Uses and Functions of Formulaic Sequences in Second-Language Speech: An Exploration of the Foundations of Fluency

David Wood

The Canadian Modern Language Review / La revue canadienne des langues vivantes, Volume 63, Number 1, September/septembre 2006, pp. 13-33 (Article)

Published by University of Toronto Press
DOI: *https://doi.org/10.1353/cml.2006.0051*

Uses and Functions of Formulaic Sequences in Second Language Speech: An Exploration of the Foundations of Fluency

David Wood

Abstract: Formulaic sequences are fixed combinations of words that have a range of functions and uses in speech production and communication, and seem to be cognitively stored and retrieved by speakers as if they were single words. They can facilitate fluency in speech by making pauses shorter and less frequent, and allowing longer runs of speech between pauses. The present study was undertaken to identify the uses and functions of formulaic sequences in the development of speech fluency in narrative retelling in English as a second language (ESL). Spontaneous spoken narrative retells by ESL learners were analyzed for ways in which use of formulaic sequences may have facilitated fluency growth over a six-month period, be they pragmatic, functional, or strategic. Five categories of formula use emerged: repetition of a formula; use of multiple formulas to extend a run; reliance on one formula; use of self-talk and filler formulas; and use of formulas as rhetorical devices. These categories are illustrated by excerpts from transcripts of learner speech.

Résumé : Les expressions stéréotypées sont des combinaisons fixes de mots qui ont divers usages et rôles dans la production du langage et la communication. Elles semblent être cognitivement enregistrées et récupérées par les usagers comme s'il s'agissait d'un seul mot. Ces expressions peuvent améliorer la fluidité verbale d'un discours en diminuant la longueur et la fréquence des pauses, et en permettant de plus longues séquences de discours ininterrompues entre les pauses. La présente étude a été effectuée dans le but de déterminer l'usage et le rôle des expressions stéréotypées dans le développement de la fluidité verbale pour les exposés redits en anglais, langue seconde (ALS). On a analysé les exposés oraux redits spontanément par des étudiants en ALS afin de trouver comment les expressions stéréotypées ont favorisé le développement de la fluidité au cours d'une période de six mois, peu importe que les expressions soient pragmatiques, fonctionnelles ou stratégiques. Cinq catégories d'expressions se sont dégagées : la répétition d'une expression ; l'enchaînement de multiples expressions ; le recours à une

© 2006 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 63,1 (September/septembre), 13–33

seule expression ; l'emploi de la verbalisation intérieure et d'expressions de remplissage ; et l'emploi d'expressions comme figure de style. Pour chacune de ces catégories, on donne des exemples tirés de la transcription des discours des étudiants.

Introduction

The phenomena of fluency and formulaic sequences have been the subjects of growing interest over the past four decades. Research on fluency has concentrated mainly on measurable temporal variables in speech such as rate or speed of speech, pause times and frequencies, and the length of fluent runs of speech between pauses. This has provided reliable measures with which to determine fluency in first-language (L1) or second-language (L2) speech production. Research on formulaic sequences, linked with the growth of corpus analysis research and computer software, has tended to deal with the nature and functions of formulaic language. A consensus has developed that formulas are fixed strings or chunks of words that have a range of functions and uses in speech production and communication and seem to be cognitively stored and retrieved by speakers as if they were single words. In recent years, research on the role of formulaic sequences in actual communication has concentrated on L1 contexts and has been rather impressionistic.

The present study was undertaken to identify the role of formulaic sequences in L2 acquisition, particularly in the development of speech fluency. The spontaneous spoken narrative retells of a group of English L2 learners were analyzed for ways in which increasing and more effective use of formulaic sequences may have facilitated fluency growth over a six-month period. Fluency-enhancing uses of formulaic sequences were marked in the data and then categorized. The categories that emerged were varied and showed that speech fluency may be enhanced by use of formulaic sequences in particular functions in discourse, or by strategic use by speakers. Five categories of formula use emerged: repetition of formulas to extend a run of speech; use of multiple formulas to extend a run; reliance on repetition of one formula to facilitate fluent speech; use of self-talk and filler formulas; and use of formulas as rhetorical devices in spoken narrative.

Fluency

The first attempts to examine speech fluency occurred when Goldman-Eisler (1967, 1972) looked at the temporal variables of speech rate,

articulation rate, and pause phenomena. With the passing of time and the evolution of technology to facilitate speech recording and analysis, a consensus has emerged that the key temporal variables linked to fluency are rate or speed of speech, pause phenomena, and length of runs between pauses.

Rate of speech, measured as syllables uttered per minute or second, tends to increase over time along with other measures of fluency or to correlate with judges' perceptions of fluency (Freed, 1995; Riggenbach, 1991; Towell, 1987; Towell, Hawkins, & Bazergui, 1996).

