattinger & DeCarrico, 1992; Weinert, 1995; Wray, 2000; Yorio, 1989) include the following:

(1) Multi-Morphemic: a form contains more than one morpheme, either free or bound.
(2) An identifiable invariance in aspects of form (morphological or lexical, for example).
(3) Retrieved from memory as a whole. This can be evidenced by latencies, fluency, repetition, intonational curves, or inappropriate use, for example.
(4) Conventionalized qualities and contribution to perceived idiomaticity. (e.g., *heyhowyadoin?* is an informal, idiomatic greeting compared to the greeting, *how goes it?* used in archaic speech).

Items 1 through 3 are perhaps the most central of the defining criteria of FEs for the purposes of this study. From a researcher's perspective, the first criterion, *multi-morphemic*, allows enough linguistic context to distinguish FEs in learner output (or in native-speaker output, for that matter) from other form-meaning pairs that are also invariable, such as single-word imperative verb forms. In brief, this criterion creates the context in which to exercise the second criterion, identifiable invariance of formal features. This second criterion permits the researcher to ascertain the speaker's knowledge of variation and fixedness in the constituents of a string of language and identify it either as a formula or as a low scope pattern.

The third criterion, relying on stipulations raised by the first and second criteria, speaks to both representation and processing. It looks for evidence that a string of morphemes is stored holistically and processed as a whole. This scenario is contrasted with morphemes being individually stored, retrieved, and assembled into sequences for language use. Evidence of holistic storage from a second language learner, particularly an early L2 learner might include three possibilities: (1) Production fluency that is greater than that of the surrounding linguistic context; (2) Lexico-grammatical error as determined by the linguistic or pragmatic context. For example, a learner might produce **I want to know <u>how are you</u>*. In this case, the learner may have learned *How are you?* as a formula yet in the current linguistic context, *I want to know* makes *how you <u>are</u>* obligatory. Furthermore, learners can create formulae that are not target-like either in form or in use yet the forms are fixed and fluently produced. For example, a learner might produce **howitgoin' ?* instead of *howsitgoin'?* but the form is still formulaic. (3) Production of multi-morphemic sequences that are not consistent with similar interlanguage structures. For example, consider a learner that consistently produces verb phrases with the verb *go* in the form GO + NP without the preposition *to (e.g., I go the market)*. Should the learner produce a singular *go* phrase such as *go to the movies* that includes the preposition *to*, we would suspect that the learner stores this phrase largely as a whole.

The fourth criterion is useful to the extent that formulae tend to lend idiomaticity to a discourse context. In the case of non-native speakers, the presence of formulae in output may

indicate a recognition of the idiomaticity of the form (i.e., appropriate use and pragmatic awareness) or attempted fluency to keep up with discourse demands.

Formulae and Acquisition

Holistic storage and retrieval of formulae have been argued from a language use perspective to afford non-native speakers idiomatic qualities to their speech (Wray, 2008) and processing advantages in response to communicative pressure (Wray & Perkins, 2000). However, from an acquisitional perspective, benefits of formulae are not so clear-cut. A developmental relationship between formulae and grammar seems largely to be pinned on the contrast between child versus adult language acquisition. While the contrast is not specifically studied in the current paper, it is hoped that the ensuing adult data will help shed light on the debate. In the child second language acquisition literature, Hakuta (1974) and Wong-Fillmore (1976) attest to a relationship between FEs and grammatical development. For example, Hakuta (1974) collected naturalistic data over thirteen months from his participant, Uguisu, an untutored 5 year-old L1 Japanese learner of English, five months after her initial exposure to English. Spontaneous speech was recorded and transcribed during the two hour sessions every two weeks. Hakuta's data showed ample evidence of "regular, patterned segments of speech [...] without knowledge of their underlying structure [...] but with the knowledge as to which particular situations call for what patterns" (Hakuta, 1974, p.288). Such patterns were observed to include configurations of the copula such as *this is*, the segment *do you* as a question marker, and *how to* as an embedded question. In accordance with the definitions of formulae and constructions provided above, it seems Uguisu had made a form-pragmatic/semantic connection with certain multi-morpheme linguistic features. She had not analyzed the constituents and had processed them as a whole.

Hakuta (1974) concluded his study with two questions of great relevance to the current study:

> "Does the rote memorization of a prefabricated pattern accelerate or decelerate incorporation into the structure? In other words, does the learning of a pattern signal or motivate search for its internal structure, or does it hinder the search because the prefabricated pattern is easy to use? And [...] do prefabricated patterns whose internal structure is finally perceived remain as convenient short-cut routes to production, or are they simply discarded, never to be employed again?" (Hakuta, 1974, p. 296)

In other words, Hakuta questioned, as we question here, whether there is a developmental relationship between holistically stored multi-morphemic strings of language and productive grammar, and how formulae and constructions relate in end-state grammars.

