

Diana Van Lancker Sittis<sup>a,b</sup>, Katherine M. Kougentakis<sup>b</sup>, Krista Cameron<sup>b</sup>, Carolyn Falconer<sup>a</sup> and John J. Sittis<sup>b,c</sup>

## **“Down with \_\_\_\_”: The linguistic schema as intermediary between formulaic and novel expressions**

**Abstract:** A special instance of formulaic expression is the linguistic schema: most of the expression is fixed, with one or more slots left open for insertion of novel words, such as *I can \_\_\_\_ with one hand tied behind my back*. This study aimed to determine whether native speakers demonstrate knowledge of the fixed portions of the schemata and flexibility for the open slots. A survey was designed with four sets of stimuli: formulaic expressions, novel sentences, schemata with their open slots left blank (schemata-novel), and schemata with open slots (schemata-fixed) in the fixed portions. Significantly fewer unique words appeared for the formulas and schemata-fixed stimuli, while more unique words were produced for novel and schematic-novel exemplars. These results, the variable provenance of schemata, and their proliferation throughout society suggest that linguistic schemata are bona fide constituents in a dual process model of language competence, holding a position intermediate between formulaic and novel language abilities.

**Keywords:** formulaic expressions; linguistic schemata; dual process model.

---

**Correspondence address:** diana.sittis@nyu.edu

## **1 Introduction**

Modern studies of formulaic language constitute a cornucopia of themes, touching on a broad range of topics within social communication: first and second language learning (Kempler et al. 1999; Locke 2006; Lieven 2007; Perkins 1999), pragmatics of conversation (Tannen 1989), theoretical linguistics and evolution of language (Code 2005), corpus studies and sociology of language (Moon 1998; Kuiper 2009), psycholinguistics (Clark 1970; Cutler 1982; Sprenger 2003; Cutting and Bock 1997) and the specific effects of neurological disease (Dieguez and Bogousslavsky 2007; Van Lancker Sittis 2006; Van Lancker Sittis and Postman 2006). Despite this considerable scholarly activity, controversies remain about how to position formulaic language in a general model of language use. These

discussions give rise to numerous proposed solutions to classifying the various types (Barkema 1996; Wray 2002). The heterogeneous array has proven difficult to tame into a convincing typology, leading to recommendations of continua to represent formulaic and novel language (Van Lancker 1988; Penttillä 2010). The strikingly pervasive presence of formulaicity in human language use has led to the suggestion that truly novel expressions are quite at the periphery (Fillmore 2007).

More basically, questions can be usefully addressed about the essential differences between novel or newly created language on the one hand, and formulaic expressions, or what we refer to as *formulemes*, on the other. *Formulemes*, by definition, have stereotyped form, conventionalized meanings, and close connection with social variables; these properties about each *formuleme* are known to the language user. Often overlooked in these discussions is this latter fact; language users know the *formulemes* – that is, they recognize a great many of them as stored and processed as a whole, while, by definition and by empirical demonstration, newly generated sentences are not recognized or handled in this way (Jackendoff 1995; Van Lancker Sidtis and Rallon 2004). *Formulemes* themselves can be freely manipulated by grammatical operations, and yet their underlying canonical shape remains constant in ordinary language use. Another way of saying this is to refer to base form (Naciscione 2006); the phraseological unit can be artistically manipulated while being sustained in form, leading to greater coherence in the entire text. The base forms “generate stable expectations” which “can be tampered with” (Kuiper, 2007: 94). The perceiver must be able to recognize the difference between the base form and its variant; this has been referred to as the “recoverability condition” (Kuiper, 2007: 96). Thus the various grammatical manipulations have as their constraint only that the known phraseological unit be recoverable by the listener.

A recent example of *formuleme* manipulation, provided by a student who was unsuccessfully pitching drink coasters into a garbage container, illustrates this point: one of the revelers said *You can lead a coaster to water, but you can't make it drink* (hilarity ensued). In this example, two words (*horse*, *him*) of a well-known proverb (*You can lead a horse to water, but you can't make him drink*) are changed for comic effect. This unmistakably exemplifies manipulating a fixed expression for humorous purposes while retaining its identity to the listeners.

The flexibility of formulaic expressions, especially idioms, has been the subject of numerous studies, reviewed elsewhere, that have attempted to classify these expressions according to principles of semantic opaqueness and/or decompositionality (Van Lancker Sidtis 2006; Van Lancker Sidtis 2010). While interesting claims have arisen from these studies, they remain controversial, and the themes they invoke are not pertinent to the study reported here.