Two aspects of pausing are usually measured in fluency research: duration and frequency, and syntactic location. Studies have shown that pause times in L2 speech are generally longer than in L1 speech (Lennon, 1984; Möhle, 1984). Longitudinal research has found that L2 pause times and frequencies reduce over time (Freed, 1995; Lennon, 1990a; Riggenbach, 1991). In studies of pause location in L1 and L2 speech (Dechert, 1980; Deschamps, 1980; Lennon, 1984), and research correlating pause location with judgments of fluency (Riggenbach, 1991), it has been shown that native speakers or highly fluent L2 learners most often pause at clause junctures or between non-integral parts of a clause, while lower-fluency speakers tend to pause within clauses. The pause profile associated with fluency may be an effect of the cognitive processing loads that accompany fluent speech; producing whole clauses and chunks of words directly from long-term memory might help avoid the slow process of producing utterances using controlled processing and word-by-word assembly.

In the present study the key variable of speech associated with fluency is the mean length of runs, measured as syllables uttered between pauses. Longer runs are a key indicator of fluency (Freed, 1995; Lennon, 1990b; Möhle, 1984; Raupach, 1980). Together with the pause profile associated with fluency, longer runs are likely a result of fluent speakers having a larger repertoire of formulaic sequences to help balance skills, attention, and planning during speech. The data in the present study indicate that certain uses and functions of formulaic sequences appear linked to increased length of runs.

Formulaic sequences

Definitions

In 1983 Pawley and Syder noted a connection between formulas and fluency; they observed that we tend not to make use of the endless lexical and grammatical potential of language. Instead, the norm is to use

standard predictable phrases such as *How are you?* Or *Will you marry me?* rather than grammatical but communicatively unlikely ways of expressing the same meaning or function such as *What is your current state of well-being?* or *Are you inclined to become my spouse?* Pawley and Syder were among the first to observe that language production may be based only partly on rule-governed formation of utterances from lexis through syntax, morphology, and phonology. In spontaneous speech, such a laborious method of language production would seem improbable, particularly in light of the limitations of human memory and attention.

Over time, developments in computer technology, corpus studies, and phraseology research have provided discourse-based evidence of how words tend to collocate and cluster. For instance, as long ago as 1988, Sinclair and Renouf (p. 153) gave the examples of the verbs *give* and *have*, which collocate with nouns in predictable ways, as in *give advice, give a look*, or *have a good look, have a deep longing*. Another example is the verb *make*, whose primary function is to accompany nouns such as *decision* and *arrangement*. Sinclair and Renouf also showed that structure-class words are often grammatically restricted. For example, *each* occurs with units of time, *of* leftward with *kind*, *part*, and *sort*. Phraseologists have pointed out that collocations cross a wide spectrum and can include phrasal verbs, prepositional phrases, and more (Mel'cuk, 1998).

Definitions of formulaic sequences centre on the notion that they are multi-word units of language and that, in spontaneous speech production, they are stored in and retrieved from long-term memory as if they were single lexical units (Nattinger & DeCarrico, 1992; Pawley & Syder, 1983; Read & Nation, 2004; Wray & Perkins, 2000). Formulaic sequences appear to be ubiquitous in speech. Altenberg (1998) remarks that 'what is perhaps the most striking impression that emerges ... is the pervasive and varied character of conventionalized language in spoken discourse ... from entire utterances operating at discourse level to smaller units acting as single words and phrases' (p. 121).

Formulaic sequences in acquisition and performance

A relatively small number of studies have examined the role of formulaic sequences in adult L2 acquisition. Yorio (1980) found that formulaic language was used by adults as a production strategy, to save effort and attention in speaking. In a case study of an adult L2 learner in a naturalistic context, Schmidt (1983) found the learner used increasing numbers and varieties of formulaic sequences while showing little development of grammatical and other aspects of language. Bolander

(1989), in a longitudinal study of adult L2 acquisition, found that learners used formulaic sequences containing particular language structures long before they could show they had actually acquired the structures themselves. The links between formulaic sequences and pragmatic competence have been researched. Coulmas (1979, p. 241) states, 'As they provide the verbal means for certain types of conventional action, their meanings are conditioned by the behavior patterns of which they are an integrated part,' and goes on to note that formulas help to facilitate unambiguous communication. Bygate (1988), in a study of adult learners of Swedish L2, found pragmatic uses of formulas to include repetition, questioning, agreeing, confirming, clarification, and focusing attention.

Formulas are probably cognitively stored and retrieved in various ways. They may be recalled on the basis of linear surface order of their parts, or by phonological units (Weinert, 1995). Consciousness, awareness, and noticing of formulas in input (Schmidt, 1990) might establish a pattern-recognition unit, which is then strengthened by frequent input, leading to single-step memory access. Frequent processing of a formula in working memory may lead to automatization or retrieval in a single-step process of remembering, as posited by instance theory (Logan, 1988). Perhaps formulaicity occurs as the result of a perception of meaning in chunks in input, which are then stored as wholes, and perhaps followed by fusion or combination of formulas, according to Weinert 1995. In the present study, in some instances it was possible to apply some of these concepts to explain why or how a particular use or function of formulaic sequences might have occurred.

