Formulaicity in Code-Switching: Theory

Kazuhiko Namba
English Department

0 Introduction

This chapter examines the possibility that the code-switched units are formulaic sequences (a term defined presently). This idea comes from a paper by Backus (2003), where he proposes that code-switching always involves a 'unit' that is produced holistically. Backus' claim contrasts with another, by Azuma (1996), that code-switching entails complete syntactic constituents. These two proposals will be compared and evaluated, with particular attention to examples in their own work and in the present dataset, that can help separate out the different predictions that their positions make. In short, are there examples that can reasonably be defined as formulaic sequences but which cannot be viewed as complete syntactic constituents? (The reverse is less likely, since there is no reason why a syntactic constituent should not also be a formulaic sequence).

A major part of the present chapter will be taken up with resolving a procedural problem that would otherwise prevent this evaluation of Backus and Azuma being done: namely, how to identify a formulaic sequence. With that in mind, the chapter is organised as follows. First, the basic claims of Backus and Azuma are reviewed and compared. Examples that they offer are examined and discussed. Next, the definition of the formulaic sequence is considered, and a means of identifying formulaic sequences in text is presented and discussed. The means of identification is then tried out on the test examples from monolingual data and on examples from Azuma and Backus, so that an interim conclusion can be reached about the validity of the two positions. In chapter seven, the identification procedure is applied to the main dataset, and the outcomes are reviewed in order to establish whether the problematic cases can, indeed, be explained in terms of formulaic language theory. At the end of chapter

seven, some reflections are offered about the identification procedure itself.

1 Does code-switching entail a syntactic constituent or a formulaic sequence?

EL insertions are a valuable test case for theories of how language is processed. Wray's Needs Only Analysis model (2002:130) proposes that virtually any kind of wordstring, continuous or a frame with gaps, can be a single lexical unit if (a) it has a reliable meaning as it stands, and (b) the input experienced by the speaker, and his/her output needs, have not required it to be broken down further. Clearly, this approach predicts that a 'unit' need not be a recognisable syntactic constituent. In contrast, standard syntactic models would assume that if a 'unit' is to be taken from another language, and that unit is more than a single word, it will be a syntactic constituent. The two sides of the argument are taken up by Backus (2003) and Azuma (1996) respectively. Before examining their claims and evidence in detail, it is worth reflecting on where Myers-Scotton appears to stand on this issue. On the one hand she speaks of the EL island is a syntactic unit, saying that an EL island shows "structural dependency relationships that make them well-formed in the Embedded Language" (2006:261). On the other hand, she recognises in her data that generally EL islands are adjuncts, formulaic or idiomatic (2002:141), and that not all of them are obvious syntactic constituents. While adjuncts, as indeed other formulaic sequences certainly *can* be complete syntactic constituents, the key issue appears to be that Myers-Scotton has found instances where they are not.

Azuma hypothesizes that "bilinguals switch at syntactically definable constituent boundaries" (1996:397). In order to verify his hypothesis, he examined code-switching examples in spontaneous speech from the literature. For instance the following example shows insertion of an English EL noun phrase "all that fish" into a Japanese ML frame.

(1) All that fish *ga* *narandenno* *Yo*

NOM lying TAG

{All that fish are[1] lying there, you know}
(Nishimura, 1985 quoted in Azuma, 1996:402)

This is an example of an EL island as a syntactic constituent[2], which supports Azuma's hypothesis. According to Azuma, naturalistic data in the literature suggests that switching occurs at the boundary of grammatical constituents, however, such data is limited because the size of the corpus might not have been large enough to detect non-constituent switching and researchers might not have recognized such switching when gathering data (Azuma, 1996:403).

In order to gather further samples, Azuma conducted an experiment in which code-switching was elicited. Japanese-English bilinguals were asked to switch language in response to randomly generated tones. The following example shows that code-switching occurs at the end of a prepositional phrase even though the tone is heard in the first word of the phrase (The arrow indicates the tone).

↓

(2) *Ano moo* *kuji* *Kara* *Rokuji* *Made* un fixed pattern and takes uh

Well 9 o'clock From 6 o'clock Till

{ Well, from 9 o'clock to 6 o'clock (it's a) fixed pattern uh (he usually) takes uh...}

(Azuma, 1996:407)

Code-switching occurs after the prepositional phrase *kuji kara rokuji made* "from nine o'clock to six o'clock" which is a syntactic constituent. This example supports Azuma's hypothesis, however it should be noted that this experiment can not elicit insertional CS

[1] There is an error in this translation. It has to be 'All that fish is lying there, you know.'
[2] Azuma's rather broad categorization of syntactic constituets includes words, phrases and clauses (1996:399).

but alternational CS. The subjects were asked to change language when they heard the tone, which is alternational CS by definition. Therefore the result of the experiment should not be generalized to insertional CS.

In his elicitation experiment, the percentage of the time that the subjects continue their speech in the same language at least one word after the tone was quite high (68 %). Only 5% of the time was the material preceding the switch judged not to be a syntactic constituent (Azuma, 1996:406). One of such examples is as follows.

↓

(3)	Living in Austin is	*totemo*	*ii*	*keiken*	*ni*	*nari*	*-masu*
		very	good	experience	RSL	become	HON

{Living in Austin is a very good experience}

(Azuma, 1996:412)

If Azuma's hypothesis is right, the switching should have happened after "Living in Austin" which is a non-finite clause acting as a subject NP. The sequence "Living in Austin is" is 'NP+ verb' which is not a syntactic constituent. Azuma suggests (1996:412) that such wordstring might be a unit in the 'performance structure'[3] which can be identified by measuring the length of pauses, i.e. a longer pause signifies a boundary between more major units. This is not very helpful because the performance structure doesn't tell us the underlying principles but observable results. Alternatively, perhaps these non-syntactic constituents have some kind of psycholinguistic and sociolinguisic/discourse status of their own. Further exploration is needed.

Backus (2003) argues that inserted EL items are not always syntactic constituents but

[3] Proposed by Gee and Grojean (1983) cited in Azuma (1996 : 412).