The literature addressing the relationship between formulae and lexico-grammatical development in L2 adult learners is conflicting and under-substantiated. Those arguing against the role of FEs in lexico-grammatical development have stated that formulae may exist in the interlanguage but not as a resource for grammatical development (Huebner, 1983; Rehbein, 1987; Schmidt, 1983; Yorio, 1989). Granger (1998) is unsupportive of a formula-productivity link given the dearth of evidence but highlights several factors to be more deeply considered before drawing overarching conclusions. These include comparing adult versus child learning, preferred learning styles (holistic and analytic), and foreign language versus second language (SL)

learning contexts. Granger recognizes that foreign-language-instructed learning contexts may need to rely more on the use of FEs than second language learning contexts because FL learners have little exposure to the target language. Therefore, there are fewer chances to develop idomaticity. Thus, she suggests, as researchers like Ellis (2002) have observed, that the quality of input may be a notable factor for consideration.

In light of Granger's argument, the condition of learning context may be unpacked even further: while foreign language learning is primarily instructed learning, second language learning contexts can be additionally contrasted by instructed versus non-instructed conditions. Each may have its own impact on the nature of input and output. Studies of FL instructed learners by Myles *et al.* (1998) and Myles *et al.* (1999) provide positive evidence of a developmental link between formulae and lexico-grammatical productivity in the FL setting. Their research participants, eleven and twelve-year-old learners of French, progressed from holistically stored interrogative forms to more analytically stored forms over the course of a year. Consistent with Granger (1998) and Ellis (2002), the two studies by Myles *et al.* suggest that the classroom context may have created input conditions that were conducive to formulaic intake, use, and subsequent analysis. It remains a question whether this pattern is evident in adult SL instructed contexts. It is noted that in such an L2 context, it is difficult to control for input and processing factors because learners are likely to receive language input outside the classroom. Therefore, SL instructed learner data may be "messier" than FL instructed learner data.

Taking a more ambivalent stance on the developmental question for adults, Krashen and Scarcella (1978) present an extensive and often cited argument *against* FEs' contribution to grammar. However, they concede in their conclusion that, while FEs are not the *sine qua non* of second language acquisition, they may serve as "intake for the creative construction process" (Krashen & Scarcella, 1978, p. 298). There are several aspects of their statement that are worthy of note here. First, it recognizes that there is a process through which creative constructions are developed. The nature of this process is a central question of the present study. Second, it raises the question of how intake for the process of developing constructions comes about. The Emergentist assumption is that preexisting static and productive language structures in interlanguage interact with each other and with novel input to result in intake. This too is a central question for the present study. Third, their observation suggests that there are multiple pathways to productive grammar; this leaves open the possibility that formulae can exist in parallel to, as well as intake for, the creative construction process.

In stronger support of a developmental link between formulae and grammar for adults, Ellis and Sinclair (1996), Bardovi-Harlig (2002), Ellis (2002), and Wray (2008) contend that formulae do indeed play a role in L2 grammar development. Ellis and Sinclair (1996) and Wray (2008) contend that memorized chunks of language serve as resources from which linguistic abstractions can be made. Ellis (2002) suggests that learning mechanisms that are active in first language (L1) acquisition (e.g., sensitivity to type and token frequency, comparison, and categorization) are also active in L2 acquisition. Therefore, he posits that, as in L1 acquisition, we might expect to see an L2 developmental path where the learner progresses from chunked multi-word units toward productive structures. This assumption, he says, is a "reasonable default" (p. 170) in investigating the development of productive structure.

Bardovi-Harlig (2002) provides partial confirmatory evidence for the formula-productivity link in adult L2 learners, but results are not clear cut. She reports on a longitudinal study of sixteen instructed adult L2 learners acquiring expressions of futurity, *will* and *going to*. She observes a somewhat complex scenario from her learners, namely that, "[...] patterns of

emergence suggest that formulaic use may occur prior to creative use as predicted. However, they also show that formulaic use may appear at the same time or even after wider use" (p. 194). Her study provides empirical evidence for the position held by Krashen and Scarcella (1978). It suggests that while a developmental relationship between formulae and productive structures may exist, the relationship is not clear cut. The remainder of this section will sketch Ellis's (2002) suggestion, the investigatory starting point, on the relationship between formulae and productive structures that forms the basis of the present study.