There is some empirical evidence that formulaic sequences have functions in L2 speech fluency. Raupach (1984), in a study of an adult L2 learner, found formulaic sequences expressing functions and operating as complete clauses, allowing the speaker time and attention to anticipate the next utterance. They appeared to create a structure in which to insert novel items, allowing time for syntactic and lexical retrieval and encoding. Dechert (1980), in a study of narrative retells in L2, noticed that the most fluent students established 'islands of reliability' of ideas and language, around which they assembled the spoken narrative. A broader presentation of the possible functions of formulas in aiding speech production is that of Wray (2002, p. 97), who sees four main categories: enabling the manipulation of information; allowing a continuing flow of speech to occur while the conscious mind is focused elsewhere in communication; shortening the processing route of speech by bypassing the need for assembly of components or use of short-term memory; and signalling the organization of spoken discourse.

While these are logical possible functions of formulas in speech, there has been very little research that attempts to elaborate categories of use and functions grounded in actual L2 speech data in a specific genre. As noted by Moon (1998), formulaic sequences exhibit a great deal of flexibility and are often genre-specific, seeming to indicate that research on the functions and uses of formulas needs to focus on actual speech data within a particular genre. The present study was undertaken to do so, using data coded to determine which types of uses and functions formulaic sequences were put to by particular participants in achieving a particular task.

Method

Participants

In this study, narrative speech samples were collected monthly from 11 full-time intermediate-level intensive English as a second language (ESL) learners at a Canadian university over six months and were analyzed for types of use of formulaic sequences that facilitated fluency development. The participants were from three L1 groups: Spanish – two female, two male; Chinese (Mandarin) – one female, two male; and Japanese – two female, two male. At the outset there were two Chinese female participants, but one of their data sets was omitted from the analysis because the recordings were corrupted. Participants were selected on the basis of their oral proficiency scores on the ESL program placement test, which elicited mostly narrative talk. They had already been enrolled in the ESL program for at least 12 weeks, and also lived with Canadian families so were provided sustained opportunities for English input and communication outside of the classroom.

Speech samples

The prompts used to elicit the spoken narratives were three short silent animated films of similar length and plot complexity, shown with a three-month interval between viewings of the same film. Silent films permitted the learners to control what to pay attention to and what to retell, rather than trying to repeat spoken language heard in the films.

The three National Film Board of Canada film prompts for the retells were:

- For samples 1 and 4, *Neighbours* (Norman McLaren, 1952: 8 minutes), the story of two neighbours living peacefully until a flower

appears between them, after which they fight savagely, eventually kill each other, and end up in side-by-side graves, with two flowers like the original marking where they lie.

- For samples 2 and 5, *Strings* (Wendy Tilby, 1991: 10 minutes), the story of a woman living on the third floor of a building above a man on the second floor. One day the woman pours a bath, and the man prepares a meal as people arrive at his place and begin to play music. Water leaks from the bath into the man's apartment. He goes upstairs and attempts a repair, but the chandelier in his apartment breaks away and crashes. The musicians leave, he returns home to play his violin, while the woman resumes her bath.
- For samples 3 and 6, *The Cat Came Back* (Cordell Barker, 1988: 7 minutes 37 seconds), tells the story of a man tortured by a stray cat that comes to his house. He plays with it, but it destroys his cherished toy, and he makes repeated fruitless attempts to get rid of the cat, finally trying to blow it up, inadvertently killing himself. His corpse falls on the cat, which dies in turn. The man's soul flies to heaven screaming in frustration as nine souls of the cat follow him.

Procedures

For each monthly speech sample, the participants were shown one film one time in its entirety without pause, and they immediately recorded their retelling on tape in the university language laboratory. The films were not introduced and no language help was given; participants did not take notes, write a script, or stop, pause, or rewind their tapes. The samples were later transcribed and SpeechStation2 speech-analysis spectrograms were used to measure pauses.

Measurement of fluency gain

To quantify the growth of fluency and the increased use of formulaic sequences, two calculations were used in the present study: mean length of run (MLR), calculated by dividing the total number of syllables uttered for a speech sample by the number of runs between pauses; and formula/run ratio (FRR), calculated by dividing the total number of formulas in a sample by the number of runs.

Identification of formulaic sequences

Native speaker judgment (NSJ) was used to identify formulaic sequences in the data. Corpus analysis computer software was not used because,

as Wray (2002, pp. 25, 27, 28) points out, its reliance on frequency counts makes it difficult to determine the distribution of some types of formulaic sequences. In addition, as noted by Moon (1998), formulaic sequences are quite variable and can be specific to certain genres or types of discourse. For the present study, the specific nature of the type of speech elicited and the relatively small number of samples from each participant meant that frequency alone could not suffice as a criterion for identifying formulaic sequences. Some formulas may be used only once or idiosyncratically in such a situation.

