FORMULAIC SPEECH IN THE L2 CLASSROOM: AN ATTEMPT AT IDENTIFICATION AND CLASSIFICATION

Marie Girard and Claude Sionis

Abstract

This study looks into one context of Formulaic Speech (FS) usage: The partial L2 immersion class.

It tries to define and classify FS according to Raupach's contextual list (1984) and lexical criteria as well as differentiating it from creative speech. FS is presented mostly as a pragmatic concept challenging the usual conceptions of language acquisition as an analytical process. Also challenged is the idea that language production is based on analysis of the input followed by production out of parsed output.

In a Second Language Acquisition perspective, FS is shown as being a temporary stage of acquisition which, among other aspects, enables the speaker to reach idiomaticity in his or her L2 and thereby efficient communication with native speakers.

Keywords: Formulaic speech, Second language acquisition, Second language production, Partial immersion, Communicative competence.

1. Introduction

1.1. *Working hypotheses and intentions*

Observers of classroom language activity, be it free or teacher-directed, often feel that children repeat sequences of words without any knowledge of the internal structure of these sequences. This impression can only become a conviction after reading books like *The Study of Second Language Acquisition* by Ellis (1994), in which the author suggests that the use of formulaic speech may be one of the early developmental stages of second language acquisition, after what he calls the *silent period stage* and before the *semantic and structural simplification stage*.

Tackling the already widely addressed issues of Formulaic Speech (henceforth FS) classified according to syntactic/lexical criteria, and as a temporary state of acquisition, certainly seems a rather ambitious plan in one paper. The present study was prompted by the hypothesis that formulaic speech used by young L2 learners could be more than dead or amorphous stretches of language imposed upon, or freely chosen, for politeness or other socially justified reason. FS could also be more than ready-made lexical short-cuts to refer to stereotypical objects or actions or any other evidence of 'lazy' language behaviour. Its supportive and facilitating role in language production was thought to be possibly

illustrated by the study of its forms and uses in L2 classes, a language context which, to our knowledge, has not been the most widely explored so far.

Resting on the analysis of young learners' productions as it does, the study aims at getting an insight into the processes at work in the production and acquisition of what is called interlanguage. The term "interlanguage" was coined by Selinker (1972) to refer to the systematic knowledge of a second language which is independent from the learner's first language and the target language. Selinker claims that learners construct a series of interlanguages and revise them in systematic and predictable ways as they move along an interlanguage continuum. Other terms, such as "*approximate system*" (Nemser 1971) and "*transitional competence*" (Corder 1967), refer to the same basic idea.

The term "interlanguage" will be used here in a broader sense, to refer to the learner's use of the second language at his or her stage of acquisition, as opposed to the native speaker's use of his or her mother tongue.

The advantage of this terminology is that it does not refer to the learner's language as a "sub-language", 'inferior', so to speak, to an ideal target language, but it also implies the notion of 'developmental stages' in acquisition, which is one of the working hypotheses for the study.

In fact, it is hoped that the study will contribute to a definition of FS in general. More specifically, we hope it will help to identify the role of FS in early interlanguage use. Another useful issue, not addressed specifically here, would be to know if L2 learners make the same use of FS in the first stages of their language acquisition as native speakers do in their everyday usage. The classroom environment probably also plays a role in the use of FS by schoolchildren.

An even more ambitious goal, which this paper will attempt to keep in its sights if not reach fully, is to demonstrate that FS is a multi-faceted learning strategy in interlanguage, and more particularly that some of its functions play a part in partial immersion interlanguage use.

In so doing we keep in mind the two basic categories of FS which young learners might resort to: The first includes the "conversational routines" (Aijmer 1998) or "situation-bound utterances" (Kecskes 2000) which are context- and/or situation bound and the second covers FS such as idioms, collocations, etc. which are not situation-bound.

There are two main directions for the study: The nature of FS is looked into through brief definitions of FS at syntactic, morphological, phonological, semantic and pragmatic levels, we then consider the difference between FS and creative speech, and discuss the relationship between them. At the core of this discussion is the question: How can the variability or fixedness of a sequence be measured? FS is then considered within the broader framework of language organization and language development.

1.2. Observation and data collection procedures

The present study is based on the observation and recording of oral communication in a class of fifteen children aged five to seven, learning English in a partial immersion program (some subjects of the curriculum are taught in English every afternoon) in a small primary school in Nantes, France. As most of these children are at an early stage of L2 acquisition, their morphosyntactic development was thought to be worth examining.

It seemed interesting to observe a group of young learners together in an immersion class, because this provides broader information than a single individual's case study, and because the children's productions in the immersion classroom were expected to approximate naturalistic second language production. Usually, total immersion consists in giving education (i.e. the whole regular national curriculum) in a foreign language. In the present case, the children learn math and science in French in the morning and have other activities in English with a twenty-five-year-old Welsh teacher in the afternoon (it is worth noting however that the Welsh-accented input did not influence the learners's output as far as FS goes). The teacher resorts to a method based on short videos and interactive role-play games. There were fifteen children aged five to seven at the beginning of the observation. All of them were French, with French-speaking parents, except for a Rumanian boy, whose mother tongue was Rumanian, who had already started learning English in an international school and was currently learning French.

The recording sessions turned out to be of variable length since they depended on holidays, extra-curricular activities, teacher's sickness, and even strikes. The recordings covered a period of two months, between the end of November 2000 until the end of January 2001.

We had to base our reflection on other researchers' data bearing on conventional L2 learning, because it turned out that the communicative situation in the classroom was closer to regular foreign language teaching in primary schools than actual immersion teaching, in the sense that it did not resemble naturalistic language acquisition as much as expected. English was introduced to the children mainly as a subject to be learned rather than as a medium through which other subjects are learned, nevertheless, the observation did provide a few insights into the various functions of FS in second language production.

As far as data collection itself is concerned, we proceeded as follows: We only transcribed sequences in which English was used and numbered them from (1) to (396) in the original corpus, the boundaries between sequences corresponded to a minimal exchange of information¹. What is meant by 'minimal exchange' is an interactive situation in which a piece of information provided by one or several speakers is complete. So each sequence corresponds to a new piece of information.