The Proposed Learning Path

Ellis (2002) has suggested that the learner's exploitation of formulae as an acquisitional resource follows a three-stage learning path that curiously resembles Krashen and Scarcella's (1978) and Towell's (1987) three distinct language structures described above. Ellis suggests that a formula is first taken in from input and stored holistically in long term memory as a form-meaning pair. At the stage of intake, the formula is unanalyzed by the learner. The formula is then subject to domain-general cognitive mechanisms and processes such as sensitivity to frequency and variation, concatenation, comparison, categorization and abstraction. Learners cautiously generalize common features over a few more types, as evidenced by low scope patterns whereby minimal variation of an original formula's constituents or contexts is evident.

Gradually, the variation of constituent forms increases as a result of cognitive processing and the same LSP develops into a fully productive construction (CONST): FE → LSP → CONST. Only some FE's can become productive: For example, *hey you!*, the call for attention, is not likely to vary even morphologically in use. However, the formula *What's up with ...?* meaning *What is wrong?* might evolve into a low-scope pattern in the interlanguage whereby a single variation into the past tense is permitted in the main verb *What <u>was</u> up with ...?* Constructions, as conceived for the present study, are fully productive in that they may reflect the fullest extent of variation permitted by the sequence of morphemes in idiomatic usage. In the case of the *What's up with* example, a learner would exhibit acquisition of the construction *what + BE + up with* by correctly inflecting the verb through permissible tense / aspect markings: *What's up with*; *What was up with*; *What's been up with,* but not **What will be up with.*

As seen from the synthesis above, it is unresolved whether adult learners use formulae as a resource for bootstrapping into grammatical productivity. In order for this to be established, it is necessary to take a longitudinal perspective, ideally from the early stages of language development. Furthermore, with beginner learners, it may be possible to document the entry of an FE into the interlanguage system and expose its relation, if any, to the development of interlanguage grammar and lexis. The current, exploratory study sought data from instructed ESL learners and makes tentative first steps in investigating the bootstrapping concerns with two questions:

1. Do L2 instructed adult beginner ESL learners use FEs, LSPs and constructions in the oral mode?
2. Is the learning path Formula → Low-Scope Pattern → Construction evidenced in participant data? In other words, is there evidence that FEs can serve as intake for the development of productive grammar?

METHOD

The Learning Context

The study was conducted in a beginner English class in the Community English Program (CEP) at Teachers College, Columbia University. The CEP classes met three days per week for two hours each day. Two instructors (L1 American English, L1 Korean) alternated teaching days, used a set curriculum and adopted a communicative language approach using the *Stand Out* textbook Level 2. Linguistic items targeted for teaching during the course included basic vocabulary for clothing, colors, family, money, and basic communicative functions such as asking for prices, stating basic comparatives, and describing daily habits and clothing choices using the simple present and present progressive. Placement into these classes was determined by a comprehensive placement exam.

Participants

Three students from the CEP class participated in the present study. They were part of a cohort of twelve beginner learners with diverse L1s (Spanish, Arabic, Japanese, Korean, and Ukrainian). Participant 1 was a 58 year-old woman whose L1 was Korean. Her length of residence in the United States was two months. Participant 2 was a 50 year-old woman from the Ukraine. Her length of residence at the time of the study was seven months. Participant 3 was a 19 year-old native of Yemen whose length of residence at the time of the study was three months. All participants reported having some very basic English language instruction at some point in their lives, though for the two older learners, it was decades ago. The participants were selected from the cohort on the basis of their relatively low-level of oral proficiency compared to other learners (utterances were generally limited to three or four words at a time with few grammatical inflections, and word retrieval was slow) and recency of arrival in the US.

Procedures

Data collection was longitudinal involving four data collecting sessions scheduled roughly one-and-a-half weeks apart over a nine-week period. On data collection days, the participants were pulled individually from the language class to perform tasks in the presence of the researcher. Data were elicited through several tasks, recorded, and transcribed per session.

Data Elicitation Tasks

Data elicitation tasks were designed to produce naturalistic oral data with a modicum of real-time communicative pressure. It has been argued by Myles, Hooper, and Mitchell (1999) and Wray (2000) that formulae afford a relief from linguistic processing burdens compared to rule-based assembly of language. Therefore, a degree of real-time communicative pressure would be expected to predispose learners towards formulaic language. Semi-structured interviews were used by Myles *et al.* (1999) to this end and became the basis for data elicitation in the present study.