The judges were graduate students in applied linguistics who had studied the background literature on formulaic sequences. They participated at the outset in a group benchmarking session in which four randomly selected retells were listened to and their transcripts read, and shared in coding formulaic sequences and building a common sense of how to apply the judgment criteria (see the criteria below). After the benchmarking, the judges took two weeks to work independently on judging and coding.

Limitations of NSJ have been pointed out by some researchers, including Wray (2002, p. 23), but the methodology used here takes a number of measures to address the limitations. One concern is that NSJ has to be restricted to smaller data sets; the corpus used in this research is comparatively small, with 11 participants and roughly 26,000 words. A second concern is that NSJ may cause inconsistent judgment as the result of fatigue or alterations in judgment thresholds over time. This issue was addressed in the present study by having judges individually listen to as well as read the transcripts, providing two parallel sources of data for judgment, and by allowing the judges several weeks to work independently on the judging. In this way, the pressure of time limits and variations in focus in the judging task were reduced. The third concern is that there may be variation between judges; this consideration was addressed by having a group benchmark identification session before individual judging in which a common set of criteria was implemented. In this way, judges participated in shared experience on which to base their individual decisions. The fourth and fifth concerns with NSJ are that there may not be a single answer to the question of what to search for and that decisions may be based on intuitions if the appropriate knowledge is not at the surface level of awareness. In response to these concerns, the group benchmark session was designed to help establish a common approach to the task. Furthermore, the judges in the present study read the most salient literature on criteria for identifying formulaic sequences, and these were used in identifying items as formulaic sequences. Features of the recorded speech such as speed and

volume changes were also used as guides. These procedures helped to avoid reliance on intuition and kept the available knowledge about formulaicity at the surface level of consciousness during judgment.

The most compelling reason for using native speaker judgment was the fact that this was a corpus of spoken language and the act of *listening* to speech and noting intonation and pause patterns cannot be done by machine. Human judgment was required if all the factors relevant to formulaicity in speech were to be determined.

Criteria for identification and classification

Five general criteria were applied in deciding whether a sequence was a formula. Wray states that, when identifying formulaic sequences in speech, 'it may simply be that identification cannot be based on a single criterion, but rather needs to draw on a suite of features' (2002, p. 43). It is important to stress that no particular criterion or combination of criteria was deemed as essential for a word combination to be marked as formulaic, and judgments were made on the basis of one, several, or all of these.

1. *Phonological coherence and reduction.* In speech production, formulaic sequences may be uttered with phonological coherence (Coulmas, 1979; Wray, 2002), with no internal pausing and a continuous intonation contour. Phonological reduction such as phonological fusion, reduction of syllables, or deletion of schwa may be present as well. All are common features of the utterance of the most high-frequency phrases in English, but are much less so in low-frequency or more imaginatively constructed utterances, according to Bybee (2002). Phonological reduction can be taken as evidence that 'much of the production of fluent speech proceeds by selecting prefabricated sequences of words' (Bybee, p. 217). This criterion was important in this research and gave the expert judges a readily identifiable characteristic of the speech samples to attend to.
2. *The taxonomy used by Nattinger and DeCarrico (1992)* includes syntactic strings such as 'NP+Aux+VP ..., collocations such as *rancid butter* and *curry favor*, and lexical phrases such as *how do you do?* and *for example*, that have pragmatic functions' (p. 36). This taxonomy is limited and not necessarily applicable in every case; it was used as a guide to possible formulaicity. For example, if a sequence that was uttered with phonological coherence matched other criteria and fitted into a category in this taxonomy, it might be marked as formulaic.

3. *Greater length/complexity than other output.* Examples include using *I would like ...* or *I don't understand*, while never using *would* or negatives using *do* in other contexts. Judges were able to see and hear the entire output of a particular participant to help in applying this criterion. In combination with other criteria, this was a key marker of formulaicity.
4. *Semantic irregularity, as in idioms and metaphors.* Wray and Perkins (2000, p. 5) note that formulaic sequences are often composed holistically, like idioms and metaphors, and not semantically. Examples were apparent in the background literature for the judges, and many formulas readily match this criterion. However, in itself it is not sufficient – sequences that appeared semantically irregular but not canonical as idioms or metaphors could simply be a sign of semantic inaccuracy on the part of a participant, and other criteria were brought to bear on the judging.
5. *Syntactic irregularity*. Formulaic sequences tend to be syntactically irregular in two ways: (1) There is a restriction on manipulation. For example, it is impossible to pluralize *beat around the bush*, or passivize *face the music*, or say *you slept a wink* or *feeding you up.* (2) Normal grammatical restrictions may be ignored. Examples are the sequences that contain an intransitive verb + direct object such as *go the whole hog* or other syntactic outliers such as *by and large* (Wray and Perkins, 2000, p. 5). This criterion was readily applied to some sequences, but, as is the case with semantic irregularity, it was important to check syntactically irregular sequences against other criteria on this list.

