

A NEW STARTING POINT?

Investigating Formulaic Use and Input in Future Expression

Kathleen Bardovi-Harlig

Indiana University

The use of formulaic expressions by second language learners has received little attention from second language acquisition research investigating interlanguage temporal systems. Instead, this field of inquiry has emphasized the productive use of verbal morphology by employing type-token analyses. This paper considers the proposed developmental sequence of *formula* > *low-scope pattern* > *construction* in the emergence of future expression in a longitudinal study of 16 adult learners of English as a second language. The findings suggest that the use of formulaic expressions may be subject to individual variation and that learners may use formulaic expressions to different degrees when developing form-meaning associations even in the same grammatical subsystem, such as the tense-aspect system. The findings also suggest, however, that the practice of favoring type over token analysis as a matter of course may eliminate valuable information about the emergence and development of temporal expression.

Ellis proposes a developmental sequence—from formula, through low-scope pattern, to construction—as “a useful starting point to investigate the emergence of constructions and the ways in which type and token frequency affect the productivity of patterns” in second language acquisition research (p. 145). The present response takes the hypothesized developmental sequence as a starting point to investigate a little-researched area of temporal expression in second language acquisition—the expression of futurity—and the ways in which

Address correspondence to: Kathleen Bardovi-Harlig, Program in TESOL & Applied Linguistics, Memorial Hall 313, Indiana University, Bloomington, IN 47405; e-mail: bardovi@indiana.edu.

instructional input may affect the acquisitional sequence. Beginning an investigation with formulaic language as Ellis advocates represents a departure from the standard analysis of freely produced language samples in L2 tense-aspect research, where type and token counts are used to exclude frequently occurring strings and to prevent them from inflating counts of appropriate use of emergent morphology. I suggest that the proposed developmental sequence is an interesting and informative starting point but not fully descriptive of L2 development, at least for the expression of futurity, because formulaic expressions do not dominate initial stages, nor are they excluded from subsequent stages.

The expression of the future is grammaticalized in English as *will* + verb (e.g., *I will study public relations*) and *be going to* + verb (henceforth referred to as *going to*; e.g., *I am going to work and study*). (Other expressions of futurity include present + adverb, as in *I leave tomorrow*, and present progressive, as in *I am leaving tomorrow*. These are not discussed further here.) The study of grammaticalized expressions of futurity offers the opportunity to examine the emergence of highly regular expressions, one of which, *will*, is morphosyntactically simple, and the other, *going to*, (potentially) complex. The future emerges relatively early, is immediately well formed with very few exceptions, and exhibits significantly more tokens than other multimorphemic tense-aspect forms, such as present perfect and pluperfect (cf. Bardovi-Harlig, 2000). Because of the relatively high number of tokens produced by learners, and the fact that all learners in the longitudinal study produced future forms, this study follows Schmidt's (1992) recommended method for studies of fluency, the multiple-case study, because it yields the extensive data that is required in order to be able to identify idiosyncratic formulaic utterances and their evolution over time.

In the future, event time (E) follows the time of speaking (S), or $S \rightarrow E$ (Comrie, 1985; Reichenbach, 1947). The differences between *will* and *going to*, which may include pragmatic as well as semantic differences, have been widely discussed (Haegeman, 1989; Leech, 1971), but it is not clear that the nuanced meanings are acquired during the first 16 months of instructed acquisition. The acquisitional data show that *will* is the dominant interlanguage marker of futurity.

In using formulaic use and input as the starting point for the investigation of the acquisition of future, this paper addresses two questions. First, is the sequence *formula* > *low-scope pattern* > *construction* in evidence in the expression of futurity? If so, do learners “unpack” unanalyzed utterances and begin to use parts of them productively, or do they drop them from their repertoire as rule-governed use takes over (Hakuta, 1974; Myles, Mitchell, & Hooper, 1999)? Second, what is the relation of formulaic language use to instructional input? As Ellis observes, “classroom environments . . . can distort the patterns of exposure, function, medium, and social interaction” (p. 170). On the other hand, classroom environments may also make it possible to track at least the broad outlines of pedagogical input to learners while tracking their production.

