Nick Riches

# What is formulaic language?

Formulaic language is language which, though it looks as if it is comprised of multiple units, appears to have been stored a single unit. Examples are

1. Word combinations, e.g. *he has high hopes* (not “tall” hopes), and *he told a tall story* (not a “high” story)
2. Phrases, e.g. *she’s larger than life*, (ADJECTIVE PHRASE), *he’s down in the dumps* (PREPOSITIONAL PHRASE), *For crying out loud* (ADVERBIAL PHRASE)
3. Entire sentences, e.g. *What’s a nice girl like you doing in a place like this?*, *Who’d a thunk it?*

The “roll off the tongue”. They appear to have been stored as a single “chunk” (unanalysed, or partially analysed unit).

# Collocation

A collocation is a group of words which tends to co-occur within the same fragment of text. Typically, the term describes pairs of words which tend to co-occur next to each other. Here are some examples:

1. ADJ + NOUN: We have **high hopes**, but we tell a **tall story**
2. VERB + NOUN: We **do the washing up** and the laundry, but we **make the bed**.

However, many collocations contain more than two words, e.g.

1. She’s **down in the dumps**
2. He’s **larger than life**

Some collocations can also contain gaps for arguments, e.g.

1. Jason **read** Angela **the riot act**
2. His boss **showed** him **the door**

Some can also be entire sentences, e.g.

1. That’s no way to behave.
2. I couldn’t agree more.
3. What’s a nice girl like you doing in a place like this.

We can detect collocations by applying complex maths. Roughly speaking, if two words exist within a collocation, then the chance of the words co-occurring is greater than the combined frequencies of the individual occurrences. There is a formula to determine this!

# Formulaic language and phonology

Formulaic chunks do not just contain lexical information. They also contain phonological information, e.g.

1. Could you **be** any more serious? (unusual stress placement signals sarcastic tone)
2. They’re **rolling** in money (unusual stress placement signals sarcastic tone)
3. What do you want to do: “wojawonadu” (phonemic reduction processes, in highly frequent chunk, e.g. place assimiliation of /d/ and /j/)

Phonological reduction appears to be a process which is used to increase efficiency (we reduce frequent chunks so that they may be processed more quickly). But there is a trade-off between efficiency and comprehesibility. If we were to start reducing all sounds in all sentences we’d quickly become unintelligible. Therefore this only occurs in highly frequent chunks.

# Idioms

Idioms are formulaic units with non-compositional semantics. i.e. the meaning is more than the sum of the parts. A classic example is, *last year, my grandmother kicked the bucket*.

Idioms are often multiword phrases, e.g. *kick the bucket*, *have your chips*. English phrasal verbs (verb + particle) are often idiomatic, e.g.

1. Batman **pulled off** the villain’s mask (literal meaning)
2. She **pulled off** a first in the exam (idiomatic - this means something like she attained unexpected success)

However, some syntactic constructions which span entire sentences can also be described as idiomatic:

1. *She danced the night away*

* S + V + Od + Noun Phrase denoting time period + “away”
* MEANING: the activity is enjoyable so time passes quickly.

1. *The more I see of humanity, the more I love my dog*

* “The more” + CLAUSE, “the more” + CLAUSE
* MEANING: There is a linear relationship between the first and second proposition.

Idioms can vary in their transparency and fomulaicity. If you were a foreign language learner of English you would probably be confused when hearing *kick the bucket* for the first time. However, you could probably guess what *spill the beans* (reveal a secret means). If an idiom is highly formulaic, it cannot be manipulated in any way. For example, we can say *grab the nettle*, but not \*\*grab the cactus*. By contrast,* spill the secret\* sounds odd, but just about acceptbale.

# Formulaic language and processing

Because formulaic units are stored as a whole, we are relatively fast to process them.

# Formulaic language and the syntax-lexicon divide

According to many accounts of language, e.g. Chomksyan linguistics, there is a categorical divide between syntax and the lexicon. Syntax is the system of rules used to make sentences. The lexicon is the repository of arbitrary / idiosyncratic information. For example, there is nothing about the combination of “d” “o” and “g” which would allow a non-English speaker to figure out its meaning. The mapping between form and meaning is therefore arbitrary. Syntactic rules take words from the lexicon as their input, and their output is a well-formed sentence.

However, sentence-level idioms (sometimes called “constructional idioms”) such as *she danced the night away*, defy this neat dichotomy. They are partially syntactic, e.g. they alow aspectual modification and argument substitution - *she is jiving the whole afternoon away*. However, they also exhibit properties of the lexicon, e.g. they have an arbitrary (non-compositional meaning). So where do they belong?

There are alternative models of language which do not presuppose strict divide between syntax and the lexicon. One of these is called “construction grammar”, and argues that all linguistic knowledge, whether lexical or syntactic, belongs within a single system.

# Clinical relevance

Some language-impaired individuals exhibit a strong reliance on formulaic language, e.g. individuals with fluent aphasia, individuals with Williams Syndrome, and also individuals with Autism Spectrum Disorders (who use echolalia).

It could be argued that there are two language systems: a rule-based system for creating new sentences, and a formulaic system which depends on chunks. Competent language use depends on the ability to master BOTH systems.

Alison Wray, one of the key researchers on fomulaic language, argues as follows:

“Without the rule-based system, language would be limited in repertoire, cliched, and, whilst suitable for certain types of interaction, lacking imagination and novelty. In contrast, with only a rule-based system, language would sound pedantic, unidiomatic and pedestrian. (Wray, 1998:64)”

# Bibliography

Aphasia, Alzheimmers, Autism - Wray, A. (2009). 11 Formulaic Sequences and Language Disorder. The handbook of clinical linguistics, 184 (available on Google Scholar).

Tomasello, M. (1998). The Return of Constructions. *Journal of Child Language*, *25*, 431–442.