Items were accepted as formulaic if two or all three of the judges were in agreement. Sequences that were idiosyncratic or deviant in some way from typical native speaker production were accepted, given that the process of judgment involved balancing a range of criteria. It was assumed that non-native-like productions that still appeared to meet all or most of the criteria were indications of several phenomena of peripheral relevance to the present study. For instance, a sequence might have been stored and retrieved as a whole with some of its features misperceived; examples are *what's happened* instead of *what happened,* or *thanks God* instead of *thank God.* Another possible cause of non-native-like production might be the communicative and cognitive stress of the retell situation as participants needed to recall events seen in the film and produce a running narrative at the same time, resulting in slips of articulation or gaps and substitutions for some elements of some formulaic sequences. Thus a sequence could meet the judging criteria

and still be deviant, idiosyncratic, misperceived, and stored with errors, or misarticulated as a result of stress.

Results: Categories and examples of uses and functions

Retells of the same film prompt by each participant were examined to determine categories of formula use. That is, for samples 1 and 4, *Neighbours*, for samples 2 and 5, *Strings*, and for samples 3 and 6, *The Cat Came Back*. This analysis focused on transcripts in which mean length of runs (MLR) and the ratio of formulaic sequences to runs (FRR) showed improvement from the first retell to the second, three months later. Table 1 illustrates that development did not always follow this pattern for individual participants, but that the whole group shows improvement from the first retell of each film to the second. Each participant is identified by first name (a pseudonym), L1 (*J* for Japanese, *S* for Spanish, *C* for Chinese), and gender (*F* for female, *M* for male).

All instances in which the use of formulas facilitated longer runs were marked in the data, and types of fluency-facilitating uses were categorized. Some uses were related to speech strategies, or a result of the communicative and cognitive load of the task. For example, a participant might rely heavily on the use of a particular formulaic sequence, string together several short formulaic sequences, or repeat one formula several times consecutively to extend a run and avoid

TABLE 1
Mean length of runs and formula/run ratio development from first to second retelling of each film

Film #	Film 1				Film 2				Film 3			
Variable	MLR		FRR		MLR		FRR		MLR		FRR	
Retell #	1	2	1	2	1	2	1	2	1	2	1	2
JF Yuka	2.8	3.7	0.22	0.40	2.5	1.8	0.45	0.31	3.0	3.0	0.33	0.38
JF Natsuko	3.7	3.6	0.36	0.31	3.1	3.6	0.23	0.27	4.0	4.1	0.33	0.29
JM Isamu	2.7	3.7	0.06	0.22	2.6	3.8	0.12	0.16	2.5	4.2	0.07	0.25
JM Jun	4.1	4.3	0.56	0.58	3.6	4.3	0.47	0.47	4.1	5.1	0.51	0.57
SF Sally	2.0	3.2	0.08	0.33	2.9	3.0	0.22	0.29	3.4	3.3	0.39	0.36
SF Lilia	5.5	4.0	0.32	0.40	3.9	4.4	0.28	0.40	3.9	5.6	0.44	0.59
SM Carlos	2.3	4.2	0.09	0.30	3.6	3.5	0.26	0.24	5.5	4.2	0.30	0.23
SM Miguel	4.9	4.1	0.41	0.50	5.0	4.3	0.44	0.48	4.3	4.3	0.45	0.41
CF Meiling	4.5	5.3	0.35	0.51	4.8	5.2	0.47	0.61	5.1	5.1	0.69	0.62
CM Liang	3.2	3.8	0.13	0.26	3.4	4.0	0.22	0.34	3.4	3.9	0.32	0.37
CM Lin	3.8	6.0	0.33	0.61	4.4	4.4	0.21	0.46	5.0	4.7	0.41	0.53
All subjects	3.6	4.2	0.27	0.40	3.6	3.8	0.31	0.37	4.1	4.3	0.39	0.41

pausing. Other uses related to functions of formulaic sequences in the discourse, such as using the sequences in self-talk, in fillers, or as a means to structure the discourse by marking chronological sequence or the beginning and ending of segments of the narratives. Overall, five categories of formula use emerged as contributing to increased length of runs. The first three are more strategic or performance-related, while the last two are more pragmatic and functional:

1. Repetition of formulas in a run
2. Use of multiple formulas to extend a run
3. Reliance on one formula or filler repeatedly
4. Use of self-talk and fillers
5. Use of formulas as rhetorical devices

Presented below are exemplars of these types of use of formulas. These are not the only such cases in the data, nor are all 11 participants represented, although all of them used formulaic sequences in these ways. Pauses are represented by their duration in seconds in parentheses in the excerpts.

Repetition of formulas in a run

In some cases, participants were able to extend the length of runs by simply repeating a particular formula consecutively within one run.