'lexical units', which are "any recurrent combinations of two or more morphemes that together exhibit idiomatic meaning" (2003:90). He proposes the 'Unit Hypothesis', saying that "Every multimorphemic EL insertion is a unit, inserted into a ML clausal frame"(2003:91). Backus's argument appears to be operating in the opposite direction from Azuma's. Azuma can predict where CS will start and stop because he can define his unit independently of his data, using standard syntactic theory. On the other hand, Backus can't predict where CS occurs at all. When it has occurred, he will be able to tell you that the EL material must be a unit for the speaker. He lets the data tell you what to take notice of, and then tries to build a theory to explain it. Backus is in a weaker position than Azuma.

In order to test his hypothesis, Backus devises criteria (2003:91) in which 'non-compositionality' is "the most important diagnostic for lexical units"(Backus, 2003:120). The criterion of 'recurrence' can be used for detecting a not particularly idiomatic combination. Other diagnostics are 'irregular morphosyntax', 'frozen form' 'pragmatic function', and 'phonological reduction'. He examines the Unit Hypothesis using his data from Turkish-Dutch code-switching. The use of plural nouns, compound nouns, adjective-noun and verb-object collocations as inserted EL items are shown. One of the examples is as follows.

(4) *Op kamer-s wonen* yap-acağ-ım

on room-pl live. INF do-FUT-1sg

{I'm going to live on my own}

(Backus, 2003:108)

The inserted Dutch EL item *op kamer-s wonen* "live on my own" is a verb phrase consisting of the main verb and a prepositional phrase and at the same time a fixed collocation, i.e. a 'lexical unit'. The tested EL items in Backus' paper (2003) are basically

all syntactic constituents therefore the role of formulaicity is not independently verified.

Backus has initially devised the Unit Hypothesis to explain inserted EL multimorphemic items but he finds out that it can apply to many cases of alternational CS. He shows (2003:114) some examples of alternational CS patterns, such as the following example (5)

(5) kültürle, kültürle dini karıştırıyorlar, *vind ik*

{they are mixing up culture and religion *I think*}

(Backus, 2003:114)

Backus (2003:116) points out the Dutch clause *vind ik* "I think" has several features of lexical units. It is fairly non-compositional,[4] its form is frozen, it is a recurrent pattern in the corpus used for the study, phonological reduction is observed, and it has a pragmatic function[5]. Since a clause is another syntactic constituent, this example supports both Azuma's and Backus' arguments and cannot, therefore, verify formulaicity independently. Another subcategory of units observed in alternational CS is a 'construction' (2003:114), i.e. units which may have open slots, may be discontinuous and which are associated with a pragmatic meaning. Examples of Ductch constructions in alternatinal CS are *het is wel* ADJ *zo* "it's ADJ though" and *die is gewoon* ADJ/NP "she's just" (2003:118). These are not constituents from the syntactic point of view but appear to have the characteristics of lexical units. If the formulaicity of these examples from 'constructions' is verified, Backus's position will be strongly supported. Their formulaicity will be examined later in this chapter.

Although Azuma's paper focuses mainly on alternational CS whereas Backus' focuses

[4] Backus explains that "the deep entrenchment of that meaning preempts the emergence of an alternative compositional meaning" (2003:116).
[5] Backus explains that "it mitigates the preceding assertion" (2003:117).

on insertional CS, both papers cover materials from the other categories as well[6]. Therefore examples of each CS category can be drawn from both papers. Our agenda was to find EL items which are not syntactic constituents but formulaic sequences. As it happens, all the examples of insertional CS in both Azuma's and Backus' papers map onto syntactic constituents, so there is no scope there for testing Backus' hypothesis. However it turned out that in both papers, several non-syntactic constituents have been observed in alternational CS. The examples which Azuma explains in terms of the 'performance structure' and the ones which Backus explains in terms of 'constructions' in alternational CS are worth exploring. Backus

If we find non-syntactic constituents, the next agenda is to examine the formulaicity of the units. The inherent difficulty here lies in the identification of formulaic sequences, whereas that of syntactic constituents is more straightforward if we use already established syntactic theories. In order to identify formulaic sequences, we need to define them. In the next section, we will try to establish a definition of formulaic sequence.

2 Formulaicity in language processing

If we turn to monolinguals' language processing, the syntactic structure and formulaicity can be explained as two processing systems. Chomskians' view of language processing (e.g. Radford, 2004:5) entails an analytic approach in which morphemes and words are combined into phrases and sentences by grammatical rules for output while input is broken down into words and morphemes. Sinclair (1991:109) calls this analytic processing the 'open choice principle'. But Sinclair also identifies another approach to another processing, the 'idiom principle', in which "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to

[6] That is, insertional CS in Azuma's and alternational CS in Backus'.

be analysable into segments" (Sinclair, 1991:110). In essence, Sinclair's 'single choices' are the same as what Wray (2002) terms 'formulaic sequences'. Both Sinclair and Wray propose that the production of these units is holistic, and therefore bypasses the analytic process of construction that occurs with the open choice principle. Sinclair claims that the idiom principle is the default mode and as an exception "whenever there is a good reason" (p114), the speaker turns to the open-choice principle. Wray (2002:14) notes the advantages of having a dual language processing system. The advantage of the analytic system is "its flexibility for novel expression and the interpretation of novel and unexpected input" (2002:18). On the other hand the advantage of the holistic system, i.e. the idiom principle, is "the reduction of processing effort" (2002:18). She argues that for native speakers it is "the accessing of large prefabricated chunks, and not the formulation and analysis of novel strings, that predominates in normal language processing" (2002:101). In addition to this processing function, Wray identifies sociointeractional functions (2002:204) which explain why the forms persist, i.e. we could in theory reduce processing on a more ad hoc basis rather than coming back to the same formulations over and over. However for these social reasons, we share these formulaic forms across the speech community.

Wray (2002:9) sets up a definition as follows and reviews the literature comprehensively.[7]

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray, 2002:9).