For each data collection session, participants performed a single task without preparation that was intended to generate as much of their own oral output as possible given their level of

linguistic development. To this end, materials were selected with the criteria of having images that were describable with basic language, depicted scenes that were culturally accessible, but that called for some degree of interpretation. Participants 1 and 2 were present for all data collection sessions. Participant 3 was present for all sessions except Session 4. Session 1 was a semi-structured interview for which participants had no preparation. The participants were shown a picture of the researcher's family and were indicated the family relationships of each of the members. The participants were then asked to describe their own family, where family members lived and what kind of work they did with a prompt "And what about your family? or "Tell me about your family," or "Where does he live? What work does he do?" The researcher attempted to maximize participant output with minimal prompting (e.g., pausing, backchanneling cues, hand gestures) in order to reduce the possibility that participants were merely copying the researcher's speech.

For Session 2, participants were presented with two pictures and asked to speak about each picture in succession. The first picture was a lemonade stand scenario with children, and the second was of humanitarian aid workers handing out books to children. In each case, participants were asked "What do you see?" to initiate participant talk, "What about this?" to guide participant attention when description lapsed and "Why are they doing this?" to motivate more talk.

Session 3 involved a movie description task and interview. The researcher showed a two-minute Quicktime movie of the researcher attempting but failing to feed his baby nephew. The movie was played only once. Participants were instructed at the beginning of the movie to describe what they saw. When the movie ended, the participants were prompted for more information about what they saw with questions such as "Why do you think ...?" and "What is the problem?" Again, efforts were made to maximize participant output with minimal input from the researcher.

Session 4 comprised two parts. Participants completed a picture description task using either the book donation picture from Session 1 or, for variation, a picture of a woman feeding ducks. Afterwards, the researcher conducted a semi-structured interview, enquiring into the participants' plans for summer.

Coding

The coding criteria attempted to capture as many instances of formulae as possible and distinguish them from productive language features. First, from a data-driven perspective, it was necessary to identify which structures or consistencies existed in the participant output, based on the assumption that participants were likely to use linguistic forms in some systematic way, and that formulaic expressions would likely be a subset of these linguistic forms. Some forms might exhibit productivity (maintaining grammatical structure across instances but showing paradigmatic variation of constituents, that is type variation, across instances), whereas some forms may be fixed across instances of use.

Second, from a theory-driven perspective, evidence was sought of linguistic forms being formulaic according to the psycholinguistic criteria provided in the previous section. Specific qualities that substantiated coding as a formula (FE) for the current study, as described fully in the previous section, "Identifying Formulae", included (1) multi-morphemic form, (2) evidence of processing as a whole (i.e., production fluency or complexity greater than that of the surrounding linguistic context; the existence of morphological errors as determined by the

surrounding linguistic context), and (3) invariance of form (i.e., two or more tokens exhibiting the exact same form regardless of context).

Productive forms, that is, those multi-morphemic forms found over several instances with some componential change across instances, were coded either as low scope patterns (LSP) or constructions (CONST). An arbitrary cutoff point was set given the need to distinguish the two constructs from each other because both entail some variation in linguistic form compared to FEs. A multi-morphemic linguistic feature was coded as a low scope pattern if, within the multi-morphemic unit, one morpheme was consistent across instances but the other morpheme varied through a maximum of two other types. For example, if the plural *-s* morpheme was found to mark the two nouns *dog* and *cat* to the exclusion of all other count nouns, then the plural morpheme would be coded as a low scope pattern. A multi-morphemic linguistic feature was coded as a construction if the stable morpheme was found with more than two types. To continue the example, if the data showed that plural *-s* marked three, four, or five count nouns, it was marked as a construction. Table 2 presents the coding scheme.

TABLE 2
Coding Scheme

Characteristics	Formulaic Expression (FE)	Low Scope Pattern (LSP)	Construction (CONST)
Multi-Morphemic	x	x	x
Processed as a whole	x	x	
Variance of form	NONE	≤ 2 types	> 2 types

It was expected through these theory-driven and data-driven approaches that the cline of fixed, semi-productive, and fully productive forms would be made evident.

Analysis

Analysis of the data was largely qualitative, entailing for the first research question the identification and enumeration of formulae, low scope patterns, and constructions from transcribed participant data. Per the second research question regarding a developmental relationship between productive and non-productive structures, comparisons were made between FEs and constructions for structural (syntactic) similarity. If the same linguistic types, such as interrogatives, negatives, or prepositional phrases, were found in both FEs and CONSTs, then argumentation for a developmental relationship was pursued. Subsequently, the concordancing software Antconc 3.20 was used to generate word and collocation lists for each participant. This would facilitate comparing the contexts of individual constituents of multi-morphemic phrases across FEs, LSPs, and CONSTs. Furthermore, based on the order of output, the software allows the researcher to see the order in which utterances were produced across elicitation sessions, thereby determining temporal relationships and permitting inference about development.