Natsuko – Japanese female – Samples 3 and 6, *The Cat Came Back*

While Natsuko is not more fluent overall in Sample 6 on these measures, she does show effective use of formulas in this sample. In particular, she uses a strategy of repeating a formulaic sequence within a run to extend it. For example, when describing one of the cat's many returns to the house in the film she uses the formula *came back* two times in one run:

And he *came back* the cat *came back* to the his house and ah

This results in a run of 13 syllables, only one of which is a filler non-lexical item, *ah.*

She also repeats a self-talk formula later:

I forget I forget the order but maybe the f he *went to the forest*

Here she appears to think aloud, buying time to recall the next event in the narrative and uses a very simple subject + verb formula to repeat her lack of clear recall. It helps her to produce a 19-syllable run.

Lilia – Spanish female – Samples 3 and 6, *The Cat Came Back*

In Sample 3, when describing the man's unintended suicide, Lilia only manages to produce two short non-formulaic runs, separated by a long 2.3-second pause:

finally he (2.3) he died

However, in Sample 6 she extends the description considerably and helps to extend one run and create a third one by repeating an adverb/adjective formula:

the dynamite ah (0.8) ah caused a (0.5) motion where the man died (0.4) and when he's *really happy* he's hysteric he's *really happy* because (0.5) he's *really happy*

Perhaps it is the nature of formulaic sequence retrieval that a single short formula can be uttered several times in sequence more or less effortlessly. This strategy can allow a large string of discourse to be produced with minimal strain on formulating capacity.

Use of multiple formulas to extend a run

A typical feature of many of the longer runs in later narrative retells in the data was the linking of formulas to create a longer run. Generally, these were two-word, shorter collocations strung together or used mixed with non-formulaic talk to lengthen runs by several syllables. For instance, in the film *Strings* several participants described the old man in the story *making music by himself in his room,* a combination of three short two-word formulas *making music, by himself,* and *in his room.* This produces a very fluent 10-syllable run.

Liang – Chinese male – Samples 2 and 5, *Strings*

Liang manages to produce a long run using two connected formulas in Sample 5, while to describe the same event earlier in Sample 2, he uses two formulas but needs to separate them with a short pause:

Sample 2: *cross the floor* (0.5) *to second floor*
Sample 5: *from her room to his room*

It is not clear why the hesitation occurs in Sample 2, but it may be that he is caught trying to recall the destination of the water crossing the

floor to the old man's room? To the other floor? Which number was it? In Sample 5 he avoids the dilemma by simply stating that the water moved from one place to another, and he utters one long and fluent run.

Liang – Chinese male – Samples 3 and 6, *The Cat Came Back*

Liang makes productive use of the practice of stringing multiple formulas in a single run. When describing the solitary music-making of the man in the film in Sample 3, he remarks only:

and *make a music*

Instead of this very simple and flawed single formula run, in Sample 6 he expands the picture and manages an 11-syllable run of considerably greater complexity and descriptive effect:

he's *make music by himself in his room*

It appears that after three months Liang is sufficiently more fluent and/or confident to take a risk and describe in a richer way, using formulaic sequences to produce longer runs.

Lilia – Spanish female – Samples 3 and 6, *The Cat Came Back*

Lilia also uses several formulas in a single run more effectively in Sample 6:

he *put outside* and ah *close the door*
always *try to far away* the cat but he *couldn't do it*

In both cases, Lilia is able to keep the discourse flowing in more complete units and avoid pausing, while producing runs of 9 and 14 syllables in these cases.

Reliance on one formula

In some instances participants created an illusion of increased fluency by relying heavily on one simple formula throughout a later narrative retell. To introduce the next action in the story, for example, it was common to use *and then*, or *and next*.

Sally – Spanish female – Samples 1 and 4, *Neighbours*

In Sample 4 Sally appears to use the formula *in the middle* as an anchor or a strategy to avoid long pauses or breakdown, while working to recall film plot content or to formulate the subsequent speech segment. This helps her to increase the mean length of runs for the segment of the narrative from 2.0 in Sample 1, to 3.2 in Sample 4, while boosting her formula/run ratio from .08 to .33.

in the middle of (0.5) the land (0.9) *in the middle* or between *in the middle of* (0.5) the land (0.9) *in the middle* or between ... continues *in the middle* (1.1) *the middle of* their houses

Lilia – Spanish female – Samples 2 and 5, *Strings*

In Sample 5 Lilia presents a clear case of repeating variations on the same formula within a run and with phonological coherence so that pausing is avoided completely:

um the old man (2.0) in in her *in his house in this moment in the same moment*

This lengthy but lexically sparse run helped Lilia to achieve the impressive gains in MLR and FRR between Samples 2 and 5.

Yuka – Japanese female – Samples 3 and 6, *The Cat Came Back*

the cat (2.0) but (0.6) still (2.0) the cat (0.6) *come back* (0.7) *came back* again (0.5) *come back* again (0.7) *comes back* again

Here in her sixth sample Yuka, while varying the form of the main verb to use in this instance, still manages to keep the rhythm of the narrative moving. This repetition strategy minimizes the length of the hesitations in the segment.