This definition aims to be as inclusive as possible but 'is, or appears to be' can be a

[7] The conclusions Wray comes to as the result of the review are : multiword lexical units are morpheme-equivalent (2002:265); they arise on account of Needs Only Analysis (ie exist mostly because they have not been broken down, not because they have been fused) (2002:130); they persist because they have socio-interactional purposes (204); they are distributed around the brain on the basis of how they are used (251).

problem in identifying a unit in a different language. A code-switched unit is so strongly marked out as different, thus might make it 'appear to be' formulaic. Another problem is how are we to tell whether a formulaic sequence is prefabricated with sufficient confidence to use for testing the hypothesis? Even here though the issue comes down to one of differentiating between definition and identification. The purpose of that definition is to draw a box round what we are going to include in our discussion, when we start trying to figure out what is going on. It's not therefore suitable for identifying examples. We need to employ this definition as a starting point and establish some criteria by which formulaic sequences can be independently detectable.

3 Eleven criteria for identifying formulaic sequences

In this section, eleven proposed criteria (Wray & Namba, 2003) are explained and the characteristics of formulaic sequences will be reviewed along the way.

Wray (2002: chapter 2) identifies four major characteristics in the existing descriptions of formulaic sequences in the literature: 'form', 'meaning', 'function', and 'provenance'. The four characteristics are not mutually exclusive but overlap. Some wordstrings which aren't marked in relation to 'form' can be formulaic from other perspectives. For example, 'very funny' is not marked from the perspective of 'form'. However it can be used when the actual event is not funny, which is marked from the perspective of 'meaning' or pragmatics of use. Wray and Namba (2003) offer eleven criteria[8] which should capture the multifaceted features of formulaic sequences. Table 6.1 shows which of the four characteristics criteria A to K cover.

[8] The criteria introduced here were developed collaboratively with Alison Wray, tested on some of the data in the present study, and published in 2003. I am grateful to Alison Wray for her input into the development of the criteria. In the account below, the criteria will be evaluated in the light of subsequent work on the data, and some modifications to them will be suggested

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Form	√					√	√		√		
Meaning		√								√	√
Function			√	√						√	
Provenance					√			√	√		√

Table 1 11 criteria's coverage of the four characteristics of formulaic sequences

The criteria support the researcher's intuitive judgement rather than being a stand-alone check-list. When a researcher judges a wordstring as formulaic, the criteria can be employed to explain why he or she feels that. The phrase "by my judgement" means it is the researcher's intuitive judgement. Therefore disagreement can happen between different researchers. The approach using intuition has weakness since it is subjective. Wray (2002: chapter 2) reviewed means to identify formulaic sequences, e.g. intuition, corpus research and phonological analysis, but there was no single criterion to identify formulaic sequences in a consistent way. The difficulty lies in the inability to distinguish them from novel strings because they can be grammatically regular and semantically transparent. In order to solve these problems the eleven criteria employ intuition as a starting point and each criterion plays a role in establishing reliable justification. Now we will look at the criteria one by one.

The eleven criteria for identification of formulaic sequences (Wray and Namba, 2003)

A : By my judgement there is something grammatically unusual about this wordstring.

The English phrase 'rain cats and dogs' is grammatically irregular[9]. In ordinary clauses, the intransitive verb 'rain' doesn't take any object NP and the NP 'cats and dogs' is not employed as an adverb. Since this wordstring has become formulaic, fossilization has occurred and the internal grammatical structure or the meaning of each word is not analyzed by native English speakers. This fossilization also preserves antiquated words. For example, a wordstring 'if I were you' contains the subjunctive form 'were' which many people no longer produce in novel constructions but only use in this wordstring.

B : By my judgement, part or all of the wordstring lacks semantic transparency.

The idiom 'kick the bucket' is a frequently quoted example which shows semantic opacity. Whether a wordstring is semantically transparent or opaque can be tested in terms of semantic non-compositionality. The meaning of the whole wordstring, i.e. 'to die' cannot be derived from the sum of the meaning of its individual parts. This example is what Moon (1998:23) calls an "opaque metaphor"[10], where the meaning is unintelligible without "general or etymological knowledge" (Wray, 2002:57). However, wordstrings might be placed in a continuum between the opaque, non-compositional end and the transparent, compositional end. Some wordstrings appearing in the middle of the continuum can be formulaic. The meaning of 'spill the beans' as a whole looks fairly non-compositional but the meaning is intelligible with general knowledge. It means 'tell a secret' and it is possible to map 'spill' onto 'tell' and 'beans' onto 'secret' whereas with 'kick the bucket' there is no such obvious mapping. In the same vein, the meaning of the whole of 'like a fish out of water' and that of its parts corresponds well, therefore this wordstring is compositional.

[9] Moon (1998:82) categorises this wordstring "ill-formed FEIs" (FEI stands for Fixed Expressions and Idioms).
[10] This corresponds to Cowie's subcategory of *pure idioms.*

Nevertheless the metaphorical message of the whole wordstring is not transparent. That is, the speaker is not talking about a fish or water. Wray suggests (2002:58) that when a wordstring has a literal meaning, it can have "a secondary, layer of pragmatic meaning". For example 'very funny' can express the opposite of its literal meaning, when the situation indicates that the speaker is talking about something not funny at all.

C : By my judgement, this wordstring is associated with a specific situation and/or register.

A specific wordstring always employed in a specific situation is formulaic. The Japanese wordstring *Itadakimasu* which means 'I'm going to eat' is bound to a situation when you start eating. As for register, social relationships can be indicated by formulaic sequences. For example, in Japanese schools, when students address their teacher in class they say *sensei* 'teacher' rather than each teacher's name. It should be noted that the association between the wordstring and the situation or register should be strong, otherwise any wordstring can be positive since it is always in some situation or resister.

D : By my judgement, the wordstring as a whole performs a function in communication or discourse other than, or in addition to, conveying the meaning of the words themselves.