Most of the time, the minimal exchange consists in a question-answer pair, or the interactive building up of a single sentence. All exchanges in French were removed, except when they revealed transfer phenomena, and all comments from the teacher that did not involve a response from the children were also suppressed, because the focus was to be on the children's speech only. The formulaic utterances selected according to criteria described in part IV of this paper are printed in bold. Sometimes there were several formulaic chunks in one sequence, so these were counted several times (see Table 1).

2. Identification and classification of formulaic speech

Researchers are often vague in their definition of FS and in the choice of the terms they use. So before considering FS in language use, it seems important to answer two basic but

¹ For reasons of space, the full corpus is not given here but is available from the authors.

essential questions: What is FS and how can the difference be made between formulaic and non-formulaic speech? These questions may be answered by looking into the nature of FS at the structural and lexical levels using examples from first and second language acquisition production.

We will first carry out a brief review of the classifications of what we will later call *formulaic speech* or *formulaic sequences* which have been proposed by researchers, then we will look into the structuring components and vocabulary of formulaic speech to reach the conclusion that formulaic speech cannot only be defined in structural or lexical terms.

Researchers in the field of second language acquisition have observed the use of formulaic utterances in the early stages of acquisition (Bohn 1986; Clark 1974; Hakuta 1974; Krashen & Scarcella 1978) but very few have described or defined the phenomenon precisely. The classification of formulaic utterances is rather difficult to establish since there is no generally agreed upon definition.

Researchers usually differentiate formulaic expressions according to their degree of variability. Various terms have been used to describe them, but they all refer to the same phenomenon: For example, while Hakuta (1974) distinguishes “*routines*” and “*prefabricated patterns*”, Nattinger and De Carrico (1992) consider all formulaic expressions to be “*lexical phrases*”, and classifies them in terms of their various functions in the organization of discourse. Many researchers follow Hakuta (1974) when he distinguishes “*prefabricated routines*”, which are memorized whole utterances or phrases that may be used without any knowledge at all of their internal structure (Krashen and Scarcella 1978: 283), and “*prefabricated patterns*” which are partly creative and partly memorized wholes: They consist in sentence frames with an open slot for a word or phrase (Krashen and Scarcella 1978: 283). Clark (1974: 3) even coined the term “*unopened packages*” to suggest that FS was like a package that had to be opened for the learner to analyse its content.

In the same manner, Bohn (1986: 188) uses the term “*formula*”, with the meaning of “magic formula” that helps convey an idea when you do not have the linguistic means to do it, for “*expressions in which no part is substitutable*”, and the term “*frame*”, with the meaning of “functional frame” that helps to place the lexical elements that convey meaning, for “*expressions which contain slots for more or less extensive paradigms of lexical elements*”.

What Bohn calls FS is made up of both frames and formulas, along with the additional category of amalgams, which are polymorphemic strings that are treated monomorphemically. Other researchers make a distinction in terms of the formulaic utterances’ semantic or pragmatic role: “*composites*” acquire a new referential meaning as the sum of their parts (e.g. *a close shave*), and “*formulae*” acquire a meaning connected with their role in discourse (e.g. *good morning*) (Cowie 1994; in Baigent 1996: 5).

Nattinger and DeCarrico (1992) classify “*lexical phrases*” according to their length, continuity and variability, and Raupach (1984: 123) makes a distinction between “*fillers and modifiers*” which give time for speech planning, and “*organizers*” that help planning.

In our study, we shall base the analysis of our data on the various criteria defined by other researchers on formulaic speech in their attempts to identify and classify formulaic sequences.

There seems to be two ways of classifying FS: Either by its function or by its degree of variability, but the “fixedness” or “variability” of an expression is difficult to determine

precisely. In the rest of this study, we will focus on the degree of variability of formulaic sequences, keeping in mind that this degree of variability is interdependent with the function of formulaic sequences, as we shall demonstrate further.

A possible entry door for the definition of the nature of FS may be offered by corpus linguistics which pays particular attention to collocations, that is, groupings or sequences of words that appear more often than others, and which Nattinger and De Carrico (1992: 20) describe as consisting of:

(...) A “node” that co-occurs with a “span” of words on either side. The span consists of particular word classes filled by specific lexical items. If it is the case that the node word occurs with a span of particular words at a frequency greater than chance would predict, then the result is a collocation. The more certain the words in the span are to co-occur with the node, the more fixed and idiomatic the collocation.

The existence of collocations shows that there are strings of words which are somehow fixed, and which seem to correspond to FS.

Another entry door may be neurolinguistics which has shown that some expressions are still used by aphasics and patients who have undergone a left hemispherectomy. The use of these expressions is generally referred to as *automatic speech* (Krashen and Scarcella 1978: 284-6). For these authors there appears to be rather reliable evidence for the existence of a formulaic type of speech. However, it is also true that the right hemisphere is able to take over a number of other functions when the left hemisphere is damaged, and that the roles of the right and left hemispheres do not seem to be very clear-cut. For the time being and for our present purpose, let us simply consider that FS might be a type of automatic speech.

There seems to be evidence from other various fields of study for a formulaic type of speech. The nature of this FS can also be studied at the structural and lexical levels in second language use.

2.1. The structuring components of Formulaic Speech: Syntax, phonology, morphology

At the syntactic level, formulaic utterances may not be, as Lyons (1968) suggests, “*unstructured or only partially structured*”. In fact, they are structured, but are actually learned and used as chunks (Skehan 1998: 53). For Klein (1986: 77), formulaic utterances are “*complex patterns of speech perceived as an entity*”.

A common criterion for determining the formulaic nature of an utterance in second language use is its well-formedness compared with the rest of the learner’s speech. For example, Homer, the Iranian 5-year-old boy observed by Wagner-Gough (1975: 71) could say “*My name is Homer*”, in one breath and “*He Fred*” in another, the former being a memorized pattern and the latter the learner’s own rule” (Krashen & Scarcella 1978: 292).

Formulaic utterances are also usually well-formed at the phonological level, but sometimes some of the phonemes are simply dropped or simplified.

