# On the Cognitive Characteristics of Language Chunks

L.H.Ma & Y.Li Zhejiang Ocean University, Zhoushan, China China Foreign Affairs University, Beijing

ABSTRACT: Language chunks refer to the fixed or half fixed structure of vocabulary words and sentences, which is stored in the brain as a whole, and exert an influence on language output. In First Language Acquisition, children experienced a stage at which unanalyzed language chunks were heavily used in a particular context; Second Language Acquisition also follows the same natural path. This path shows that language chunks acquisition is the center of language acquisition. This article has discovered that language chunks have strong language generating capacity: the use of them can accelerate the rate of language processing, improve the quality of the language output and promote learners language fluency and idiomaticity.

KEYWORD: language chunks; open-choice principle; retrieval; idiomaticity; generating capacity

## **1 INTRODUCTION**

A formulaic sequence is a sequence of words with a relatively stable structure form, sentence function, purpose of meaning, and is used frequently in linguistic communication. It is directly applied in language as a whole without syntactic analysis. Some formulaic sequences are unable to perform syntactic analysis, and other times it is usually not necessary or efficient to do so. Use of formulaic sequences is defined by Sinclair (1991) as the idiom principle, abided by all language users. Another principle is named open-choice principle, referring to constructing sentences from scratch with grammar and vocabulary. Sinclair proposed the priority of the idiom principle over open-choice principle by arguing the latter involves the complex and unnecessary process of breaking down sentences. His research was later recognized by linguists that language embodies dual attributes: analytical feature and procedural feature, which means language is both an analyzable system based on grammar and a formulaic system based on memory (Skehan, 1998). Therefore the linguistic expression is a process of submitting to syntax rules and retrieving phrasal units from memory. Despite countless combination of phrases allowed by grammar, neglected and unanalyzed formulaic sequences repeat the most in linguistic communication. It is according to corpus statistical analysis that most communication between human are realized through fixed and semi-fixed formulaic chunks. The high frequency of formulaic language appearing in both spoken and written discourses far exceeds that of sentences constructed on grammar.

#### 2 STRUCTURE CLASSIFICATION OF FORMULAIC SEQUENCES

Many Chinese and foreign linguists have suggested different rules of classification (Natiinger & DeCarrico, 1992; Lewis, 1997; Wray, 2002, 2008; Lifei Wang, 2006; Shiping Duan, 2008). The author observed that proposals of Wraya and Nattinger & DeCarrico share traits of similarities in that they both identify formulaic sequences by the structure and function of meaning, which shows more practicality in research. Four types of formulaic sequences are classified as follow:

A. Multiword items

Set phrases composed of one or more words,

e.g. once in a while,

Fixed idiom expressions,

e.g. nuts and bolts.

B. Institutionalized utterances

Proverbs, mottos and social gambits,

e.g. How are you?

Happy New Year!

C. Phrasal constraints

Sentences with changeable inside structure, which allows for vertical and horizontal replacement and

insertion of new words,

e.g. If I were you,...

D. Sentence-builders

Lexical phrases and quotations that build up sentence structure,

e.g. My point is...;

Firstly...

Judging by the structure characteristic, formulaic sequences are not distinctly different from general language. Its changes in expression form from fully fixed to semi-fixed expression actually realize a gradual transition to general language, from vocabulary level to sentence level. Diminishing fixity of sequences, increasing fragments between language components, more flexibility of expression and larger tolerance of syntactic variations make possible the creative use of language.

## 3 COGNITIVE CHARACTERISTICS OF FORMULAIC LANGUAGE

Research on language process and output revealed that language users usually rely on repeated use of existing word chunks and pre-stored sentence structure in the brain, instead of newly created structures to fluently and accurately express things.

From the view of economics, human needs to store or semi-store situations and things that are repeatedly sensed and experiences in forms of concepts, categories and graphic to achieve highest economic efficiency of cognition. Formulaic sequences can relieve cognitive pressure to a large extent as a stored or semi- stored graphic language. In other words, by spending minimum process time on existing language sequences, the human brain saves cognitive resources to perform other task such as constructing discourse. Native speakers are fluent in their mother tongue because they store in memories plentiful and readily available language sequences. There are two major cognitive features of formulaic language.

# 3.1 Easiness to retrieve

The chunking function of formulaic language is highly efficient to process language on-site. Word chunks, instead of a string of scattered words are stored inside the brain and retrieved directly. The process of handling word chunks from the first to last reflects the advantage of whole words retrieval, much faster and easier compared to other process of generating word strings on-site. Nattinger & DeCarrico (1992) believed a speaker's language proficiency is not dependent on number of syntax rules he masters, but on the number of pre-stored word chunks in the brain. Retrieving word chunks during a fluent linguistic communication greatly reduce the workload for the brain to organize language, because there's no need to breakdown the internal structure of every sentence to express ideas. Further research indicates that many rules reflected in formulaic language are fixed in the formulaic language and no longer restricted by the contextual syntax rules, which proves formulaic sequences' high grammar compatibility in the deep structure language system. Scholars believe word chunks' internal components are highly correlated and a single missing component can cause loss or misrepresentation of semantics and pragmatic information. Strong superiorities in semantic attribute, contextual restriction and compatibility attribute make word chunks easier to retrieve and appears frequently in linguistic behavior, which is referred to as the frequency effect advantage: high frequency language units take less time to process efficiently, and therefore are activated earlier and at a more frequent level in language out. And because formulaic sequences usually possess specific and clear semantic, contextual and pragmatic function, situations of inconceivable semantics or delayed activation in other language forms does not exist.