RESULTS

In response to the first research question, formulae were evident in the oral output of the participants, as well as low scope patterns, and constructions. Results are shown in Table 3. Formulae were present for all participants to varying degrees. Participant 1 exhibited ten formulae, of which four were found in noun phrases, one in a determiner position, one in a locative prepositional phrase, and four entailing subject/verb combinations. Participant 2 exhibited eight formulae of which two entailed noun phrases, one was a locative prepositional phrase, three were interrogative structures, and two were subject/verb combinations. Participant 3 had two formulae in his data, one in a verb phrase and one in a noun phrase. There were relatively few items coded as low scope patterns, which basically entailed a subject and its copula for Participant 1, and negation and S+V for Participant 3. Forms coded as constructions were found for all participants. These included S + V constructions, noun phrase constructions (*very many* + N), adjective phrase constructions (*very* + ADJ), recursive constructions (*x and ... y and ....*; *maybe x ... maybe y),* and negation constructions (*not x, not y*). In response to the second question, the data did not support the developmental path of FE → LSP → CONST. First, for all participants, there was little overlap of comparable forms from formulae to constructions, with the exception of locative prepositional phrases found in the data of Participants 1 and 2. Second, there were relatively few low scope patterns, and certainly none that bore sufficient structural or lexical similarities to other linguistic forms.

Following a temporal analysis of the locative phrases[2] for Participant 1 (see Figure 1), it was observed that both a construction and a formula using the locative *in* were produced in Session One: *in United State* (CONST) and *in class* (FE). It is not possible, due to the lack of relevant data, to determine whether there was a developmental relationship between the FE and the construction. At the very least, this co-occurrence suggests that formulaic and productive forms exist simultaneously.

Subsequent locative forms by Participant 1 seemed to generally adhere to the same form, *in* + NP. There were no articles and no determiners (though there was a possessive pronoun) suggesting that the learner was productively using the locative preposition *in*, especially with geographic nouns. Because this construction was present in the earliest session, it was not possible to determine whether it had formulaic origins. On a final note, Participant 1 also produced a locative phrase without the preposition *in* where it was obligatory: *she's play garden.* Though the data fails to be confirmatory, it is conceivable that the FE, *in class,* could be a developmental precursor for *in* being productive with non-geographic nouns such as *garden*.

In a similar analysis, Participant 2 produced twenty-five instances of locative *in* + NP. Figure 5 illustrates her data. Almost all instances lacked the definite article in a manner that was consistent with obligatory contexts in the L2 and/or the fact that articles are absent from her L1, Ukrainian. Thus, like Participant 1, *in* + NP seems to be a productive structure in her

---

[2] As shown in Figure 1, Participant 1 generated the temporal preposition phrases *in the spring, in the summer,* and *in the fall* during a class in the middle of the term. For this class, participants were able to prepare written output that was read to the class. Since this output was not spontaneously generated language, it entailed a shift to written mode, and did not permit control of the conditions under which the language was generated, this data were not used for the current study.

interlanguage. Because the participant was productive with *in* + NP from the very first session, it was not possible to show whether that construction originated through the developmental path.

Table 3
Linguistic Structures by All Participants

Participant	Formulae, LSPs & Constructions	Classification
#1	CONST (or LSP) S + V	SUBJ + V
	CONST [in + NP]	Locative prepositional phrase
	LSP [she's; he's]	SUBJ + copula
	LSP [I'm; I am]	SUBJ + copula
	FE [shortpant]	noun phrase
	FE [shortsleeve]	noun phrase
	FE [jeanskirt]	noun phrase
	FE [Red Cross]	noun phrase
	FE [in class];	locative prepositional phrase
	FE [excuse me]	SUBJ + V
	FE [alot of]	determiner
	FE [I hope so]	SUBJ + V
	FE [I think]	SUBJ + V
	FE [it's fine] (meaning "I'm fine)	SUBJ + V
#2	CONST S + V	
	CONST [in + NP]	Locative
	CONST [very many + noun]	noun phrase
	CONST [very + noun];	noun phrase
	CONST [very + adj];	adjective phrase
	CONST [SUBJ + have] ;	
	CONST [maybe + X]	verb phrase;
	CONST [not + X]	noun phrase (negation)
	FE [whoisit]	interrogative
	FE [two weeks]	noun phrase
	FE [economic crisis]	noun phrase
	FE [in the kitchen]	locative prepositional phrase
	FE [what is this]	interrogative
	FE [what is your name]	interrogative
	FE [how old are you]	interrogative
	FE [I would like] (meaning "I like")	SUBJ + V
	FE [I don't know]	SUBJ + V
#3	CONST [number + {one, two}] (meaning "first, second")	noun phrase;
	LSP [they want; I want]	SUBJ + V
	LSP [NO+ X] [no don't + X]	negation
	FE [give you]	verb phrase
	FE [another one]	noun phrase