Radford (1997: 124) mentions the case of a child’s use of the formula *Whasat?* (i.e. “*What’s that*”, an essential component of the child’s repertoire in picture-book situations), in a number of “*variant pronunciations, many of which are so contracted that they show no trace whatever of the wh-word “what”*”:

/wAsAt/ /wAsA/ /AsAt/ /AsA/ /sAt/ /sA/ /At/ /A/”

Sometimes the sequence is reduced to a single syllable as evidenced in our corpus:

(1) LOUIS: [Is] (it's) six o'clock.

(2) AGATHE: [ez] (there's) a car, car.

In other cases, the pronunciation is slightly simplified: There are no internal pauses and some diphthongs are reduced to a single sound, as shown in the following table based on analysis of our corpus:

Table 1: Phonological features of some formulaic sequences

FORMULAIC SEQUENCE	EXPECTED PRONUNCIATION	ACTUAL PRONUNCIATION
How are you	[haʊ aə ju]	[ha:ju]
What time is it?	[wɒt taɪm ɪzɪt]	[wɒtəmaɪzɪt]
How old are you?	[haʊ əʊld aə ju]	[haʊda:ju]
Are you_____?	[aə ju]	[a:ju]
I'm not_____.	[aɪm nɒt]	[a:nɒt]
Here you are.	[hɪə ju aə]	[hi:jʊa]
Brush our teeth	[brʌʃ əʊə ti:θ]	[brʌʃəti:θ]

These examples show that errors in FS are more likely to be of a phonological nature than of a syntactic nature. These structures may be variable and unstable in their phonological form (Radford 1997: 124), but most of the time they are recognizable by their well-formedness compared with more creative constructions.

Nattinger and De Carrico (1992: 26) even suggest that formulaicity also exists at the phonological level:

In acquiring phonology, children may use sequences far in advance of normal development; for example, they may use particular CCV syllables before they use expected CV or reduplicated CVCV syllables. Cruttenden (1981) found children using advanced consonant cluster forms like [sz:] *shoes*, and [dridθ] *do it*, before they had acquired the “earlier” CV forms. Such CCV clusters should be considered phonological idioms comparable to syntactic idioms and, like them, learnt as unanalysed units. Later on, children re-analyse these phonological chunks and learn to segment and reconstruct them in accordance with the phonemic principle.

Formulaic elements appear at the morphological level too, in the form of amalgams, defined by Bohn (1986: 188) as “*seemingly polymorphemic strings which are treated monomorphemically*”. For example, in our corpus, Luca assimilates the two morphemes *do* and *I* to the single pronoun *I*:

(3) LUCA: But **I'm** do like that

- (4) T: No, it's not time for sports. Luca, are you standing up?
 LUCA: No, **I'm sit down**, like this!#

In the same manner, he uses the two morphemes *it's* for the single pronoun *it*:

- (5) LUCA: **It's** can be hand clock [], sand clock [], can be, **it's**
 can, can be and some sand clock [], you know sand clock

The use of such forms corresponds to the first stage of morphological development. Early research on developmental stages (Ervin-Tripp 1973; Dulay & Burt 1975) has shown that learners first tend to use contracted negative forms such as *don't* or irregular verbal forms like *went* without recognizing their morpheme constituents, and then they analyze them, which sometimes results in ill-formed productions such as *goed* (a typical form of 'overgeneralization' in interlanguage terminology), and finally learners start using the correct form. Research (Dulay & Burt 1975) also shows that, although there is variability between learners, there are similar developmental patterns for all learners. Further research might find out a general, universal pattern of development. In that sense, amalgams belong to the first stage of morphological development.

2.2. The vocabulary of Formulaic Speech

The issue has already been given extensive coverage (Arnaud & Bejoint 1992; Aijmer 1998). Within the connectionist theory of multiple lexical storage, the lexicon is thought of as being composed of both single words and fixed phrases (continuous or discontinuous) (Taylor 1989). In fact, FS seems to be stored in the lexicon just like single words (Kiefer 1985, 1996). At the semantic level, this results in formulaic expressions being generally "rounded off" (Klein 1986: 77), which means that they have lost the referential meaning of their constituents. Therefore, learners sometimes misuse FS because they do not fully grasp its meaning.

- (6) T: How old are you?
 LOUIS: **How old are you?**

In this example, Louis only repeats the formula without fully grasping its meaning, since the teacher expected him to answer by giving his age.

- (7) T: The bathroom. What is he doing, the little boy?
 AGATHE: **Brush your teeth**
 MATHIEU: **Brush your teeth**

In this other example, the children use the formulaic "*brush your teeth*" without realizing that they are using a pronoun that doesn't correspond to the question.

The next sequence is interesting since it shows a progressive degradation of the formulaic arrangement, corresponding to a loss of meaning.

- (8) DAPHNE: Can I listen to clock? Can I listen to [] clock? Can I li-?
 Can I li- Est-ce que... comment on dit j'peux écouter ?
 (Does it... how do you say 'I can listen')

The fact that Daphne “cuts” the sequence at random places (even in the middle of the verb “*listen*”) might be an indication of her not having full knowledge of the morphosyntactic structure of the sentence.

- (9) ALINE: Agathe, **do you like chocolate cake?**
 AGATHE: Emm **No I'm not**

- (10) T: Ah, six and a half. Are you six and a half? Yes, yes I am, yes, six and a half. Aline, are you four?
 ALINE: **No, it is not.**

Here, the pronouns used in the short answers do not correspond to the questions which shows that Agathe and Aline did not grasp the meaning of the pronouns within the formulaic passages. Obviously, these answers only meant “*no*” for them and did not involve a person.

Sometimes, FS is not retrieved from the lexicon but from previous speech. In that case it is easily recognizable since it just copies the previous utterance.

Clark (1974: 3) gives a very good example of this phenomenon:

- Mother: We're all very mucky.
 Child: I all very mucky too.

At the pragmatic level, FS embodies a societal knowledge shared by a community (Ellis 1994: 85). In that sense, it occurs in socially-defined situations, and there is a correspondence between the situation and the FS used (cf. Coulmas 1979, 1981; Kecskes 2000). In the corpus analysed here, most FS samples are shared by all the class (though some of the children seem to misuse them more often than others). So FS is likely to reflect some kind of **shared classroom knowledge**. This element is certainly characteristic of our corpus and language observation environment. For example, the phrase “*chocolate cake*” has become formulaic because it is used as a “private joke” within the class, so that any mention of “*cake*” or “*chocolate*” or “*food*” triggers the use of this phrase.