LONGITUDINAL STUDY

Sixteen learners representing four language backgrounds (five Arabic, six Japanese, two Korean, and three Spanish) participated in a longitudinal study; the mean length of observation was 11.5 months, ranging from 7 to 17.5 months. The learners were enrolled in the Intensive English Program, Center for English Language Training at Indiana University. Learners were placed in the first level of instruction and attended classes 23 hours a week, receiving instruction in listening and speaking, reading, writing, and grammar. They were in a mixed-language environment, receiving instruction in the host environment, with access to the ambient language, although they differed individually in their patterns of contact with native speakers and with other nonnative speakers of different L1 backgrounds (R. Ellis, 1990). The learners also showed individual differences in rates of development and eventual proficiency (Bardovi-Harlig, 2000).

The progress of the learners was monitored for the purposes of data collection during each 7-week session that they attended classes. During the observation period, 1,576 written texts and 175 oral texts were collected. The majority of the written texts consisted of journal entries (1,101); an additional 370 texts were compositions, 73 were essay exams, and 32 were elicited narratives based on silent films. The oral texts consisted of 102 guided conversational interviews supplemented by 73 elicited narratives based on silent films and the ensuing conversations.

These language samples comprise authentic language use, produced by the learners in the course of their ESL studies. The journals, which represent the majority of the written samples, were completely up to the learners; the topics of the essays and film retells were determined by the teachers and the researcher, respectively, but what learners wrote and how they wrote it was up to them. The same holds true for the oral samples. There were no tasks that controlled the production of any particular form (as opposed to topics). As a result of the topics that the learners wrote or talked about, as well as the number and length of journal entries and compositions completed, the number of tokens of tense-aspect forms varies across learners.

A second source of data was the teaching logs completed by participating grammar and writing instructors. The teaching logs recorded the topic(s) of instruction, classroom activities, type of feedback, homework, and page numbers of lessons in the textbooks, as well as copies of original materials created by the instructors (Bardovi-Harlig, 2000.)

Emergence of *Will* and *Going to*

In the acquisition of the grammaticalized expressions of the future *will* and *going to*, *will* emerges first and greatly outnumbers the use of tokens of *going to*. The early emergence of *will* is not the only cause of the higher number of uses: *Will* spreads rapidly to a variety of verbs, quickly yielding a large num-

ber of *will* + verb tokens and types. The rapid spread of *will* throughout the verbal lexicon suggests that, for most learners, there is either little initial formulaic use of *will* or that it is so brief that it cannot be detected in this corpus. All the learners use *will*. In 1,576 written texts, learners produced just over 1,400 tokens of *will*; in 175 oral texts, they used more than 700 tokens.

The emergence of *going to* presents another face. Relatively small numbers of tokens are distributed throughout the language samples. Learners produced 249 tokens of *going to* in the written sample and 52 tokens in the oral sample, roughly one-seventh the number of tokens of *will* in the same texts. Half of the learners (8 of 16) used *going to* 10 or fewer times. The other half showed more than 20 uses. No learner's use of *going to* reached his or her frequency of use of *will*. In the course of the emergence of *going to*, the string *I am going to write (about) + NP* appears in the written texts of about half of the higher frequency users (5 of 8) and meets the following recognized criteria for formulaic use (Myles et al., 1999; Weinert, 1995).

1. Frequency and invariance in form: The fixed string *I am going to write (about) + NP* has a slot for an NP and optional use of *about* and shows variation of neither subject nor verb. It is realized as *I am going to write two paragraphs* or *I am going to write about my father*.
2. Situational dependence: *I am going to write (about)* functions to announce a topic and typically appears in the opening paragraph of a composition but may be generalized by learners to their journal entries as well. Because of its function, there tends to be only one use per text.
3. Well-formedness that may be grammatically advanced compared to the rest of the learner's language: *Going to* constructions are uniformly well formed, and the earliest uses seem to precede productive use of the *to*-complement.

