



Formulaic Language for Communicative Competence: Native Speakers and ESL Learners

Dilafroz Abdulkhakova Shamsiddinovna

Teacher, Uzbekistan state world languages university, Faculty of English Philology, Uzbekistan, Tashkent

Abstract: This article emphasizes the speech patterns in native speakers' and language learners' speeches. There are many definitions of formulaic language that have been suggested so far. The author has conducted the research in terms of lexical chunks and the findings might suggest that formulaic language (expression) is one part of the communicative approach to the vocabulary system of a language. The speech patterns may vary according to the use of English.

Keywords: Formulemes, ESL, EAP, L2 learners, mutual information (MI), corpus linguistics, native speakers, formulaic sequence, conventionalization, learner's corpora / British National Corpora.

Recurring formulaic word patterns are frequently used in natural language. The construct of formula is triangulated in this article from the viewpoints of education, psycholinguistics, and corpus linguistics. It explains the corpus linguistic extraction of formulaic sequences for academic speech and writing that are pedagogically beneficial. It establishes instructors' assessments on the pedagogical significance of English as a Second Language (ESL) and English for Academic Purposes (EAP) instruction. It summarizes three studies that demonstrate how different characteristics of formulaicity impact how accurately and fluently advanced L2 English learners and native English speakers understand these formulas. The tasks for language processing were chosen to represent a variety of spoken and written, production and comprehension abilities that are ecologically valid. Numerous metrics obtained from corpora, such as length, frequency, and mutual information (MI), had an impact on processing across all tests, though to varying degrees depending on the populations. For native speakers, the MI of the formula mostly dictates processability; for non-native language learners, the frequency of the formula primarily determines processability. These findings have implications for (a) the psycholinguistic validity of corpus-derived formulas, (b) a model of their learning, (c) the prioritizing of which formulas to teach in ESL and EAP classes, and (d) ESL and EAP instruction.

There has been a great deal of scientific works related to formulaic language and variety of linguists' viewpoints in terms of formulemes or chunks of the vocabulary system. Surprisingly, scientists' opinions include not only linguistic definitions or explanations, but also psychological even medical ones. To begin any discussion of formulaic language, it is important to establish some foundations and establish the terminology that is used to refer to it, and to look at a definition or definitions. *Formulaic sequence* is generally used to refer to one such item, *formulaic language* is the uncountable noun referring to these items as a collective, and *phraseology* is a term often used to refer to the study of formulaic language.

The general consensus on a definition of formulaic language seems to be that the items will be (Wood 2015:2):

1. Multi word
2. Have a single meaning or function
3. Be prefabricated or stored and retrieved mentally as if a single word

From my point of view, formulaic expressions are the ready-made chunks of a lexicon that can be either formal (*Good day Sir/Madam, Good morning*) or informal (*What`s up, Hey (man), How are ya?*). These speech patterns are considered as formulas as there is no change in their overall structure and cannot be altered by adding other parts of speech or even a word. However, it is possible to substitute the words by utilizing the same formula or pattern to create a new word phrase which is used in various discourse like *have a good day-have a good journey-have a good class-have a good flight-have a good weekend-have a good trip-have a good holiday-have a good vacation-have a good party-have a good night..* Furthermore, the formulaic language might be consisted of one word (*Attaboy, Okay, Bye, Howdy?*) two words (*Help yourself, Dig in, See you/ya, Take care, Hi there*) and several words even a whole sentence (*Long-time no see, How are things?, How have you been?, It`s been a while*). Hence, formulaic language is increasingly becoming popular due to the communicative approach to the study of a lexis (vocabulary system of a language).

According to W. Langacker such natural-sounding expressions should be recognized as linguistic units in their own right so that they can be used to capture native-like selection in L2 development. Due to the fact that they cannot be easily defined in terms of conventional language subsystems like grammar, vocabulary and phraseology, these expressions are often ignored in L2 research and instruction (Langacker 2008). Usage-based approaches define language as an inventory of symbolic units with differing degrees of specification, with no strict separation between grammar, lexicon and phraseology. All linguistic units are seen as conventionalized pairings of form and meaning, regardless of their size and internal structure (Langacker 2008). Frequency of forms in the input is seen as one of the main factors driving acquisition (Ellis and Cadierno 2009) and each learner is assumed to discover the regularities and patterns of an L2 through exposure and experience with the language.