FIGURE 1
Usage of *in* + NP by Participant 1

```
 yes uh my two brother livis in united state ah...and two children  Kor
y garden teacher student     [in class]          Teacher and student [
rom Korea  [??] [the weather] in the spring is warm and sunny the weath
is warm and sunny the weather in the summer is uhm hot and rainy the w
uhm hot and rainy the weather in the fall is cool and windy and cloudy
 windy and cloudy the weather in is cold and snowy [I like] winter bec
 born and grew up and married in Korea I have three daughters and one
nter name is yong ? she lives in New Jersey with her husband and daugh
 oh no no no thirty She lives in Los Angeles with her husband Her husb
ears old and very cute. Other in my... family are in Korea. Sometimes
te. Other in my... family are in Korea. Sometimes I sad because they a
five ??? who are still living in my country. I am fifty eight years ol
. I was teacher and counselor in Korea. I can do yoga and play piano a
```

FIGURE 2
Usage of IN + NP by Participant 2

```
Ukraine Oksana live in village

an um  financist yes in Ukrain not job very [econ
now stud stud study in professional childrens sch
mmm                I see in picture four child maybe
hey need to sell   In America             Yes I
I see Not so much in New York            No Asp
e I was born in uhm in [?] in  West Ukraine. It

uh fast river and uh in mountains  lakes. The wed
divided.  Uhm I live in West Ukraine seven months
months.  Uh I live in Odessa. Sea black sea. Ho
e is .... chemistry in fabric in alcoholic [?]
chemistry in fabric in alcoholic [?] fabric...wo
She live she lives in in  village. She is in Od
e live she lives in in  village. She is in Odess
in  village. She is in Odessa. They are they li
They are they live in Odessa  I work for ...yes
I work for ...yes? in Odessa I work in academy
es? in Odessa I work in academy food technology
ating finish ate...[in the kitchen]... [whoisit
ur and half year     in home I kilbase make.
y white t shirt have in her in his hand in his r
t shirt have in her in his hand in his right har
e in her in his hand in his right hand some books
```

However, Participant 2 did produce one FE, *in the kitchen*, in which we see the preposition with an obligatory definite article and a noun. This was coded as a formula because of its fluency of production as well as its divergence in form with other locative phrases which

did not have articles. The current study was not able to show a breakdown of this formula over time.

DISCUSSION

The purpose of the current study was to explore whether formulaic expressions were present in instructed ESL adult beginner language and whether a developmental relationship between formulae and constructions was evident. The data collected over nine weeks showed that FEs were present to a limited degree in beginner learner interlanguage but there was no evident relationship between FEs and lexico-grammatical productivity as framed by Ellis's developmental path. The remainder of this section will attempt to address why formulae were not more widespread in the learner data, arguing for a pre-formulaic phase in learning. The section will then address the lack of evidence for a developmental relationship between formulae and constructions, provide a rationale for the co-occurrence of formulae and constructions of similar forms, and speak to the limitations of the present study.

A Pre-Formulaic Phase

Given the low output of formulae by the participants, there is the possibility that the learners were in a pre-formulaic stage. Participant 3 produced the least formulae and least amount of output overall. Participants 1 and 2, although more verbally communicative, produced only a few more formulae each. Overall, the participants' language was halting and labored, with a narrow range of lexical and morphosyntactic structures produced in their data. It leads one to speculate that there may be requisite situational and cognitive preconditions that enable a learner to store and subsequently use formulae. This supposes the low number of formulae in the participant data as a function of the relatively low level of language ability. Therefore, it is suggested that unless preconditions for acquiring formulae are met, bootstrapping into grammar will not follow the developmental path FE → LSP → CONST.

The following preconditions are posited for learners to acquire formula that may serve as a bootstrapping resource for productive language structures: (1) sufficient language experience, (2) developmental readiness, and (3) opportunities for form-meaning mapping. These need to be empirically validated in future studies. Per the first condition, learners need sufficient experience with language input to build up a basic store of language structures. The beginner learners in this study were relatively recent arrivals to the United States with very little or no instruction in English prior to arrival and few interactions in English outside the class. Consequently, they did not have extended or meaningful exposure to the language input necessary for development prior to the study.