Five learners exhibited five or more uses of *going to write about* in the introductory function. This group includes L1 speakers of Spanish, Japanese, and Arabic. Examination of the role of L1 is beyond the scope of this commentary, but it should be pointed out that learners from all the L1s in the study (Arabic, Korean, Japanese, and Spanish) were represented in the group that showed high use of *going to*. The learners' use of *going to* is sketched briefly below and in Table 1. Production is reported in half-month intervals; T2.0 indicates the first half of the second month, T2.5, the latter half.

Carlos begins to use *going to* in his second month of instruction. *I am going to write* occurs in four of the first five tokens, although it has the introductory function only once. Of the five written tokens in T2.5, three represent formulaic use, followed by a single formula used at T3.0, when it dies out. In the oral sample at T3.0, four tokens and types of *going to* appear, all in the first person. No further uses of *going to* were produced in the corpus until T6.0, when Carlos uses four additional tokens and types (three oral, one written), varying person and number as well as verbs. He goes on to produce six additional types and tokens between T7.0 and T8.5, two in writing and four additional types and tokens orally.

Table 1. Emergence of *going to* and *I am going to write* in the interlanguage of five learners

	Carlos	Guillermo	Satoru	Eduardo	Khaled
L1	Spanish	Spanish	Japanese	Spanish	Arabic
Total <i>going to</i>	14 (W), 13(O)	25 (W), 0 (O)	31 (W), 7 (O)	30 (W), 7 (O)	41 (W), 16 (O)
Examples	5 T2.0 W: write (5 = 4g + 1f), lose; O: make, die T2.5 W: write (3), read, withdrawn T3.0 W: write; O: die, do, study, know T6.0 W: be; O: perfect, improve, do T7.0 O: play, destroy T8.0 W: break; O: thief [steal], appear T8.5 W: study T10.0 Last sample	6 T.5 W: do T5.5 W: play (2), start, share, write (2) T6.0 W: write (2) T7.0 W: have, do, give, show T7.5 W: write (2), play (2), get married, call, share T8.0 W: call, have, give, be T8.5 W: write T8.5 Last sample	28 T3.0 W: going to (about, move T3.5 W: write T5.0 W: write, stay T6.0 O: go T7.0 W: write (2); O: study, go T7.5 W: write (6); O: study, apply, begin, take T8.0 W: write (3) T8.5 W: write (5) T9.0 W: write (4) T9.5 W: write (3) T10.0 W: write (2) T11.5 W: write T11.5 Last sample	6 T1.5 W: take (2), see (2), spend, go, study T2.0 W: write (4) T2.5 W: write (3), tell (2), die, eat, see, continue T3.0 W: visit, spend, give, applicate [apply] I will write (2); O: write (1g), go T5.0 W: be, say, tell, cost T6.0 W: spend, make; O: have, happen, talk T7.5 O: study, be T8.5 Last sample	18 T.5 W: visit T2.5 W: make T5.0 W: write (1g); O: gonna read, try T5.5 O: try, speak T6.0 W: write (2g), study, be; O: do, try T7.0 W: change, try, be T7.5 W: write (2 = 1g + 1f), brush T8.0 W: write, use; O: study (2), take T9.0 W: write (10), be T9.5 W: write (2); O: transfer, send T10.0 W: write (2), be (4) T10.5 W: write T11.0 W: write, leave (2), give, work; O: leave (2), eat, study, work T11.5 W: try T11.5 Last sample

Note: W = written, O = oral, (xg + yf) = x general uses + y formulaic uses

Guillermo shows his first use of *going to* at T.5 in Level 1 following a class-assigned model letter (see also Khaled at T.5). His first productive uses appear at T5.5, the five tokens showing four verb types, one of which is *going to write*. After the second *going to write*, two more follow immediately (T6.0), resulting in three uses of *going to write* in three consecutive compositions in six days. The next five uses show five verb types, and *going to write* surfaces again twice in nonexclusive use in T7.5 and once in T8.5, along with nine additional tokens (seven types). No oral uses were recorded.