As A. Pawley & F. H. Syder pointed out there are many grammatically correct ways of expressing a notion beyond word level, there are only some that are preferred choice among proficient users of the language (Pawley & Syder 1983). Native speakers are able to select the right combination of words out of the countless option allowed by traditionally described grammar and lexicon of the language, but L2 learners often make the mistake of combining words in grammatically correct but awkward-sounding ways. Native-like word combinations are generally referred to as chunks such as formulaic sequences, multi-word units, lexical phrases, etc., which are more or less fixed word sequences characteristic of fluent native-like language use.

The exact definition of a chunk differs across theoretical and methodological approaches depending on their research focus and purpose. S. Granger and M. Paquot (2008) present a comprehensive overview of the field of phraseology (the study of multi-word units). They distinguish between two major approaches to multi-word units: traditional phraseological approach and frequency-based approach, both of which have a rather different scope. Traditional phraseological approach (Nesselhauf 2004, Cowie 1998) is mainly concerned with distinguishing between different types of multi-word units on the basis of linguistic criteria such as non-compositionality (idiomaticity) and fixedness (e.g. *spill the beans*), and separating fixed multi-word units from free combinations (e.g. *to have skeletons in one`s cupboard*).

According to J.M. Sinclair and D. Biber, the frequency-based or corpus-driven approach (Sinclair 1991, Biber 2000) describes frequently occurring word combinations based on corpus-derived measures such as frequency and collocational strength, which may not fit linguistic categories (e.g. *the back of the*, Biber 2000). Finally, psycholinguistic approaches and SLA approaches define

chunks mainly as units processed and stored as a whole (Schmitt and Carter 2004, Wray 2002, 2008).

Even within these many valid perspectives, there is lack of agreement on which particular word sequences count as fixed chunks and which do not. This is partly because chunks are an overlap between traditionally defined language subsystems, such as morphology, lexicon, grammar, phraseology and discourse, and partly because their defining features (e.g. non-compositionality, fixedness, function) form a continuum, not discrete categories. Some word sequences are widely recognized as fixed chunks because they are prototypical examples of recognized categories; but there is some controversy about word sequences that lie at the end of the continuum of characteristic features. As J. Bybee stated that these units are predictable and regular in terms of syntax and transparent in their meaning and may not be considered fixed. Nevertheless, Wray singles out that they are intuitively formulaic and native-like because they represent the preferred ways of saying things out of all the options in principle allowed by the grammar and the lexicon of the language.

By taking a usage – based (UB) perspective H. Smiskova, M. Verspoor and W. Lowie carried out the research in terms of specific linguistic units which can be defined as the term *conventionalized ways of saying things* (CWOSTs). According to the above mentioned three researchers the term itself consists of three parts, the first one is *conventionalized (language)* refers to the expressions established as the preferred formulations of a particular notion. The other part are the linguistic form: multi-word expressions which are deemed as *ways of saying* and the last part is about the things: notions, themes, concepts beyond word level. Nevertheless, there might be the opposite term to the CWOSTs which is used to express ideas in relatively strange ways and this can be defined as *awkward ways of saying things* (AWSTs). The word *awkward* here stands for expressing the thoughts *in principle possible but not established as a preferred expression*. Hence, this phenomenon refers to the context of L2 development, and CWOSTs are considered as natural sounding or native-like expressions which play a significant role in the differentiation of students between high and low-input learners in the category of preferred ways of saying things. Clearly, this category of expressions captures an important difference in L2 development and deserves more attention.