Functions in communication can be accounted for in terms of the 'interpersonal' function in Halliday's (1994) terminology. For example, Butler (2003:182) examines several corpora[11] and identifies that "a large group of sequences have an 'interpersonal' function, relating to speech acts being conveyed or the attitudes of the addresser". Moon (1998:218) proposes several functions of formulaic sequences in relation to Halliday's interpersonal function, such as, 'evaluative' conveying speaker's evaluation and attitude, e.g. 'kid's stuff'

[11] The corpora consists of his own study on Spanish corpus and other studies in corpus linguistics (e.g. Moon 1998, Altenberg 1998, Biber *et al* 1999).

and ‘modalizing’ conveying truth values, advice, requests, etc. e.g. ‘you know what I mean’. In my corpus, Toshiya frequently used “I’ll tell you what” at home when he started going to school[12]. This wordstring functions as a turn claimer in conversation. The turn-claimer “I’ll tell you what” also manages the flow of the discourse. Butler (2003:183) also establishes that “many other sequences have what Halliday calls a ‘textual’ role, in terms of what we might call information management in the text” (Butler, 2003:182).

Altenberg (1998), also sees the significance of this kind of formulaic material. Laying out his proposal that the beginning part of the clause is “a thematic ‘springboard’ containing given information and some frame setting element” (1998:111) and that the rest of the clause is “the propositional core conveying new information”, he observes that the items in the first group are drawn from “a restricted store of frequently utilized items” and those in the second group from “an open set” (p.111). Discourse markers, e.g. ‘on the other hand’, are archetypal models which fit this criterion.

E : By my judgement, this precise formulation is the one most commonly used by this speaker/writer when conveying this idea.

This criterion focuses on the fixedness of formulaic sequences. Occurrence in the same form can be strong evidence that the wordstring is stored as a whole. However, in order to apply this criterion, additional data from the same informant or detailed knowledge of the speaker’s idiolect is needed. This is available to a certain extent in this data but might not be available in some data. Erroneous and idiosyncratic wordstrings can be identified with this criterion.

F : By my judgement, the speaker/writer has accompanied this wordstring with an action, use of punctuation, or phonological pattern that gives it special status as a unit, and/or is repeating something s/he has just heard or read.

[12] When they lived in the UK for one year.

Some socio-interactional routines are expressed with an action. For example in Japanese kindergartens, pupils say *Sensei ohayo-gozai-masu minasan oyaho-gozai-masu* 'good morning teacher, good morning everybody' with a specific way of bowing, intonation and rhythm. There are orthographical cues to formulaic sequences, such as hyphenation, e.g. 'pick-you-own vegetables'. Wray (2002:36) points out that if formulaic sequences are retrieved as units from the memory, they should be produced more fluently. Pauses around the formulaic sequence might be a demarcation of the boundary of formulaic sequences. The articulation of formulaic sequences may also be less precise than novel strings (Wray, 2002:37). In the case of a repetition of what the speaker has just heard, prosodic patterns, i.e. intonation and rhythm, may be retained or may not – repeated material can sometimes take on the phonology of a formula, that is, a multiword string may carry the same stress and intonation as a single word. In the same way as criterion E, contextual and additional information is needed in order to apply this criterion.

G : By my judgement, the speaker/writer, or someone else has marked this wordstring grammatically or lexically in a way that gives it special status as a unit.

A wordstring, 'spin dry' shows its formulaicity in the passive and past forms: 'this shouldn't be spin dried', 'I spin dried it'. They don't appear as *'this shouldn't be spun dry' (which means it was spun in order to dry it, but not in a spin drier) and *'I spun it dry'. With criteria F and G, the use of a different language could be viewed as an indication of formulaicity since items from different languages stand out grammatically and phonologically. The purpose of this study is to examine whether CS material is formulaic. Therefore we should exclude language switch as a symptom in these two criteria.

H : By my judgement, based on direct evidence or my intuition, there is a greater

than-chance-level probability that the speaker/writer will have encountered this precise formulation before, from other people.

Wray (2002:106) categorises formulaic sequences in child language into two types according to the way they are acquired. One is underanalyzed strings which have been borrowed from other people's speech, the other is fused strings[13] which have been created by the child and stored. This criterion enables the researcher to identify underanalyzed strings as long as one can detect evidence in the data that the speaker previously encountered the wordstring. For example, in the corpus of the current study, the second sibling Ellis says "Look I did it all by yourself" when he completes something by himself in his play. He has heard the wordstring 'all by yourself' in his mother's speech, i.e. "Good boy! You did it all by yourself!". The fact that he keeps using 'yourself' instead of 'myself' is strong evidence that this wordstring is stored as a whole. Otherwise the judgement should rely on the norms of the speech community surrounding the speaker.

I : By my judgement, although this wordstring is novel, it is a clear derivation, deliberate or otherwise, of something that can be demonstrated to be formulaic in its own right.

The idiom 'kill two birds with one stone' has an exact translation in Japanese. *Is-seki* 'one stone' *ni-chou* 'two bird'. It is commonly observed that people change *ni* 'two' into *san* 'three' or other numbers. One can mention the actual number of achievements with the expression. The latter phrase is not formulaic however this is clearly a derivation of the formulaic sequence, *is-seki ni chou*. In order to be sure that the wordstring is a derivation of a formulaic sequence, the whole set of criteria should also be applied to the expression from which the wordstring is believed to be derived.

J : By my judgement, this wordstring is formulaic, but it has been unintentionally applied inappropriately.

[13] Fused strings will be detected using criterion K.

In the example “I did it all by yourself” (see criterion H), the use of ‘yourself’ instead of ‘myself’ fits this criterion as well. Since the speaker has learned the wordstring formulaically from another speaker, he doesn’t realize that it shouldn’t be used about himself but about his interlocutor. Therefore the form ‘all by yourself’ is correct but it isn’t applied appropriately. The wordstring for his addressee has been applied to himself.

K : By my judgement, this wordstring contains linguistic material that is too sophisticated, or not sophisticated enough, to match the speaker’s general grammatical and lexical competence.