- (11) T: Look, it's a birthday party, there's a birthday cake, how many candles can you see on the cake?
 C1: **#Chocolate cake!#**
 C2: **#Four!# Four!**

Other samples of FS directly dependent on shared classroom knowledge can be identified in our corpus: The recurrent teacher's phrase: “*Can I have?*” (*orange juice/bananas/apples, etc*) to elicit answers bearing on countable/uncountable items acquires its formulaic character by dint of being repeated and associated with different teaching contexts and lessons, in different time periods. Likewise “*Can I?*” is associated

with "go to ..." as well as a single verb (e.g. "drink"), its ubiquitous character makes the pupils use it alone, without combining it with a complement as in "Can I have some chips, can I" where the second *can I* is in no way a question tag or confirmation question. In other words, FS depends on the pragmatic context, the ongoing activity and the characteristics of the conversational partner (Rescorla and Okuda 1987: 283). These authors (1987: 287) noticed several differences in the use of FS by their subject, a five-year-old Japanese girl, depending on whether the conversation was held with adults or with peers. They noticed that the child talked more and used more repetitions with adults than with her peers.

Being essentially a communicative device, FS is most often pragmatically appropriate. There are cases however when FS appears in non-appropriate contexts more as a result of a reflex attitude than a deliberate strategy. For example, when the teacher mentions a measurement of quantity (*half*) for a recipe, it triggers the child's use of the FS phrase used to tell time (*half past*):

- (12) T: So I have two packets of eight hundred, plus half a packet, la moitié de la moitié# (half the half)
MATHIEU: #**Half past**#

This "triggering effect" seems to be due to the fact that the children are encouraged to use formulaic expressions by the teacher's prompts.

Nattinger and De Carrico (1992: 19) suggest that formulas are form/function composites. A formulaic form is learned in terms of its function. In FS, function has precedence over form, therefore it is more relevant to consider formulas from a pragmatic point of view rather than from a syntactic one.

Ill-formedness in FS then seems to be due to phonological simplifications or semantic and pragmatic misunderstandings which influence the form of the FS phrase. In other words, there are no structural errors in FS. This is in keeping with the definition that Wray (2000) gives of the formulaic sequence: "A *sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: That is, stored and retrieved whole for memory at the time of use, rather than being subject to generation or analysis by the language grammar*".

However useful an analysis of the structural components of formulaic speech and its vocabulary can be, it fails to fully grasp the nature and function of FS. The latter is at the moment being investigated more specifically by the same authors and will soon materialize in print. The former however can certainly be further defined here by looking at how FS can be differentiated from creative speech

3. Formulaic speech and creative speech: A few differentiating criteria

Most experts (Hakuta 1974; Rescorla and Okuda 1987; Bygate 1988; Miller & Weinert 1998), agree on the fact that formulaic utterances are "*unanalyzed wholes*". But the most difficult thing is to detect whether a sentence or a phrase is "*unanalyzed*". Raupach (1984) offers some psycholinguistic insights into how to solve this problem: He found that formulaic utterances are most often delimited by pauses or hesitation phenomena, and usually lack internal pauses. Following this, he managed to select utterances that were then

processed automatically in his data. But the problem is that creative sentences are delimited by pauses too.

There is another type of linguistic phenomenon that helps identify the boundaries of formulaic utterances in our observation: Code-switching generally appears around formulaic utterances, as when Lisa (who has memorized the sequence “*Can I have a ___*” instead of “*Can I have ___*”) says:

- (13) LISA: **Can I have a__** comment on dit ça (how do you say that)

In actual fact, children happen to fill in the “slots” within FS with words transferred from their native language:

- (14) ALINE: **I like** yahourt
T: Yoghurt
ALINE: **Yoghurt**

In the following example, the transfer phenomenon in the “slot” of the sequence “*Do you like ___?*”, with the use of the word “*pig*” to designate meat, is worth mentioning, since it is similar to the code-switching phenomena described above.

- (15) AGATHE: **Do you like, do you like** pig?
T: Oh no#
MATHIEU: #yeah!#

Radford (1997: 131-2) argues that a formulaic utterance is repeated monotonously several times to determine its formulaicity. Even if that occurs only occasionally, it appears to be a very useful element in our data. Indeed, some formulas are usually repeated in sing-song manner:

- (16) LISA: **Chocolate cake, chocolate cake**
CLASS: **Chocolate cake, chocolate cake, chocolate cake, chocolate cake, chocolate cake**, Catherine

- (17) AXEL: Nooo emm yes, yes **yes I am, yes I am, yes I am, yes I am**

Klein (1986: 77) suggests that formulaic utterances are recognizable by the fact that they are fairly short, but this is also very difficult to define, since the notion of length is relative. In fact, some very long sequences can be considered as formulaic, like poems or songs learned by heart.

Another criterion for recognizing formulaic utterances is their frequency. However, since formulas are most of the time context-bound, some are more frequent in certain contexts than others. The wide currency of a given formulaic utterance does not necessarily warrant its appropriacy in a given type of discourse, in other words “frequency” does not translate as “acceptability” in all cases, especially if the formulaic expression used in speech is borrowed from a non-speech, or non-conversational context, indeed, as Wray (2000) notes, “*there are undoubtedly some formulaic sequences that are widely accepted as such*

by native speakers but which are actually not very frequent in normal discourse”, such as expressions like “*God save the Queen*”. And the opposite is also true: Wray (2000) again mentions the fact that:

Another difficulty with using frequency as a means of spotting formulaicity is that it forces us to assume that any sequence of words that is repeated a few times is formulaic, that is, that we will not generate the same sentence from scratch very often without then keeping a copy whole for later use.