Satoru uses *going to* 31 times in his written texts, 28 of which are *I am going to write about* + NP. The five earliest uses of *going to* (T3.0–T5.0) include two uses of *going to write* and three other verbs. The next 26 written uses (T7.0–T11.5) are exclusively devoted to *going to write*. Satoru's first oral use was recorded at T6.0 with an additional seven tokens (five types) produced between T7.0 and T7.5.

Eduardo shows what appears to be early creative use of *going to*. Journals from T1.5 display Eduardo's initial seven uses, showing five tokens and three person-number forms (more variation than any other learner). At T2.0 Eduardo's four uses are exclusively devoted to *going to write*. At T2.5, three additional uses of *going to write* appear and the use generalizes in two directions: to additional verb phrases that have the same function as *write about* (e.g., *I am going to tell a story, tell about, continue the story*) and to more general verbs (e.g., *eat, die, and see*). Those uses are the last that Eduardo shows of *going to write*, although *going to* is used with other verbs through T7.5. The function of *going to write* is assumed by *will write* immediately after the abandonment of *going to write* but is only used twice. Eduardo produces seven tokens (seven types) of *going to* in the oral sample, the first of which is *going to write* at T3.0.

Like Eduardo, Khaled begins to use *going to* with a range of verbs. The earliest uses are widely spaced at T.5, T2.5, and the first use of the introductory *going to write* at T5.0. Between T6.0 and T7.0, seven tokens (five types) appear. At the same time (T5.0–T6.0) *I'm gonna* + verb appears in Khaled's oral sample in a tape journal, an example of planned oral speech. Although the verbs vary, the use of *gonna* serves the same function orally as *I am going to write* does in writing, announcing the forthcoming topic, *I'm gonna read, try to explain, speak about*. At the end of the eighth month Khaled begins his exclusive use of *going to write* and this accounts for the 15 uses of *going to* in the next month and a half. Four uses of *going to be* are observed in the tenth month in response to an essay exam prompt, "What is life going to be like in 2040?" Four new verbs emerge simultaneously in a personal account that appears in both the oral and written corpus at T11.0. Three additional uses of *going to write* appear in T10.5–11.5.

These 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. Satoru's and Khaled's initial use of multiple verbs with *going to* followed by exclusive use of *going to write* shows that use of a formula may displace what seems to be relatively creative

use. To a lesser degree, Carlos's and Eduardo's use of the formula also causes a temporary cessation of *going to* with other verbs when it first appears.

The formulaic use of *I am going to write* seems to break down to a smaller—but likely still unanalyzed—unit, *going to*, as learners use different verbs and vary person and number. Only Satoru maintains his exclusive use of *I am going to write* to the end of the observation period from T7.0 to T11.5. His oral interviews show wider use of *going to* with four tokens and types attested at T7.5, just after the emergence of *going to* in writing. Those learners whose use of *going to* extends to other verbs apparently abandon *going to write*. This latter development may be due as much to rhetorical as to grammatical development and should be interpreted with caution.

How can interlanguage afford the exclusive (if temporary) use of *going to* for the *going to write* formula? One answer might be that because *will* emerges first and establishes itself relatively quickly, the system already has a future form that bears the communicative pressure of future meaning. On the other hand, learners may wonder what good a second form is when the target and interlanguage already have an invariant and transparent form, *will*, for the expression of the future, in accordance with the one-to-one principle (Andersen, 1984). By assigning *going to* to the *going to write* formula and function, learner systems may solve the problem of apparent semantic equivalents. The broader range of combinations in the oral corpus may indicate a further division of functions: Khaled's oral use of *gonna* with a range of verbs while reserving written uses exclusively for *going to write* may further distinguish functions and forms of the strings in the oral and written texts. (Khaled is the only learner in the study to use *gonna*. Whether it serves as the contracted form of *going to* or an independent third expression of futurity in his interlanguage is a question for further investigation.)