This criterion should be applied to children and learners. Since a formulaic sequence is not analyzed, learners wouldn’t know if it contains grammatical or lexical items that are too advanced for their age. In our corpus, Toshiya reproduces some phrases which he has learned in school[14], e.g. “If you don’t know it, your Mum or Dad will write it for you”. He says this to his little brother Ellis when they are playing. At this age Toshiya still called his parents ‘Mummy’ and ‘Daddy’. If he had constructed this clause from scratch, he wouldn’t have used ‘Mum’ or ‘Dad’ which would be products of an older child. Toshiya has obviously learned this whole sentence from his teacher’s instruction (matches criterion H). Alternatively, if a learner creates a non-native wordstring and ‘fuses’ it into a formulaic sequence in order to save on future processing (Peters, 1983), it may remain in that learner’s lexicon even when his/her language develops enough to generate the correct form.

In addition to the eleven criteria, fixedness of formulaic sequences should be noted. Some formulaic sequences are highly fixed or frozen and don’t show much variability. For example, ‘by and large’ always appear in this form, i.e. syntactic or morphological changes never occur. Other formulaic sequences can be inflected. For instance ‘kick the bucket’ can

[14] When he lived in Wales, the medium of education was English.

be used in different tenses but cannot be passivized. On the other hand, 'spill the beans' can be passivized and shows more variability. Furthermore a number of formulaic sequences usually have a slot in them. Open class items, in many cases referential noun phrases, can be put there.

In order to cater for different types of data[15], guidelines are set as follows (Wray & Namba, 2003:28). If the example is error free, criteria A to I and K are applied. Criterion J is for the wrong use of a formulaic sequence and is therefore not applied to error free data. If the example contains one or more errors in its 'form', i.e. incorrect word or grammar, criteria E to J are applied to the actual wordstring. At the same time, the corrected form is examined with criteria A to E, H, I and K. If the example contains one or more errors in its 'usage', i.e. inappropriate context, all the criteria are applied to the actual usage. The appropriate form for the context is examined with criterion H and K (See Table 6.2).

		A	B	C	D	E	F	G	H	I	J	K
Criteria		Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Error Free		√	√	√	√	√	√	√	√	√		√
Error in Form	Original					√	√	√	√	√	√	
	Correct	√	√	√	√	√			√	√		√
Error in Usage	Original	√	√	√	√	√	√	√	√	√	√	√
	Appropriate								√			√

Table 2 Application of criteria to different data types (based on Wray & Namba, 2003)

The judgement is made based on a five point scale, i.e. 'Strongly agree', 'Agree', 'Don't know/Not Applicable', 'Disagree' and 'Strongly Disagree'.

[15] This is an application for children and learners. Regarding the application of the criteria for adult native speakers, see Wray & Namba, 2003:28.

4 Testing the eleven criteria

Before we apply these criteria to our data, they need to be tested for their ability to identify formulaic sequences in monolingual data. Examples already recognized as formulaic in the literature are examined first. Next, examples within their context will be examined (6.4.2). Then the criteria will be applied to examine key examples from Azuma's and Backus' papers (6.4.3).

4.1 Formulaic sequences in the literature

Formulaic sequences in the literature will be a good test for the criteria. Since they are usually classic citation examples, contextual information is not available. With regards to criteria E, F, H and K, direct evidence is not available but if I can judge from my intuition, 'Agree' at least will be given. They are not error forms, therefore criterion J is not applied. First, the idiom 'kick the bucket' (that is, in its meaning of 'die') is checked as follows (see Table 6.3).

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	SD	SA	NA	SD	NA	NA	SD	SA	SD	NA	NA

Table 3 formulaicity in 'kick the bucket'[16]

The word string is non-compositional and the meaning is totally opaque from its parts. Therefore it is scored as 'Strongly Agree' on criterion B. If the speaker hasn't encountered

[16] SA= 'Strongly agree', A= 'Agree', NA= 'Not applicable' / 'Don't know, SD='Strongly disagree', D= 'Disagree'. Positive judgments (SA and A) are written in blue, negative ones are in red, N/A is in black.

this string before, it will not occur in this form, since it is so idiosyncratic that it could not possibly be a novel construction - 'Strongly Agree' on criterion H. There is no grammatical irregularity -'Strongly Disagree' on A. There might be a lexical indication if someone said 'and he kicked the bucket, so to speak' 'Agree' on G. It doesn't seem specific to any situation or register but it is specific to colloquial usage. This might be slightly positive in criterion C but not as strong as 'Agree' thus 'Not applicable'. It doesn't have any pragmatic function- 'Strongly Disagree' in criterion D. There are no signs of deviation from another formulaic sequence - 'Strongly disagree' on criterion I.

The collocation of a verb + a noun, 'meet the demand' is termed "restricted collocation"[17] by Cowie (1998:215). Since one element 'meet' is semantically opaque and the other element 'demand' is transparent, 'Agree' is selected on criterion B. If the speaker hasn't encountered this collocation, he/ she wouldn't use it- 'Agree' on H. It is grammatically not unusual, is not specific to any situation or register, doesn't have pragmatic functions, doesn't have any grammatical or lexical indication and there is no sign of derivation-'Strongly disagree' on A, C, D, G, and I.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	SD	A	SD	SD	NA	NA	SD	A	SD	NA	NA

Table 4 formulaicity in 'meet the demand'

The institutionalized routine 'Happy birthday!' is marked 'Strongly agree' on criteria C,

[17] In a restricted collocation, one of the elements is used in a figurative sense.

D, E, F and H (see Table 6.5). It is said on a specific day (C), it has a function of congratulating the addressee on their birthday (D). It is often said with a gesture, facial expression, special prosodic features or indeed is often sung (F). Even without evidence, one can assume that this wordstring is learned as a whole from other people, probably family members, and the speaker will always use this form or another with a similar formulaic status, e.g. 'many happy returns', 'congratulations'(E).

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity / register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	A	SD	SA	SA	SA	SA	SD	SA	SD	NA	NA

Table 5 formulaicity in 'Happy birthday!'