Wong-Fillmore (1976) also noticed that formulaic utterances are community-wide, so they are frequent for all learners of the same community. Lastly, a most agreed-upon feature of formulaic utterances is their well-formedness compared with more creative speech. For example, the use of negation in Mathieu’s speech illustrates this fact:

- (18) T: It’s time to play let’s say football
MATHIEU: No, no **play football!**

This sequence shows that his developmental stage for negation is of the type *no+VP*, which is the earliest stage for the development of the negative pattern. However, at the same time, he is able to produce:

- (19) MATHIEU: **I don’t like** this pizza

This example proves that Mathieu is using a frame to express dislike. If he had created his sentence from scratch, he would probably have said “*No like (this) pizza*”. However, here, he has the formulaic tool to express dislike, so he directly retrieves it from his lexicon and completes it with the new information he wants to convey.

With regard to these elements, it is now possible to draw up a tentative list of the factors that allow the definition, identification and classification of FS (adapted from Raupach 1984:122):

- High frequency
- Repetition (either monotonous repetition of the formulaic utterance or repetition of the previous utterance)
- Relatively short sentence length
- Syntactic correctness
- Influence of the “context” (topic, language task, etc.)
- Lack of variability
- Community-wide use
- Delimitation by pauses or hesitation phenomena
- Lack of internal pauses.
- Delimitation by code-switching phenomena

To conclude this list, it seems that utterances can be considered as formulaic not by their nature, but rather by the way they are perceived and by the use that is made of them. In fact, the factors in this list should only be considered as clues to recognizing FS. Perhaps a sequence should be considered formulaic if it exhibits more than one of these factors at the same time, but once again, this is difficult to establish precisely.

The point at issue here is that FS appears to be determined at the processing level, not at the spell-out level. In a way, formulaicity is a choice. Although most of the evidence gathered in the present study points to formulaic speech being used when the creative route to a given structure is not available, it is also possible to consider that a learner can choose to use the same sequence holistically or analytically, that is, to resort to FS or not, depending on his or her needs. However, it seems more likely that there is a “compulsory stage” for the use of FS and that the acquisition process is based on an evolution towards a freer use of FS.

In the following chart, we selected 186 sequences that we classified as formulaic, according to their frequency (they all occur at least four times within the transcription).

Table 2: Features of selected formulaic sequences.

FORMULA	Selected sequences	Repetition of previous sequence	Repeated monotonously	Syntactically correct	Phonologically correct	Appropriate	Delimited by hesitation	Lack of internal pause	Delimited by code-switching
<i>Thank you</i>	4	0	0	4	4	4	0	4	0
<i>Good afternoon</i>	6	4	0	4	4	4	0	4	0
<i>How old are you?</i>	7	2	0	5	6	5	3	6	0
<i>How are you?</i>	8	5	1	8	7	7	1	7	0
<i>What's your name?</i>	4	1	0	4	4	3	0	4	0
<i>Can I go+(to)VP?</i>	9	1	1	6	6	9	0	9	0
<i>Can I have a +N?</i>	5	0	0	5	5	5	1	5	1
<i>Can I have +NP?</i>	13	3	3	10	10	13	5	9	1
<i>Can I (+VP)?</i>	24	20	10	21	24	20	1	24	2
<i>What time is it?</i>	7	3	2	7	7	5	2	6	1
<i>(It's) ___ o'clock.</i>	20	2	2	17	20	20	2	19	3
<i>(It's) half past ___.</i>	15	5	2	11	12	13	4	10	1
<i>No, I'm not</i>	4	2	0	4	3	3	1	4	0
<i>Yes, I am.</i>	4	0	1	4	4	4	1	4	0
<i>Do you like+NP</i>	5	1	3	5	5	4	4	4	1
<i>I like+NP</i>	17	1	3	17	17	13	5	17	1
<i>I don't like+NP</i>	7	0	1	7	7	7	0	6	1
<i>Brush (...) teeth</i>	5	1	0	5	4	4	0	1	0
<i>Chocolate cake</i>	14	3	4	14	14	12	3	14	1
<i>Don't touch it</i>	8	0	0	8	7	8	0	0	0
Total	186	54	33	166	170	163	33	157	13
Percentage		29%	18%	89%	91%	88%	18%	84%	7%

What this chart shows is that a majority of the selected sequences include various

components of formulaic utterances. Applied to our data, Raupach's list of criteria (1984) enables us to distinguish four main characteristics of nearly equal importance:

1) they are phonologically correct, this may result from the fact that both teachers and pupils repeated those short (3-4 words) phrases a high number of times, correcting and cross-correcting one another so often that the formulas acquired the nearly musical ring of familiar radio jingles.

2) the formulas are syntactically correct, probably mostly because they are both learned and used as unanalysed chunks (Skehan 1998). This may seem paradoxical, but the lesson is perhaps worth remembering for the teaching of grammar in general within the big communicative vs analytic debate.

3) They are appropriate, this may seem the most unexpected characteristic, but the relatively limited context of utterance in which they are used (typically within the narrow limits of an adjacency pair) added to the teacher's prompts which act as triggers and context pointers, may account for this relative high rate of appropriacy.

4) they lack internal pauses, an aspect which reinforces the "learned-as-whole-chunks" characteristic of the production of formulas, but point both to the relatively unanalysed element of formulas as pragmatic devices used spontaneously and their role as near punctuation markers, a crucial aspect identified by Raupach (1984) to measure the intervals between sequences precisely.

In this particular case, the hesitation phenomena delimiting formulaic utterances do not seem a proper criterion for determining the formulaic status of an utterance, since hesitation phenomena can also occur before or after creative strings of words. The repetition factor is less predominant and occurs occasionally but must not be disregarded as illustrated by the "*can I...?*" example mentioned above.

Code-switching phenomena, which seem to appear at the periphery of the most "flexible" sequences, do not appear very often. However, code-switching is naturally less frequent than other phenomena such as hesitations, so maybe a reasonable conclusion would be to state that it is one of the factors that helps identify formulas, but neither the most important nor the most reliable one.

Formulas are generally said to be easily recognizable by their well-formedness, their frequency, and the fact that they are community-wide. Our study bearing on young L2 learners introduces a few variations in this identification process, but confirms the community-wide characteristic which we called "shared classroom knowledge".