Use of self-talk and fillers

A particularly noticeable feature of the speech samples in this data set is participants' increased use over the six samples of self-talk and filler formulas. Such formulas include self-referential collocations as *I know*, or *I think*, or *I guess.* Also included in this category are long strings used

for self-talk or circumlocution such as *I don't know*, or *I don't know the thing's name.*

The following are some examples of how participants used self-talk and filler formulas to lengthen runs.

Lilia – Spanish female – Samples 2 and 5, *Strings*

While discussing the moral of the story in depth and at great length in Sample 5, Lilia uses a large number of formulaic sequences not present in her Sample 2. Among these are the self-talk sequences *I don't know* and *I think so.* The latter is particularly noteworthy as a formulaic sequence because she means to say *I think (it is ...)* but produces *I think so (it is ...).* This is likely a sign that she has acquired the three-word *I think so* sequence as a whole and is not using *I think that*, or *I think*, as syntactic rules would require.

Liang – Chinese male – Samples 3 and 6, T*he Cat Came Back*

As was the case in other later samples, Liang uses the self-talk formulaic sequence *I think* to his fluency advantage in Sample 6, as compared to Sample 3. In fact, he uses it twice in two adjacent runs at one point, extending both runs by two syllables, likely with minimal effort:

> because all of the souls follow his ah souls *I think* (1.2) *I think* a cat's souls always to follow

It may be that the 1.2-second pause between these runs is a reflection of the need to formulate the last part of the second run; perhaps Liang has difficulty with this approach, as the syntax error in it suggests. The 1.2 second pause combined with the formula *I think* before and after the pause may have given him enough time in real-time speech to compose the last part of the second run.

Use of rhetorical devices

In some cases, participants were able to organize the information in later retellings by using rhetorical marker formulas. They used beginning formulas such as *at the beginning*, narrative move markers such as *when the story is go ahead*, and endings such as *that is the end of the story.* All add greatly to the length of runs as well as to the effectiveness of the storytelling.

Sally – Spanish female – Samples 1 and 4, *Neighbours*

In Sample 1 Sally produces quite dysfluent speech with few formulaic sequences of any type. However, in Sample 4 she uses two common rhetorical formulas to mark the movement of the narrative effectively: *beginning of the ...* and *and this is the end of the story.* She may have had greater control of the task and the speech required and was able to retrieve these rhetorical formulas to frame the narrative.

Carlos – Spanish male – Samples 1 and 4, *Neighbours*

Like Sally, Carlos uses few formulas in Sample 1 and produces choppy text of very short runs. Three months later in Sample 4, however, he uses a starter formula *the start the history*, and later the following:

when the history (.3) *go ahead*

Carlos appears to have begun to acquire a repertoire of formulaic sequences to mark the progress of the narrative. In the first case they are simple and inaccurate, and in the second case preceded by a short pause, but they show the beginning of an ability to use such units of speech with some effectiveness.

Discussion and conclusion

The use of formulaic sequences played a clear role in facilitating the development of speech fluency over time in this study. In particular, the participants employed a range of uses and functions of the sequences to extend the length of runs between pauses – a key indicator of increased fluency. One theme arising from these data is the complexity of human speech and the varying routes available to arrive at the same speech goal. The variability is explained by the fact that the data here are samples of real-time, real-life performance under the constraints of cognitive load, external situational factors, and socio-cultural issues.

The use of formulaic sequences occurs in certain classifiable ways in these data. Participants often repeated formulaic sequences to extend the length of runs, or they strung together multiple formulas to extend runs. In some cases it may be that recall of one lexical item at the beginning of a sequence triggered the recall of several more sequences beginning with the same lexical item; perhaps this is evidence of a cohort effect in lexical retrieval in which retrieval or exposure to the initial phoneme of a word activates words in the lexicon with that initial phoneme (Ellis,

2002). Sometimes the speakers used one particular formula repeatedly, perhaps to allow for some controlled processing or conceptualizing or recall of film content while continuing to speak. Formulas were used in some instances for self-talk or fillers in the discourse, perhaps showing what McCafferty (1994) calls a tendency to convert outer or social speech to inner speech. Formulas were also used as rhetorical devices to organize the discourse of the narrative retells, perhaps serving as anchors for the language and content of the following chunk of the story, similar to Dechert's (1980) notion of 'islands of reliability.'

These categories of formula use expand somewhat on those of Wray (2002, p. 97). Wray's summary of the value of formulaic sequences in speech production focuses mainly on their roles in controlling the flow of information, allowing time for mental processing of other aspects of speech and signalling the organization of speech. The analysis in the present study adds the functions of self-talk and extending the length of runs of fluent speech.