4.2 Examples from a monolingual speech with the context

In this section we will test the criteria on monolingual's speech data with the context. The following conversation occurs on a busy roadside on the outskirts of York. The motorist asks the pedestrian how to get to the city centre.

Finding a direction-a transcript

1 A : Erm (.) I seem to be (.) a bit los::t ↑I'm↑ trying to get to Yor::k

2 B : Oh (.) ↑oh↑ well that's quite straight for::ward from here (.)↑if↑ you just

3 carry on:: down this road this is Heslington Lane:: (.) just

A : Yeah

B : carry on straight ahead:: [(.) the] road

A : [Yeah]

B : forks to the left but (.) ignore that just go straight ahead:: (.) and that's Broa::dway (0.5) when you come to the end of Broadway there are a set of traffic lights::=

A : = Ye::s ↑how↑ far's that

B : Oh:: (0.5) mile (.) probably

A : Go straight ahead for a mile =

B : = Yes:: =

A : Ignore: the left [fork]

B : [Ig] nore the left for::k

A : Yeah (.) then I get to some traffic lights =

B : = You ↑get↑ to some traffic lights (.) turn right at the traffic lights::

A : [(.) huhuh] carry

B : on down there that's the main:: road into Yor::k (.) just sort of carry on down there ↑you'll come↑ to some traffic lights:: ↑keep↑ in the right hand lane::

A : [(.) Yeah] and

B : there are some traffic [lights::]

A : [What is] it a dual carriageway then =

B : = Yes it's a dual carriageway (.) part of the way anyway

A : OK

B : If you carry er ↑keep↑ on in the right hand lane:: [(.) Yeah] you

A :

B : come to some more traffic lights there's a ↑round↑about there:: you'll ↑see↑ (.) erm (.) Clifford's Tower on the right::=

A : = Wha- what's:: Clifford's Tower =

B : =Yes it's a big tower on a mou::nd ↑you'll be actually::↑ riding alongi-on the road alongsi::de (0.5) Clifford's tower (.) [You're O k::] in York then

A :

B : ↑There's also↑ by Clifford's Tower there's a (.) car park you can park your car there and then:: =

38 A : OK So I look out for:: a tower on top of a mound and
39 he[ad] towards that:: [(.)]
40 B : [Ye:s::] [Yes::]
41 Ye[s::]
42 [Wh] en I get in /tu::/ =
43 A : =↑Yeah↑ when you get into[/ði::/]
44 [City]
45 B : ↑Into↑ the city (0.5) er at the traffic lights you'll be able to see:: Clifford's
46 Tower from the traffic lights::=
47 A : = OK (.) Fine (.) Thanks a lot (.)↑Bye

(Langford, 1994:158)[18]

A phrasal verb 'carry on' is observed four times in the text, i.e. 'carry on down this road' (line 3), 'carry on straight ahead the road' (line 5), 'carry on down there' (line 18-20 and again in line 20). The structure of the wordstring is 'carry on' + adverbial + [place]. Four time occurrence from the same speaker indicates 'Strongly agree' on idiolect (E). This is specific to the situation of explaining a direction- 'Strongly agree' on C. With regards to Criterion B, the meaning of 'carry' doesn't directly correspond to the meaning of the whole wordstring, i.e. 'continue', however 'on' corresponds to it. Therefore 'Agree' is given to this criterion. There is no evidence of previous encounter but we can guess she has encountered this wordstring before hence 'Agree' is given.

[18] Capitalization = loudness, underline = stress, ↑ = raised pitch height, colons = stretched segment (Langford, 1994:49)

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SD	A	SA	NA	SA	A	NA	A	SD	NA	NA

Table 6 formulaicity in 'carry on+ [Adv]+ [place]' (line 3)

The adverbial phrase 'straight ahead' is observed three times, i.e. 'carry on straight ahead the road' (line 5), 'go straight ahead' (line 7 and 12). The pattern can be described as '[Verb for movement] + straight ahead'. We will examine this wordstring in line 12. The speaker A is confirming the speaker B's explanation in line 7, which is direct evidence for criterion H- 'Strongly agree'. It is specific to the situation of direction-'Strongly agree' on C.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SD	D	SA	NA	A	NA	NA	SA	SD	NA	NA

Table 6.7 formulaicity in '[Verb for movement]+straight ahead' (line 12)

With regards to another phrasal verb 'look out for [NP]' in line 38, the prolonged part 'for::' can be an indication that there is a gap after this wordstring-'Agree' on 'performance indication' (F). The meaning of the wordstring would be the same without 'out', i.e. 'look for' and 'look out for'- 'Strongly agree' on B.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity / register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SD	SA	NA	NA	NA	A	NA	A	SD	NA	NA

Table 6.8 formulaicity in 'look out for [NP]'

The wordstring 'sort of' in line 20 is a conversation filler. It gives time for the speaker to think- 'Strongly agree' on D. Grammatically an NP should follow the preposition 'of' but 'sort of' does not specifically attract an NP (ie. 'of' is not playing is normal independent role)— the verb phrase 'carry on down there' follows here. This is a grammatical indication that 'sort of' is formulaic. 'Strongly agree' on A and G. There is no direct evidence but she must have encountered this wordstring before-'Agree' on H, and will use it again-'Agree' on E.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity / register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SA	NA	NA	SA	A	NA	SA	A	SD	NA	NA

Table 6.9 formulaicity in 'sort of +VP' (line 20)

'Thanks a lot' (line 47) is a conversational formulae. It has a pragmatic function of showing his gratitude- 'Strongly agree' on D. He must have encountered and will use in this form- 'Strongly agree' on E and H. Grammatically the combination of the noun 'thanks' followed by the adverbial phrase 'a lot' is idiosyncratic- 'Strongly agree' on A.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	SA	SD	NA	SA	SA	NA	NA	SA	SD	NA	NA

Table 10 formulaicity in 'Thanks a lot' (line 47)

Although we haven't analyzed the whole text exhaustively, it has been demonstrated that with the context, more robust analysis can be done.