Input and Intake

How are the patterns of emergence related to instructional input? In the more general case of the emergence of *will* and *going to*, neither timing nor frequency of input seems to affect learner production. According to the instructional logs, *be going to* is introduced a full month before *will* in Level 1 (O'Neill, Anger, & Davy, 1981, *AKL: Beginning*) and is separated from the introduction of *will* in Level 2 by the program syllabus and the textbook (O'Neill, Kingsbury, Yeadon, & Cornelius, 1978, *AKL: Intermediate*). This textbook series maximizes the differences between *will* and *be going to*. *Be going to* is introduced as a general future cued by adverbials such as *tomorrow* and *next week*. *Will* is introduced in polite requests and offers, then four chapters later as a future form. In addition to the textbooks, the interviews suggest that the interviewees (ESL teachers in the same program) may prefer the use of *going to* to *will* in some contexts, as in (1):

- (1) Interviewer: *What are you going to do for Christmas Break?*
Carlos: *I will go back to my country on December 8.* [T3.0]

If timing and frequency of instructional input do not determine order of emergence or frequency of use of *will* and *going to*, what does? From a formal perspective, both are completely regular in English, although *will* might have an advantage if learners regard it as lexical rather than grammatical, because lexical devices precede morphology in expression of temporal reference (Bardovi-Harlig, 2000). Using *going to* may be a harder task just because it is longer than *will* and so involves more phonological units to be sequenced (cf. Ellis, 1996). Learners must also inflect the auxiliary *be*, which may add to the formal complexity of this construction. From the perspective of the association of form and meaning, it is also interesting to note that the emergence of *will* before *going to* follows what Leech (1971, p. 64) identified as the “order of importance” of future expressions, suggesting that *will* represents a “neutral” future in English (p. 52).

As Ellis observes, instructional input often does not reflect language use outside the classroom. The *I am going to write (about) + NP* formula is clearly an example of that. However, there is an experimental value in this difference because some classroom input can be traced in learner output. Additionally, although the classroom experience is a shared one, not all learners had the same reaction to the input.

The use of the formula *going to write* seems to be more susceptible to instructional input than the use of *going to* more generally, although its use too seems to be tempered by the acquisitional stage of the learners and other variables. All the higher rate users who use formulas began instruction together and were in the same classes for the first two levels. Instruction on *going to* began at T1.5 in Level 1. The writing logs and records show that at T2.0 the Level 2 writing teacher used three model compositions with the *going to write* formula at the end of the opening paragraph. The first paragraph of the model composition about admiring one’s father concluded, “Although it is difficult to list all the reasons why I admire my father, *I am going to write about* two of them in this composition.” The other two models on “Mother” and “Halloween” were identical. Two of the learners, Carlos and Eduardo, showed immediate use of the formula coincident with instruction. As described earlier, they both abandoned the formula after limited use when *going to* generalized to other verbs.

Satoru and Khaled were in the same classes in Levels 1 and 2 as Carlos and Eduardo, but their use of *going to write* is delayed by three or more months compared to the first instructional period and emergence in Eduardo’s and Carlos’s samples. Eduardo, Khaled, and Satoru also shared the same Level 3 classes. According to the logs, future was not taught in Level 3, and there is no evidence from the process writing book of the *going to write* formula. (It is possible that the teacher used a model not included in the logs.) Khaled passed from Level 3 to Level 4, whereas Satoru repeated Level 3. However, it is in this period from T7.5 to T10.0 for Satoru and T9.0 to T10.0 for Khaled that the most concentrated use of the formula is found. There is no obvious link to input in this later stage, but, like Eduardo and Carlos, Satoru and

Khaled were exposed to the formula in Level 2. It is possible that individual teacher comments on their essays led them to use the formula, but that information is not included in the corpus because no corrected essays were collected.