Types of uses of formulaic sequences may follow a path of development, as posited by Raupach (1984). At early stages of fluency, frequent hesitation and use of non-lexical fillers are used to give time for processing to be completed. Later, the more advanced speaker is likely able to use chunks of formulaic speech and lexical 'hedges' as ways of buying time for processing. Further research on the uses of formulas at various stages of fluency development could provide insight into mental processing in spontaneous L2 speech, and aid in the elaboration of pedagogical plans.

Future research needs to establish more clearly what constitutes a formulaic sequence. With written data or L1 speech data, corpus analysis software can help. For formulas in L2 spoken data, we have the criteria and classifications set forth over the years by Coulmas (1979), Nattinger and DeCarrico (1992), Wray and Perkins (2000), and others, or phraseological classifications. But we still must depend on listener judgment rather than a firmer set of standards. Researchers need to work toward clear definitions and criteria for identifying particular subsets of formulaic sequences in speech, especially in L2.

David Wood has been working in language education for many years, most recently teaching applied linguistics at Carleton University, University of Victoria, and Naruto University of Education in Japan. He is the author of the textbook *Making the Grade* and has a PhD from University of Ottawa. His research interests are formulaic sequences, spoken language, and teacher education.

Acknowledgements

I thank the editors of this volume and the two anonymous reviewers for extremely helpful comments and advice on earlier drafts of this article. Thanks also go to the participating students.

References

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 101–122). Oxford: Clarendon.

Barker, C. (Director). (1988). *The cat came back* [Motion Picture]. Canada: National Film Board of Canada.

Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L.K. Obler (Eds.), *Bilingualism across the lifespan: Aspects of acquisition, maturity, and loss* (pp. 73–86). Cambridge: Cambridge University Press.

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

Bygate, M. (1988). Units of oral expression and language learning in small group interaction. *Applied Linguistics, 9*(1), 59–82.

Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics, 3*, 239–266.

Dechert, H.W. (1980). Pauses and intonation as indicators of verbal planning in second-language speech productions: Two examples from a case study. In H.W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 271–285). The Hague: Mouton.

Deschamps, A. (1980). The syntactical distribution of pauses in English spoken as a second language by French students. In H.W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 255–262). The Hague: Mouton.

Ellis, N.C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*, 143–188.

Freed, B.F. (1995). What makes us think that students who study abroad become fluent? In B.F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123–148). Philadelphia: John Benjamins.

Goldman-Eisler, F.G. (1967). Sequential temporal patterns and cognitive processes in speech. *Language and Speech, 10*(2), 122–132.

Goldman-Eisler, F.G. (1972). Pauses, clauses, sentences. *Language and Speech, 15*(2), 103–113.

Lennon, P. (1984). Retelling a story in English. In H.W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 50–68). Tubingen: Gunter Narr Verlag.

Lennon, P. (1990a). The advanced learner at large in the L2 community: Developments in spoken performance. *International Review of Applied Linguistics in Language Teaching, 28*, 309–321.

Lennon, P. (1990b). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*(3), 387–417.

Logan, G.D. (1988). Towards an instance theory of automatisation. *Psychological Review, 95*, 492–527.

McCafferty, S.G. (1994). The use of private speech by adult ESL learners at different levels of proficiency. In J.P. Lantolf & G. Appel (Eds.), *Vygotskian approaches to second language research* (pp. 117–134). Norwood, NJ: Ablex.

McLaren, N. (Director). (1952). *Neighbours* [Motion Picture]. Canada: National Film Board of Canada.

Mel'cuk, I. (1998). Collocations and lexical functions. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 23–53). Oxford: Clarendon.

Möhle, D. (1984). A comparison of the second language speech production of different native speakers. In H.W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 26–49). Tubingen: Gunter Narr Verlag.

Moon, R. (1998). Frequencies and forms of phrasal lexemes in English. In A.P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 79–100). Oxford: Clarendon.

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

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R.W. Schmidt (Eds.), *Language and communication* (pp. 191–226). New York: Longman.

Raupach, M. (1980). Temporal variables in first and second language speech production. In H.W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 263–270). The Hague: Mouton.

Raupach, M. (1984). Formulae in second language speech production. In H.W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language production* (pp. 114–137). Tübingen: Gunter Narr Verlag.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 23–35). Amsterdam: John Benjamins.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes 14,* 423–441.

Schmidt, R.W. (1983). Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137–174). Rowley, MA: Newbury House.

Schmidt, R.W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140–160). New York: Longman.

SpeechStation2 [Computer software]. (1997). Somerville, MA: Sensimetrics Corp.

Tilby, W. (Director), & Verrall, D. (Producer). (1991). *Strings* [Motion picture]. Canada: National Film Board of Canada.

Towell, R. (1987). Variability and progress in the language development of advanced learners of a foreign language. In R. Ellis (Ed.), *Second language acquisition in context* (pp. 113–127). Toronto: Prentice-Hall.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84–119.

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

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 and Communication, 20*, 1–28.

Yorio, C. (1980). Conventionalized language forms and the development of communicative competence. *TESOL Quarterly, 14*(4), 433–442.