The interface between creative and FS is not clear-cut: Some sequences such as *Can I+VP* might be considered as half-formulaic, half-creative. Wong-Fillmore (1976) was criticized by Bohn (1986: 192) because she pushed the concept of formulaicity so far it became synonymous with syntactic structure.

4. Formulaic speech and the analyzability of speech

The difference between formulaic and creative speech is in fact very vague, and FS itself seems to have various degrees of formulaicity. **Three elements may play a part in the formulaicity of an utterance: Linguistic fixedness, communicative fixedness and socio-cultural fixedness.**

For example, while frames are as linguistically fixed as formulas, they are more

flexible at the communicative and socio-cultural levels. In the same manner, idioms (i.e. “a string of words which has an idiosyncratic meaning”, Radford 1997: 510) differ from formulas in that their socio-pragmatic function is less flexible, but the boundary between those “categories” is very thin. How can the variability of an utterance be measured?

Moreover, the fact that some sequences of speech are not analyzed by learners can simply be accounted for by the fact that they are not analyzable. Many idioms belong to that category, since their meaning cannot be inferred from the referential meaning of their constituents. Some of them are even ungrammatical, like “by and large”. In the same manner, as Wray (1999) suggests, “it is necessary to allow for the possibility that word sequences may be formulaic even though they do not need to be, that is, even though they are semantically transparent and syntactically regular”. Then, one might wonder how speech becomes formulaic. It seems that the origins of FS are related to their functions, since they seem to appear when needed in routine situations related to community identity or discourse organization. In all cases, FS might be the result of the high-frequency input of a particular sequence. But other factors are likely to be important, such as, in some cases, transfer from the first language (e.g. between relatively close languages like French and English: Thee two structurally parallel constructions: “*Quelle heure est-il?* / *What time is it*”, used formulaically by the teacher) or the learner’s communicative need to perform the language functions encoded in FS with ease and fluency (Ellis 1994: 272). Some sequences learned analytically may also become formulaic in a process called *fusion*, according to Peters (1983), probably for reasons of processing economy. The “genesis” of FS has not been explored much and certainly needs to be further investigated.

Therefore, FS can be seen more as a production strategy than a real linguistic form. In that sense, repetitions can be considered as formulaic as long as they correspond to a production strategy, that is, when they have a clear function in the speech structure.

Based on the different types of fixedness we identified previously, we defined a continuum ranging from the most independent structures of creative speech to structures that are tightly bound to context (both intralinguistic and extralinguistic). There are various degrees of formulaicity rather than clear-cut categories of formulaic utterances: FS is not a discrete category.

Figure 1: The creative-formulaic continuum.

The question of the nature of FS, be it at the structural or lexical level, brings up a series of much broader questions concerning the neurolinguistic aspects of language, lexical storage or language structure and organization.

5. Formulaic speech in language organization and development

Formulaic utterances are parts of language which are not fully processed at the structural level, “rounded off” semantically and best defined in terms of their pragmatic function. They are situated in a continuum from a mostly flexible to a mostly fixed type of language. How is that reflected in the whole language organization? What type of relationship is there between FS and creative speech? We shall attempt to answer these questions by focusing on the notions of competence and performance. The idea will then be pursued a bit further by questioning the organization of language with regard to language development.

5.1. Performance versus competence

Radford (1998: 251), like most researchers working within the Chomskyan theory of Transformational Grammar, considers FS as a sort of “anomaly” in children’s speech, because it is not representative of their competence. Here is the definition he gives of performance as opposed to competence:

Performance is a term which denotes observed language behavior, e.g. the kind of things people actually say when they speak a language, and what meanings they assign to sentences produced by themselves or other people. Performance can be impaired by factors such as tiredness, drunkenness, etc. Performance is contrasted with competence which denotes the fluent native speaker’s knowledge of the grammar of their native language.

Nevertheless, it is worth noting that Chomsky himself has maintained that grammar rules are not psychologically real, and “*just because a sentence can be explained by the application of a particular linguistic rule does not mean that the speaker has applied it each time.*” (Larsen-Freeman et al. 1991: 69).

Widdowson (1979; cited in Nattinger and De Carrico 1992: 3) suggests that “*knowledge of a language does not mean only knowledge of the rules which will generate an infinite number of sequences, but a knowledge of the rules which regulate the use of sentences for making appropriate utterances.*”

Most of the time, linguistic phenomena of a pragmatic kind are attributed to performance, not competence, and the Chomskyan model, by creating “*a dichotomy [between performance and competence] which is unnecessarily severe and exclusive*” (Nattinger & De Carrico 1992: 2), lacks a description of pragmatic competence.

For example, Error Analysis involves a set of procedures for identifying, describing and explaining errors in learner language. An “error” is a deviation in learner language which results from a lack of knowledge of the correct rule. It contrasts with a mistake, which is a deviation in learner language that occurs when learners fail to perform at the level of their competence, a lapse that reflects processing problems (Ellis 1994: 701). The

problem with this type of approach is that it is based on the opposition of performance (mistakes) and competence (errors). As mentioned previously, errors within FS do not reflect the same type of competence as errors within creative speech, but they cannot be given the status of a mistake either.

So competence has many facets, ranging from the grammatical dimension of the Chomskyan model to a more pragmatic type of knowledge. Thus, although FS seems to belong to performance, it does reflect some kind of competence, which is neither of a syntactic kind, nor of a purely pragmatic kind, but of a pragmalinguistic one. This “*pragmalinguistic competence*” is the interface between grammar and the more general principles of language use (Nattinger and De Carrico 1992: 11). It is also referred to as “*communicative competence*”, defined as such by Ellis (1994: 696):

Communicative competence consists of the knowledge that users of a language have internalised to enable them to understand and produce messages in the language. Various models of communicative competence have been proposed, but most of them recognize that it entails both linguistic competence (e.g. knowledge of grammatical rules) and pragmatic competence (e.g. knowledge of what constitutes appropriate linguistic behaviour in a particular situation).

Nattinger and De Carrico (1992: 16) represent the interface between linguistic and pragmatic competence required for the learner to use lexical phrases in the following chart. The solid lines indicate processes involved in grammatical competence, and the dotted lines those involved in pragmatic competence:

FS is thus a proof that there is interaction between the linguistic and pragmatic aspects of competence, and, as Bachman (1990) suggests, the term “communicative competence” should be used to refer to that interface.