Conventionalization is closely related to frequency of occurrence in learner corpora. All linguistic units, including sentences such as *You know what I mean* have a token frequency and conventionalized units are assumed to occur with high frequency. As N. Schmitt pointed out that one defining characteristic of a CWOST could be its high token frequency in reference corpora, preferably defined as a frequency band with a strict threshold. Stefan Th. Gries (2008) also argues strongly in favour of using Corpus Linguistic methods for SLA research from Cognitive Linguistics perspective. However, there is evidence that conventionalized expressions do not necessarily have high token frequency, as is the case in these pure idioms like *break a leg – frequency 17, under the weather- frequency 31, call it a day – frequency 55*. Bybee argues that even a simple two-word collocation such as *experience delays* is not highly frequent, even though it is established as the conventionalized way of expressing a certain notion. This suggests that the token frequency of an expression may be related to the frequency of the notion in question: if a notion occurs less frequently, the conventionalized expression (CWOST) will also occur less frequently. Finally, longer word sequences will automatically have lower token frequency than shorter word sequences. This suggests that a purely corpus-derived frequency criterion may not be sufficient to define CWOSTs in general. However, token frequency should be enough to identify the conventionalized formulation out of a range of possible ones, because it will likely be the most frequent one. Possibly, the frequency distribution of all expressions within that range might be the most frequent expression. This could potentially interfere with the identification of a CWOST, as the most frequent expression could simply be a prototype, i.e., the most generic in meaning and therefore most frequently used.

Corpus-derived frequency information is only a reflection of how linguistic forms are being used by language users: conventionalized expressions are shared in a speech community and the creators of conventionalization are language users themselves. Psycholinguistic research shows that fluent native speakers are highly sensitive to frequencies of linguistic forms (Ellis 2008) and “store probabilistic relations between words”. Also, S. Wulff (2008) presents evidence that non-expert NS

judgment of idiomaticity is highly consistent. These findings are consistent with usage-based approaches where frequency plays a crucial role and leads to conventionalization. From my point of view having the inner intuition of a human being may play a role in defining the expressions as frequently used among both native speakers and proficiency level users of language. The formulaic expressions might not be easily identified from the corpora as they are multi-word expressions.

According to my own investigation, English native speakers are prone to utilize lexical chunks (bundles), collocations, idioms and other speech patterns which I refer here to formulaic expressions (language). Hence, students must get immersed in formulaic expressions to master the English language and sound like a true Londoner or American. In fact, many language learners have great difficulties to produce formulaic expressions thus their utterances sound unnatural and foreign to English natives. Formulaic language is of great importance since this fast-moving world is increasingly becoming formulated by clipping and shortening the words even whole phrases vs sentences. Conventionalized ways of saying things assist to express one's thought beyond-word-level notion which many L2 learners find arduous, so some of them use their L1 to say something in L2. This matters to misunderstanding between the participants of communication as well as decrease L2 development. By selecting topic specific vocabulary to every topic and creating a particular context (dialogue) during the L2 class it is possible to boost the learners' lexical resource and collect certain natural-sounding formulemes. These phrases are quite beneficial as learners may not change the structure of the speech formulas and straightly adapt them to their conversation. Furthermore, the frequency of these formulemes might be identified in learner's corpora / British National Corpora. It can be concluded that formulaic set expressions are, in fact, beneficial for language learners, as this phenomenon assists to maintain a natural balanced conversation like native speakers of the target language.

References:

1. Bybee J.L. Usage-based grammar and Second language acquisition ,2008, P.236.
2. Biber D. Lexical bundles in academic speech and writing. In Lewandowska-Tomaszczyk, B. (ed.). Practical Applications in Language and Computers (PALC 2003), 2004, P.165–178. Frankfurt am Main: Peter Lang.
3. Wray, A. Formulaic Language and the Lexicon. 2002, Cambridge: Cambridge University Press.
4. Schmitt, N. and N. Carter. 'Formulaic sequences in action: An introduction' in N. Schmitt (ed.): Formulaic Sequences, Amsterdam: John Benjamins, 2004, pp.1-22.
5. Spears R.A. NTC's Dictionary of American slang and colloquial expressions, 2000. 560p.
6. Pan, Fan & Hu, Rulan. (2017). David Wood, Fundamentals of formulaic language: An introduction. London: Bloomsbury, 2015. Pp. vii + 198.. Journal of Linguistics. 53. 1-5. 10.1017/S0022226717000081.