4.3 Examples from Azuma and Backus

Lastly we will discuss the key examples from Azuma and Backus. Our agenda is to find examples which are not grammatical constituents but are formulaic. Examples (6), (7) and (8)[19] are counter-examples to Azuma's hypothesis because switching doesn't occur between syntactic constituents. Azuma suggests the performance structure might explain this. We will examine the role of formulaicity now.

↓

(6)	Living in Austin is	*totemo*	*ii*	*keiken*	*ni*	*nari*	*-masu*
		very	good	experience	RSL	become	HON

{Living in Austin is a very good experience}

(Azuma, 1996:412)

The wordstring 'Living in Austin is' is grammatically regular, semantically transparent, not specific to situation or register, doesn't have pragmatic function-'Strongly disagree' on A, B, C, D. There is no direct evidence to mark 'Agree' on E, F,G and H. We don't detect formulaicity from this wordstring alone.

[19] Example (6) was already introduced as (3) in section 1.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity / register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	SD	SD	SD	SD	NA	NA	NA	NA	SD	SD	SD

Table 11 formulaicity in 'Living in Austin is'

The next example (7) has a similar basic structure, 'NP + is' to example (6). The inner structure of the NP, i.e. consisting of one pronoun is simpler than that of the NP in (6).

↓

(7) He's uh MBA *no* *kurasu* *o* *totte-* *ite*

GEN Class ACC take PROG

{He's uh taking a class for his MBA}

(Azuma, 1996:412)

The contracted form 'He's' and the filler "uh" after it[20] can be an indication of formulaicity- 'Agree' on F. One might argue that language changes after each wordstring might be an indication of formulaicity therefore at least 'Agree' should be scored in criteria F and G. However since we need independent evidence for the relation between formulaicity and code-switching, language change alone is not counted as evidence for formulaicity. In other criteria, we can't find any positive judgment for the formulaicity.

	A	B	C	D	E	F	G	H	I	J	K

[20] However, more reasonable explanation will be that the role of the filler is to take time for switching.

Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SD	SD	SD	SD	NA	A	NA	NA	SD	SD	SD

Table 12 formulaicity in 'He's'

With the next example, the switching occurs after a couple of words.

↓

(8) and then he'll go to *otearai* *ni* *ikimasu*

toilet to go

{and then he'll go to the toilet}

(Azuma, 1996:412)

The wordstring 'go to' doesn't have evidence for formulaicity in most criteria. He will use this combination for the same meaning again, therefore 'Agree' on E. This example shows 'portmanteau structure' (see chapter 5). That is, the meaning of 'go to' are repeated in Japanese. The structure of the Japanese part looks like a mirror image of the English part.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / register specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with
Judgment	SD	SD	SD	SD	A	NA	NA	NA	SD	SD	SD

Table13 formulaicity in 'go to'

The examples which entail the performance structure don't show strong formulaicity with

our current criteria. However it should be noted that Backus (2003:p115) cited these two particular examples of Azuma (7) and (8) as 'constructions' therefore 'lexical units' so he at least considered them formulaic. Next we will examine Backus' example of 'construction'.

> (9) *die is de slechte persoon,* ondan sonar coğunlunkla yapıncık, mesela altı kişi yapıncık o zaman artık *o ja* normal görünür, ama o ilk kişi *DIE IS GEWOON, ja,* en berbatı
>
> (Turkish in normal font, Dutch in italics, lexical units are capitalized)
>
> { *she's the bad person,* and then the majority will do it, for example if six people will do it, from then on it's seen as "*oh sure*," as normal, but that first person, *SHE'S JUST, well,* the worst}
>
> (Backus, 2003:118)

According to Backus (2003:119) the Dutch adverb gewoon has a basic meaning of 'normal' but in this construction its meaning is 'just'- 'Agree' on semantic opacity (B). It is used "if one wishes to make an emphatic statement about someone, stating a quality that is either surprising or characteristic, but in any case, according to the speaker, very noteworthy" (2003:119). At least 'Agree' can be given to D. There isn't direct evidence but this construction might have been acquired from someone else and might be used again- 'Agree' on E and H.

	A	B	C	D	E	F	G	H	I	J	K
Criteria	Grammatical irregularity	Semantic opacity	Situation / specificity / register	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch with maturation
Judgment	SD	A	SD	A	A	NA	SD	A	SD	NA	NA

Table 14 formulaicity in *die is gewoon*

5 Discussion

In the last section the three different groups of wordstrings are tested to verify the eleven criteria. Whereas the criteria work fine for the examples in 1) already recognized formulaic sequences without the context and 2) monolingual dialogue with the context, they don't detect strong formulaicity in the examples in 3) non-syntactic constituents from Azuma's and Backus' papers. In this section we will assess the eleven criteria and propose solutions to problems.

The analysis has not been done exhaustively but the result of this small experiment shows that formulaic sequences are multi-faceted phenomenon. Out of the eleven criteria, criterion B 'semantic opacity' and D 'pragmatic function' seem to be strong ones. Even when other criteria are not marked, marking on either of these two alone can be evidence for formulaicity. Criteria C, E, F and H can be stronger if more contextual information is available. Criterion C can be a strong criterion if the situation or register is specific such as the dialogue of finding a direction in 6.4.2. Criteria E (idiolect) and H (previous encounter) can be judged by a researcher's intuition and at least 'Agree' can be scored. When direct evidence is available 'Strongly agree' can be given. With criterion F (performance indication), contextual information is crucial. Criteria E, H or F alone cannot be strong indicators of formulaicity but they will support other criteria. Formulaic sequences marked by criteria A, I, J, K have not been encountered, but this might be attributed to the limitation in the number of samples we tested.

Some gaps have been noticed. In the literature 'genre' has been mentioned, e.g. Gläser (1998:143) examines the use of formulaic sequences in a variety of genres from popular science articles to literary texts. She finds specific roles of formulaic sequences in specific

genres, for example in text books they are employed to "enhance the intelligibility and memorability of a text". Perhaps we should include 'genre' in criterion C. It will be modified as follows.