Per the second condition, learners must be developmentally ready to intake formulae that may serve for future language development. They need to have automatized certain lower-level language processes such as phonological processing and lexical access so that memory resources are freed up to process and take in multi-morphemic units holistically (Ellis, 2001; McLaughlin, 1987; Miller, 1954). For instance, given the halting nature of the participants' output (e.g., labored pronunciation of forms, searching for words, non-target-like morphosyntax, utterances sequenced in two or three word groups) it is likely that much of their working memory resources

were occupied with unautomatized phonological, morphological, and syntactic processing tasks (Hakuta, 1974). Consequently, few formulae may have been taken in up to this point.

Finally, per the third condition, learners need opportunities for mapping forms to meanings through interactive contexts that lend perceptual salience to FEs and their form-function connections. With such opportunities, FEs can be committed to rehearsal and memory, that is to say, to be processed for intake. The third condition is moderated by the current state of the learner, described by the first and second conditions, as well as the nature of the learning environment. Early stages of language learning, as exemplified by the participants in the current study, are observed to have a limited processing span and little previous L2 knowledge; form-meaning mapping is an arduous task in even the most hospitable of language learning environments. Consequently, early stages of learning might only exhibit formulae whose forms are salient, either through frequency of input or by instruction, or whose forms fall within the perceptual span of the learner, and whose meanings are readily retrievable from the context.

Despite these conditions, certain types of instruction methods might differentially influence the manner, span, and type of formula to be taken in by the learner due to unnatural frequencies of use. For example, in the present study, there was very little rote method in the classroom. Any instructional repetition of forms was strictly at a word level and occurred three or four times during the entire observation period. Concurrently, though as noted not necessarily consequentially, there were relatively few formulae. In contrast, the Myles, Mitchell, and Hooper (1999) study on interrogatives shows learner chunks to be more pervasive. As acknowledged in that study, repetition and drilling were a notable part of the instructional method and attributed to the build up of FEs.

The Developmental Path: Formulae, Low Scope Patterns, and Constructions

Overall, the nine-week observation period did not support for the learning path where a single productive form emerges in the following stages: Stage 1: Formulaic Expression → Stage 2: Low Scope Pattern → Stage 3: Construction. For example, Participant 2 produced what was identified as a formula, *whatisyourname,* when referring to the third person, but data did not show any evolution of the form. Instead the present study found a scenario that is less organized. Learners exhibited formulae of one type, constructions of other types, and in the case of locative expressions, formulae and constructions of similar types (Bardovi-Harlig, 2002).

This last phenomenon was observed from Participant 1 and Participant 2. Participant 1 exhibited the formula *in class* as well as productive locative prepositional phrases such as *in Korea* and *in my country*. Participant 2 exhibited the formula *in the kitchen* as well as productive forms exemplified through *in Odessa*, *in village*, and *in fabric* (factory). In the literature, there is support for this from an empirical perspective: Bardovi-Harlig (2002) observed that productive and formulaic forms of the same type co-existed in her learner's data. From a theoretical perspective, Wray (2008) has asserted that the mental lexicon is heteromorphic; this condition allows chunked forms and analyzed forms co-exist.

However, it must be argued that such observations do not, as yet, preclude the possibility of the developmental path. Emergentists (Ellis, 2006; Ellis & Larsen-Freeman, 2006) submit that learners store a variety of productive and static language forms (L1 and L2) that interact with each other and with novel input via cognitive processes; all are resources for language development that might feed into and result from the developmental path. The present study observes that learners are in the first stages of developing stores of productive and formulaic L2

language structures. Consequently, the interactions between such structures are not manifest yet and we see no movement from FEs to LSPs and then to constructions. Furthermore, the developmental path makes no claims about what happens to formulae after the learner has analyzed them. Second language learners (Bardovi-Harlig, 2002), and even native speakers (Sinclair, 1991), do not necessarily discard formulas once analyzed forms are derived.

To illustrate the overall dynamic, a longer study might capture formulae giving rise to productive article systems within the context of a prepositional phrase. A formulaic prepositional phrase like *in the kitchen* might provide a contrast and therefore interact with the learner's model of other noun phrases, giving rise to a more productive, more variable, though not necessarily a more target-like article system. The abundant construction IN + NOUN would hypothetically be compared to IN + PROPER NOUN and the formula IN + DEFINITE ARTICLE+ NOUN in cognitive processes. A learner's L1 which lacks articles might skew the adoption of formulae and the development of productivity towards a no-article solution. Evidence in the L2 such formulae like *in class* and *in America* might reinforce this. Yet these need to be reconciled with formulae such as *in the kitchen* and others like it, perhaps internalized through the instructional context. Variability in learner production of articles ensues. This has been described as the learner "shifting in a slow and gradual manner [...] toward the maintenance of an internal consistency within the structures which the learner possesses" (Hakuta, 1976, p. 331). Over time, we might expect to see increased, though not necessarily accurate, use of the definite article in the learner's output; the interlanguage becomes composed of non-target-like formulae, target-like formulae, and analogous analyzed structures.