Observation of language learners proves that use of formulaic language increases proficiency of their linguistic communication. Language proficiency is not defined by swiftly application and organization of rules, but a result of coding example word chunks into complete unit according to context (Robinson, 1993). Without pre-stored word chunks, the concept of proficiency is nonexistent. The effect of formulaic language works in at least two ways:

- (1)Formulaic sequences are stored and retrieved with a complete language form, which don't require new grammar restriction in expression. Coding cost of the brain is therefore reduced to minimum to ensure swift reaction of brain and shorter time of language output. Under limited time, language acquisition and communication is accelerated to the fullest.
- (2)Language chunks possess semantic attribute and doesn't require much grammar or pronunciation adjustment to generate language due to their contextual constraint and compatibility characters, making it easier to retrieve to meet the needs of immediate communication.

In general, formulaic language represents a language structure stored in the long-term memory of language users as a complete unit, which embodies great advantages in expressiveness, pragmatics and retrieval speed. It is easy to retrieve directly from the brain and to make small adjustments for later retrieval and use based on fixity. Without pre-store or semi-stored word chunks to help process and understand language, proficiency and accuracy of expression is hard to achieve.

### 3.2 Strong language generating capacity

The concept of formulaic sequences covers habitual collocation of words and phrases, sentence structures, as well as complete discourses. The structural stability and semantic predictability make word chunks easy to learn and apply while the dispersion in the structure (the replaceable component) provides space for creativity. As ideal language memory units, formulaic sequences provide raw materials for language acquisition. Researches discover the high frequency word chunks in native speakers' corpus generally have high generativity. These word chunks can further generate dozens of similar ones based on certain syntax rules, e.g. "Generally speaking" generates "Honestly speaking", "Frankly speaking", etc. By that analogy, memorization of one word chunk means mastering the matching word chunks. As pre-made language components, language chunks themselves include inner syntax rules, representing a new dialectical view of language that vocabulary and grammar are not distinctively separated. Language is not a system of semantic grammar, but a system of grammatical semantics. Language learners memorized semi-fixed word chunks structure, and retrieve them on demand to generate endless possibilities of semantic groups. During the process, they memorize the inner syntax rules to effectively integrate vocabulary and grammar, thus avoiding distortions from the mother tongue for more accurate and authentic expressions. Moreover, with increasing language proficiency, learners will come across more complex methods to express different functions through expansion use from one formulaic word chunk. In this sense, word chunks have generating capacity as well. In second language acquisition, vocabulary mismatch caused by negative transfer of mother tongue greatly undermines fluency of expression. Memorizing fixed word chunks, sentence structure, lexical phrases and fostering habits of using word chunks reduce the influence of mother tongue in communication by improving the learners' ability to correctly match up vocabulary and create discourses from an English way of thinking.

# 4 CONCLUSION

Second language learners tend to overlook the fact that native speakers' most frequent used collocation of words and structures often disregard syntax rules and put too much weight of study of grammar. Research on formulaic sequences revealed the true characteristics of language and language habit of native speakers: a combination of format and function. Formulaic sequences are crucial to facilitate language acquisition and improve fluency of language output. Learning formulaic sequences represents a necessary stage that fits the cognitive characteristics of second language acquisition. The interlanguage system, which is often regarded as something based on language rules, is in fact a summary of the previous memories and experiences of words and sentences. Language development is a long-term process of continuous accumulation and reinforcement based on word usage, where memorizing formulaic language is more important than simply following rules of the target language. During the process, syntactic analysis isn't always necessary. Learners naturally come across syntactic structures and consciously generalize the underlying syntactic rules in repeated use of word chunks.

## REFERENCES

- Ellis, R. The study of second language acquisition. Shanghai: Shanghai Foreign Languages Education Press, 1994.
- [2] Sinclair, J. M. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
- [3] Duan, S. P. Review on Domestic Teaching of Second Language Chunks. *Foreign Language in China*, 2008 (4).
- [4] Ding, Y. R. & Qi, Y. Use of Formulaic Language as a Predictor of L2 Oral and Written Performance. Journal of PLA University of Foreign Languages, 2005(3).
- [5] Ding, Y. R. Second Language Acquisition for English Majors. Shanghai Foreign Language Education Press, 2007.

#### INFORMATION OF THE AUTHORS

Name: Lihua Ma, Title: Associate Professor, Employers: Zhejiang Ocean University; Yi Li China Foreign Affairs University Undergraduate Research Area: Linguistics, Applied Linguistics, Business English Writing, Second Language Acquisition Theory, English Listening Address: 41<sup>st</sup> Building, Beitang Community, Zhujiajian, Zhoushan, Zhejiang Province Post Code: 316100 Mobile: 15957078860 Email: zsmalihua@126,com