There is one other thing to consider in the case of Khaled. At the time of instruction and modeling, Khaled was using another formula with the same function. Between T2.0 and T7.5, Khaled began 43 texts with *I want to speak about*, supplemented by five uses of *I will write about* at T3.0. There is a slight change in the formula after winter break at T5.0, when the final 23 tokens are reduced to *I want speak about* (perhaps simplified by analogy to *will?*). T8.0 sees his first use of *going to write* (in a nongraded journal), followed by two uses of *I want speak about*, which dies out completely after T8.5. At T9.0 there are 10 uses of *going to write* in 2 weeks. It appears that the earlier formula might have blocked the acquisition of another one with the same function. No other learner in the corpus showed use of an introductory formula that preceded the *going to write* formula.

DISCUSSION

In answer to the first question concerning acquisitional sequences, the data show that some learners do show formulaic use in the acquisition of future expression. For 5 of the 16 learners, the use of *I am going to write* stands out. Their production over the months of observation show that the formula breaks down into smaller parts, from the full *I am going to write about* to the core *going to* where not only the verb but also person and number vary. This seems to be an example of learner production moving along the formulaic-creative continuum (Weinert, 1995).

In answer to the second question concerning instructional input, we see that learners showed different responses to instruction, as expected from the literature. Eight of the 16 learners shared the same grammar and writing class, but three of them used *going to* rarely (fewer than 10 tokens), whereas the remainder were higher frequency users and used the *going to write* formula for at least a limited time. In contrast, three learners from the other instructional group were also high-frequency users of *going to* but showed no formulaic use. Instruction is just one source of input in a mixed environment, and it, like input from other sources, is subject to influence of salience, complexity, timing with respect to learner developmental level, and communicative need.

Examining written and oral production adds depth to the study of formulaic use. Through the study of written texts we see that formulas can be used to increase fluency in different modes, illustrating how some learners may learn discourse structure for compositions. Learners who used *going to write* may have enjoyed not only a grammatical advantage but also a rhetorical advantage in that the use of the formula at the beginning of a composition or journal entry may have freed planning time for the propositional content of the text.

The data also suggest that formulaic use may be found in interlanguage beyond the earliest stages. Learners can employ formulas at any stage—although they might be harder to detect—especially as new form-meaning associations are forged. In terms of the tense-aspect system, we might expect to find periodic use of formulas by some learners as new morphology emerges. One question for further research is whether the same learners rely on formulas in different areas of grammar or within a grammatical subsystem, such as tense and aspect. One learner who showed formulaic use of the present perfect in the longitudinal study produced no tokens of *going to*; thus, this question remains for further research.

This brief account of formulaic language in the tense-aspect system suggests that, when we adopt as a starting point in SLA research the hypothesized sequence of development from formula through limited scope to construction, we look through a wider lens than one that excludes formulaic use. Although the use of formulaic language seems to play a limited role in the expression of future, its influence is noteworthy. Incorporating the investigation of formulaic use and its relation to input in future inquiry will yield a fuller account of L2 acquisition.

REFERENCES

- Andersen, R. W. (1984). The one-to-one principle of interlanguage construction. *Language Learning*, 34, 77–95.
- Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Oxford: Blackwell.
- Comrie, B. (1985). *Tense*. New York: Cambridge University Press.
- Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, 18, 91–216.
- Ellis, R. (1990). *Instructed second language acquisition*. Oxford: Blackwell.
- Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning*, 24, 287–297.
- Haegeman, L. (1989). *Be going to* and *will*: A pragmatic account. *Journal of Linguistics*, 25, 291–317.
- Leech, G. N. (1971). *Meaning and the English verb*. London: Longman.
- Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2. *Studies in Second Language Acquisition*, 21, 49–80.
- O'Neill, R., Anger, L., & Davy, K. (1981). *AKL: Beginning*. London: Longman.
- O'Neill, R., Kingsbury, R., Yeadon, T., & Cornelius, E. T., Jr. (1978). *AKL: Intermediate*. London: Longman.
- Reichenbach, H. (1947). *Elements of symbolic logic*. Berkeley: University of California Press.
- Schmidt, R. (1992). Psychological mechanisms underlying language fluency. *Studies in Second Language Acquisition*, 14, 357–385.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16, 180–205.