Limitations

It is duly noted that the amount of elicited data in the current study were too limited to substantiate the learning path under investigation. This seems to be the culmination of the participants' low language proficiency and relatively unsystematic data elicitation techniques. First, a future study might establish selection criteria for participants within the cohort of beginners, taking care that language proficiency is minimally adequate to yield interpretable data. Second, a future study might systematically solicit more data through more controlled task dimensions across written and oral mode, and through intensive spacing of data collection. An intensive spacing of data collection would be appropriate to document the ephemeral nature of L2 formulae and constructions as they are situated in early language acquisition. Third, a greater control of the input factor would be desirable. No data was collected regarding the participants' L2 input outside the classroom in the present study which severely curtails any speculation of its influence.

Last, it is worthwhile to speculate on the type of data that would strongly confirm the processes and products that are suspected to be at work here. One might hope to see, for example, a learner intake no less than two multi-morphemic strings that are structurally similar, readily traceable to their input, and which first emerge in their output in within some proximal amount of time, perhaps even within the same data collection session. These multi-morphemic strings would have the trappings of holistic processing as prescribed in this paper and repeatedly manifest themselves in such a manner over another data session to show stability as holistic forms. Subsequently, one would hope to see increasing amounts of variation of their constituents in a way that reflects the similarities and differences of their structures until finally, in the course of time, each string exhibits the full variation that their lexico-grammatical constituents permit.

CONCLUSION

This exploratory study investigated the following questions: (1) whether formulaic expressions were evidenced in beginner adult learners' speech and (2) whether the putative developmental path of Formula → Low-Scope Pattern → Construction was in evidence. The study found that formulae were used but, due to a lack of evidence, it was unable to resolve whether formulae contribute to productivity. Instead, it found a scenario that is less neat than the proposed learning path: learners exhibited formulae of one type, constructions of other types, and in the case of locative expressions, formulae and constructions of similar types. However, these conditions did not preclude the possibility of bootstrapping. In light of these considerations, the scope of applicability of the developmental path needs also to be better defined in addition to the exact mechanisms that are expected to give rise to it. With an improved study design in future, it may be possible to capture such a phenomenon.

ACKNOWLEDGMENTS

REFERENCES

Bardovi-Harlig, K. (2002). A new starting point? *Studies in Second Language Acquisition*, *24*(2), 189-198.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*(2), 215-221.

Ellis, N. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33-67). New York: Cambridge University Press.

Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143.

Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive creed. *AILA Review, 19*, 100-121.

Ellis, N. C., & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes*, *13*(2), 307.

Ellis, N., & Sinclair, S. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology, 49A*(1), 245-250.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics- introduction to the special issue. *Applied Linguistics, 27*(4), 558-589.

Goldberg, A. (2003). Constructions: a new theoretical approach to language. *TRENDS in Cognitive Sciences, 7*(5), 219-223.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. *In Cowie, A. P. (ed.) Phraseology: Theory, analysis and applications, 145-160. Oxford:* Clarendon Press.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, *24*(2), 287-297.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language Learning, 26*, 321-351.

Krashen, S., & Scarcella, R. (1978). On routines and patterns in language acquisition and performance. *Language Learning*, *28*(2), 283-300.

McLaughlin, B. (1987). *Theories of second language learning*. London: Arnold.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning, 48*(3), 323-364.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French l2: A basis for creative construction? *Studies in Second Language Acquisition, 21*(1), 49-80.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Rehbein, J. (1987). Multiple formulae: aspects of Turkish migrant workers' German in intercultural communication. In K. Knapp, W. Enninger & A. Knapp-Potthoff (Eds.), *Analysing intercultural communication* (pp. 215-248). Berlin: Mouton.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Towell, R. (1987). A discussion of the psycholinguistic bases for communicative language teaching in a foreign language situation. *British Journal of Language Teaching, 25*(2), 91-99.

Van Lancker Sidtis, D. (2003). When novel sentences spoken or heard for the first time in the history of the universe are not enough: toward a dual-process model of language. *International Journal of Language & Communication Disorders, 38*, 1-44.

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

Wong-Fillmore, L. (1976). The second time around: Cognitive and social strategies in second language acquisition. Stanford University, Stanford, CA.

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

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

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

Yorio, C. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstam & L. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp. 55-72). Cambridge: Cambridge University Press.