C : By my judgement, this wordstring is associated with a specific situation, register and/or genre.

Biber *et al* (1999:989) point out that idioms such as 'kick the bucket' are used occasionally in fiction but rarely in other genres. Therefore 'kick the bucket' is given 'Agree' on criterion C (it was previously NA- see table 6.3).

The non-syntactic constituents observed in alternational CS in Azuma's and Backus' papers, namely the 'performance structure' and 'constructions' are good examples to verify formulaicity independently. However the test doesn't show strong formulaicity in them. Pawley and Syder's (1983:210) lexicalized sentence stems are similar to 'constructions'. An example of a sentence stem is 'NP be-TENSE sorry to keep-TENSE you waiting'. An actual form can be 'I'm sorry to have kept you waiting', for example. It looks like a novel sentence but there is an underlying frame which is formulaic. In the current criteria, there is no way to ensure that this sort of example is captured. We need another criterion to capture underlying frames.

L : By my judgement, there is an underlying frame and one or more gaps in this wordstring. The frame is formulaic and the gaps can be filled with any lexical items.

This criterion alone is not strong enough to verify formulaicity. If other criteria, such as pragmatic function (D) or semantic opacity (B) are marked, this criterion will be more robust. The example 'NP be-TENSE sorry to keep-TENSE you waiting' will be analyzed

with the modified criteria as follows.

	A	B	C	D	E	F	G	H	I	J	K	L
Criteria	Grammatical irregularity	Semantic opacity	Situation /register/genre specificity	Pragmatic function	Idiolect	Performance indication	Grammatical/ lexical indication	Previous encounter	Derivation	Inappropriate application	Mismatch maturation with	Underlying frame
Judgment	SD	SD	SD	A	A	NA	SD	A	SD	NA	NA	SA

Table 6.15 formulaicity in '[NP] be[-TENSE] sorry to keep[-TENSE]you waiting'

Several examples in the dialogue of direction (6.4.2) will be scored on this criterion. The wordstrings, 'carry on + [ADV] + [Place]' (line3), '[Verb for movement]+straight ahead' (line 12),'look out for + NP' (line 38) and 'sort of +[VP]' (line 20) are all scored 'Strongly agree' on criterion L. With regards to Azuma's example (6), (7) and (8), 'Agree' on criterion L will be given to all of them. Example (6) 'Living in Austin is [ ]' and (7) 'he's [ ]' can be seen as an 'X is Y' formula. More specifically (6) can be an '[Activity] is [Attribute]' formula and (7) can be a '[Person] is [Activity]' formula. The switch occurs at the gap in both cases. With example (8), '[NP] go to [Destination]'can be a formula. Here it is weakened by the fact that 'the toilet' is not a destination and 'go to the toilet' is formulaic[21] however the Japanese part *otearai ni ikimasu* 'go to the toilet' is formulaic, which indicates that CS occurs at the boundary of a formulaic sequence. Switch can be either before or after a formula. Backus' example (9) *die is gewoon* [ ] 'she's just' will be given 'Strongly agree' since it has pragmatic function which supports formulaicity strongly[22]. This criterion seems to be a useful and potentially robust one, so it will be

[21] 'Agree' on B, 'Strongly Agree' on C and E.

[22] I have to depend on the judgment done by Backus (119:2003), since I don't have much intuition about Dutch/Turkish. It is inevitable that in any approach that uses intuition, there will be constraints when you don't know the language.

adopted in the remainder of the analysis.

It has turned out that alternational CS is a good place to explore such examples, but these examples are too few and are, of course, the very ones that Azuma and Backus themselves found problematic, so let us reserve judgement on the question of whether CS occurs at formulaic or constituent boundaries until we have examined the data from the main dataset.

References

Altenberg (1998) On the phraseology of spoken English: the evidence of recurrent word-combinations. In A.P. Cowie (ed.) *Phraseology: theory, analysis and applications.* Oxford: Clarendon Press, 101-122

Azuma, S. (1996) Speech production units among bilinguals. *Journal of Psycholinguistic Research* 25 (3) :397-416.

Backus, A. (2003) Units in code switching: evidence for multimorphemic elements in the lexicon. *Linguistics*. 41 (1): 83-132

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999) *Longman Grammar of Spoken and Written English.* Harlow: Pearson Education Limited

Butler, C. S. (2003) Multi-word sequences and their relevance for recent models of Functional Grammar. *Functions of Language*. 10 (2): 179-208

Cowie, A.P. (1998) Phraseological Dictionaries. In A.P. Cowie (ed.) *Phraseology: theory, analysis and applications.* Oxford: Clarendon Press, 209-228

Gläser, R. (1998) The stylistic potential of phraseological units in the light of genre analysis. In A.P. Cowie (ed.) *Phraseology: theory, analysis and applications.* Oxford: Clarendon Press, 125-143.

Halliday, M.A.K. (1994) *An Introduction to Functional Grammar.* London: Edward Arnold

Langford, D. (1994) *Analyzing Talk*. London: The Macmillan Press Ltd.

Moon, R. (1998) *Fixed Expressions and Idioms in English.* Oxford: Clarendon Press.

Pawley, A. & Syder, F.H. (1983) Two puzzles for linguistic theory: native like selection and native like fluency, In J.C. Richards & R.W. Schmidt (eds.) Language and

Communication. New York: Longman. 191-226.

Peters, A.M. (1983) *Units of Language Acquisition.* Cambridge: Cambridge University Press.

Myers-Scotton, C. (2002) *Contact Linguistics.* Oxford: Oxford University Press.

Sinclair, J. (1991) Corpus, Concordance, Collocation. Oxford:Oxford University Press.

Radford (2004) English Syntax: an introduction. Cambridge: Cambridge University Press.

Wray, A. (2002) *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Wray, A. and Namba, K. (2003) Use of Formulaic Language by a Japanese-English Bilingual Child: A Practical Approach to Data Analysis. *Japan Journal of Multilingualism and Multiculturalism.* 9 (1): 24-51