5.2. Language organization and language development

Creative speech and FS are often sharply opposed. For example, Krashen and Scarcella (1978: 285) provide neurological evidence that FS and creative speech are fundamentally different: “*Routines and patterns are often preserved in cases of non-fluent (syntactic) aphasia and after left hemispherectomy*”. Hatch (1983: 211) suggests that FS is governed by the right hemisphere (while most of language is governed by the left hemisphere), which would account for cases of

bilingual aphasia patients who, after left hemisphere damage, lose their first language but do remember tenth-grade, high-school French, using these high frequency utterances, parts of dialogues, and vocabulary items to carry out their communicative needs.

However, one might wonder how these English-speaking patients could handle their “*communicative needs*” using high-school French. Moreover, this theory was criticized by Nattinger and De Carrico (1992: 30), who argue that FS is of a different nature than automatic speech, and that generally automatic speech is pragmatically inappropriate, whereas most of the time FS is not. A possible solution to this problem might be to consider the possibility that two modes of language processing can co-exist.

Bolinger (1975; cited in Skehan 1998) offers the formulaic approach as a useful way of bypassing doing the rule-based processing of sentences when a word string is used many times, he accepts that much language use is repetitive and not particularly creative. In fact, it would seem plausible that a good deal of native speaker linguistic behavior is just as routine as the formulaic language of learners. Corpus linguistics has found initial evidence of lexical rather than grammatical organization of language with such phenomena as: “*The clustering of words related to areas of meaning, the delexicalisation of some very frequent and generally applied words (e.g. put), the role of lexical repetition across speaker boundaries in discourse and (...) certain patterns of collocation suggesting idiomatic sequences*” (Baigent 1996).

This conception of language is now widely accepted and no longer seen as a return to behaviorist views of language.

The idea of a lexical organization of language tallies with the notion of multiple lexical storage, which suggests that whole sentences could be memorized as single lexical units and then need only as much processing effort as a single word to produce. FS could then be seen as a short-cut route to speech production co-existing with more complex creative speech. The idea goes back as far as Bolinger (1975) and has been substantially developed by later writers.

Along that line, researchers like Raupach (1984) or Peters (1983) have suggested that there are two different modes of language organization, corresponding to two modes of learning that the speaker chooses to use according to his or her own cognitive mode and to the situation. These two modes could be summed up in the following chart (adapted from Raupach 1984):

Table 3: Poles of language organization and development

GRAMMATICAL ORGANIZATION OF LANGUAGE	LEXICAL ORGANIZATION OF LANGUAGE
Creative speech	Formulaic speech
Propositional speech	Automatic speech
Referential language	Expressive language
Word production	Prosodic arrangement
Analytic learning	Gestalt learning
System learning	Item learning

This chart is not meant to refer to a drastic opposition but rather to a continuum

from a mostly automatic to a mostly propositional use of language. Indeed, the opposition formulaic / non-formulaic is not clear-cut. For Rescorla and Okuda (1987), “*we should move beyond misleading dichotomies such as prefabricated formulas versus creative constructions in our attempt to describe second language learning*”. Indeed, there can be flexibility and variability within formulaic utterances, and “*learners seem to make use of patterns which are varied to a greater or lesser extent through lexical substitution*” (Ellis 1994). Rescorla and Okuda’s (1987) conception of a language made of a few structures or patterns that generate a large number of different English utterances seems plausible, and creative speech would then be made up of a set of formulaic modular elements assembled according to pragmatic and semantic rules.

This organization of language is reflected in the acquisition process. Peters (1976; in Krashen and Scarcella 1978: 288), opposes “*analytic*” and “*gestalt*” styles of foreign language development, which roughly corresponds to what Cruttenden (1981) calls “*system-learning*” and “*item-learning*”. The analytic style, which is used for referential, labeling functions, is the “one-word-at-a-time” development described in most studies of child language acquisition, while the gestalt style is the attempt to use whole utterances in an appropriate situation, and is thus used in more “*conversationally defined*” contexts. Although Peters (1983) studied first language acquisition by children, not second language as in our corpus, his findings suggest that there is individual variation among children as to which style will predominate. A given child may use one style for one situation and the other for a different situation. Moreover, children differ in the extent to which they employ FS: Some children develop their language using a word-based analytic approach, whereas other children seem to rely more upon a gestalt strategy based on the acquisition of FS (Rescorla and Okuda 1987: 283).

For Hatch (1972; in Krashen and Scarcella 1978: 290), these two modes of learning are independent and the analytic mode eventually predominates, with gestalt speech primarily serving only as a short-cut or pragmatic tool to allow social interaction with a minimum of linguistic competence.

Rescorla and Okuda (1987: 282) suggest that the use of the gestalt mode may be peculiar to second language acquisition, since it is not present in their subject’s first language acquisition, but their findings are based on a case study, and since there seem to be large individual differences, it is not sure that these modes of learning always vary from one type of acquisition to another. Further research needs to be carried out in this field.

6. Conclusion

FS challenges the usual conception of language acquisition as an analytical process, and the idea that language production is based on an analysis of the input followed by a production out of parsed output (Bygate 1998; Miller & Weinert 1998). FS is linked to many facets of language production, and is therefore likely to play a part in learner production at the linguistic, psycholinguistic and communicative levels.

It is important not to under-estimate the children’s knowledge of the language, and to be careful not to mistake creative structures for formulaic frames. It is unfortunate that the only conclusion that can be reached for now is that formulaicity, even if it can be identified by the few criteria previously mentioned, is mainly a pragmatic concept which

depends on the speaker's choice in processing language, and thus cannot be fully identifiable by its form.

Therefore, the problem is still open, since all those criteria could be applied to creative speech all the same. It seems that learners' errors in FS are different from those in creative speech, being of a more pragmatic kind. Nonetheless, the criteria can only be taken as occasional clues, not as actual proof. As a further paper will try to demonstrate, formulaicity is a way of processing language, and its manifestation at the level of form is very subtle and difficult to identify with certainty.

FS is situated in a continuum from a mostly flexible to a mostly fixed type of language, the "fixedness" of which can be found at linguistic, communicative and socio-cultural levels. This continuum corresponds to two poles of language development, one being analytic and the other "gestalt". So, as a real component of language use, FS is not only a temporary stage of acquisition (Pawley & Syder 1983; Aijmer 1998), but also a processing strategy that benefits both the speaker and the hearer. It benefits the speaker's production by giving him or her time for discourse planning or other types of processing effort, and it also facilitates processing by the hearer (Wray 2000), who processes a sequence more easily if it is formulaic than if it is more creative. Thus, FS helps the learner reach idiomaticity and thereby efficient communication in his or her second language with native speakers.

Since FS is at the interface between linguistics and pragmatics, it is an interesting subject to investigate in psycholinguistics and neurolinguistics too, to determine the processes at work behind the development of communicative competence. The present study is based on short-term observation, and all the problem areas could not be investigated in depth. Another field of study to be investigated may be the role of the input in the use of FS. Indeed, it seems that FS as used by learners is encouraged by highly formulaic input.

References

- Aijmer, K. (1998) *Conversational Routines in English*. London: Longman.
- Arnaud, P., & H. Bejoint (1992) *Vocabulary in Applied Linguistics*. London: MacMillan.
- Bachman, L.F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Baigent, M. (1996) *Speaking in Chunks: An investigation into the use of multi-word phrases in spoken language by advanced learners of English*. MSc Thesis in Teaching English, Aston University. <http://www.les.aston.ac.uk/lisu/diss/m.baigent.html>
- Bohn, O. (1986) Formulas, frame structures, and stereotypes in early syntactic development: Some new evidence from L2 acquisition. *Linguistics* 24: 185-202.
- Bolinger, D. (1975) Meaning and memory. *Forum Linguisticum* 1: 2-14.
- Bygate, M. (1988) Units of oral expression and language learning in small group interaction. *Applied Linguistics* 9.1: 59-82.

- Clark, R. (1974) Performing without competence. *Journal of Child Language* 1: 1-10.
- Corder, S.P. (1967) The significance of learners' errors. *International Review of Applied Linguistics* 5: 161-9.
- Coulmas, F. (1979) On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics* 3: 239-266.
- Coulmas, F. (1981) *Conversational routine: Explorations in standardized communicative situations and prepatterned speech*. The Hague: Mouton.
- Cowie, A. (ed.) (1994) *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press.
- Cruttenden, A. (1981) Item-learning and system-learning. *Journal of Psycholinguistic Research* 10: 79-88.
- Dulay, H., & M. Burt (1975) A new approach to discovering universals of child second language acquisition. In D. Dato (ed.), *Developmental psycholinguistics* (Monograph Series on Language and Linguistics), 1977. Washington, D.C.: Georgetown University Press, pp. 209-233.
- Ellis, R. (1994) *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Ervin-Tripp, S. (1973) Some strategies for the first two years. In A. Dil (ed.), *Language Acquisition and Communicative Choice*. Stanford: Stanford University Press.
- Hakuta, K. (1974) Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning* 24: 297-297.
- Hatch, E.M. (1972) Some studies in language learning. *UCLA Workpapers on Teaching English as a Second Language* 6: 29-36.
- Hatch, E.M. (1983) *Psycholinguistics. A second language perspective*. Rowley, Mass.: Newbury House.
- Kecskes, I. (2000) A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics* 32.6: 605-625.
- Kiefer, K. (1985) How to account for situational meaning? *Quaderni di Semantica* 2.85: 288-295.
- Kiefer, K. (1996) Bound utterances. *Language Sciences* 18.1-2: 575-587.
- Klein, W. (1986) *Second Language Acquisition*. Cambridge: Cambridge University Press.
- Krashen, S., & R. Scarcella (1978) On routines and patterns in second language acquisition and performance. *Language Learning* 28: 283-300.
- Larsen-Freeman, D., & M.H. Long (1991) *An introduction to second language acquisition research*. London: Longman.
- Lyons, J. (1968) *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Miller, J., & R. Weinert (1998) *Spontaneous Spoken Language. Syntax and Discourse*. Oxford: Oxford University Press.
- Nattinger, J., & J. De Carrico (1992) *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nemser, W. (1971) Approximate systems of foreign language learners. *International Review of Applied Linguistics* 9: 115-123.

- Pawley, A., & F.H. Syder (1983) Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J.C. Richards & R.W. Schmidt (eds.), *Language and Communication*. London: Longman, pp. 191-226.
- Peters, A.M.. (1976) Language learning strategies: Does the whole equal the sum of the parts? *Language* 53: 560-573.
- Peters, A.M. (1983) *The units of language acquisition*. Cambridge: Cambridge University Press.
- Radford, A. (1997) *Syntactic theory and the structure of English*. Cambridge: Cambridge University Press.
- Radford, A. (1998) *Syntactic Theory and the Acquisition of English Syntax*. Cambridge: Cambridge University Press.
- Raupach, M. (1984) *Formulae in Second Language Production*. In H. Dechert et al. (eds.), *Second Language Productions*. Tübingen: Narr.
- Rescoria, L., & S. Okuda (1987) Modular patterns in second language acquisition. *Applied Psycholinguistics* 8: 281-308.
- Selinker, L. (1972) Interlanguage. *Interlanguage Review of Applied Linguistics* 10: 209-231.
- Skehan, P. (1998) *A Cognitive Approach to Language learning*. Oxford: Oxford University Press.
- Taylor, J.R. (1989) *Lexical Categorization*. Oxford: Oxford University Press.
- Wagner-Gough, J. (1975) Comparative studies in second language learning. *CAL ERIC/CLL Series on Languages and Linguistics* 26.
- Weinert, R. (1995) The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16.2: 180-205.
- Widdowson (1979) *Explorations in Applied Linguistics*. Oxford: Oxford University Press.
- Wong-Fillmore, L. (1982) The second time around: Cognitive and social strategies in SLA. *Language Learning* 32: 53-68.
- Wray, A. (2000) Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics* 21.4: 463-489.
- Wray, A. (2002) *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.