A viable approach to modeling the structural properties of novel and formulaic language is to view expression-types as occurring at two extremes, from fixed,

representing the known formuleme (which, of course, is subject to manipulation), to novel, for which lexical choices are dependent on grammatical constraints and creative lexical selection only. In this conceptualization, one encounters an interesting, intermediate type of linguistic object: the linguistic schema, first described by John Lyons (1968: 177–178). Lyons' example was *Down with \_\_\_\_*. In psychology, a schema is defined as “cognitive framework or concept that helps organize and interpret information.”<sup>1</sup> Lyons' use of the term *schema* in language use refers to a linguistic framework that, in its essential format, officially allows for insertion of novel material.

Our topic in this study is the linguistic schema. Schemata carry the characteristics of formulaic expressions: they have basic canonical form (usually with distinctive intonation contour and often with a signature voice quality and articulatory detail); they utilize specialized connotational and social meanings, conveying attitudinal nuances; and they are known with these properties (form and meaning) to the native speaker. But schemata possess an additional versatility. A schema differs from the typical formuleme in having, as part of its phraseological base, one or more free open slots.<sup>2</sup> Schemata mandate that novel lexical items be inserted into one or more slots. The open slot(s) provide(s) the thematic point of the utterance. For example, *I'm not a \_\_\_\_ person* (Figure 1) is used to express a

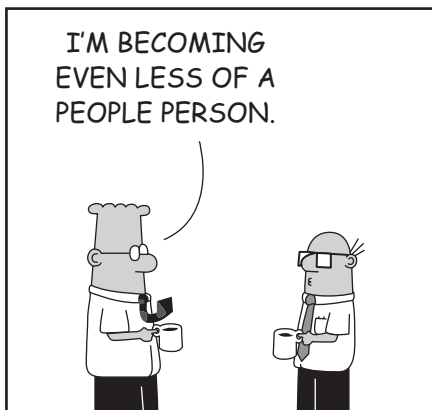


Figure 1. Dilbert uses a schema<sup>3</sup>

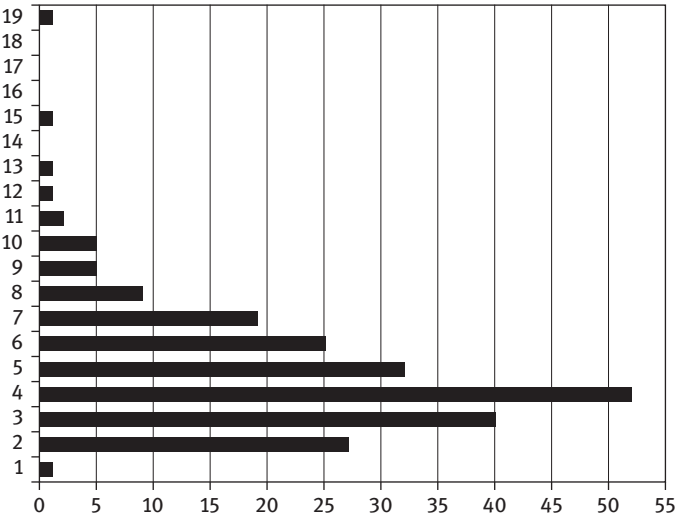
1 Schema theory as a psychological concept was elaborated by Bartlett (1932).

2 Verb plus particle or complement phrases with the object position as the missing slot, such as “Take \_\_\_\_ to task,” are not included in this grouping (Kuiper 2007).

3 Dilbert © 2011 Scott Adams. Used by permission of Universal Uclick. All rights reserved.

strong personal preference that is asserted to form part of one’s identity, as in *I’m not a morning person, I’m not a city person, I’m not a war movie person, I’m not a touchie-feelie person*.

The expression *That was a \_\_\_\_\_ and a half* signals strong amazement, usually with approval, about whatever referent is in the slot: e.g., *That was a movie and a half. If you had my \_\_\_\_\_, you’d be \_\_\_\_\_, too* carries nuances of self-pity, self-congratulation, and high complaint: *If you had my job (wife, house, car), you’d be (drinking, overeating, crying), too*. The items may be phrases, as in *A walking \_\_\_\_\_*, which communicates concern, intensity, and/or inevitability, as in *A walking idiot, wonder, disaster, genius, time bomb*; or sentences: *I wouldn’t be caught dead \_\_\_\_\_*, which indicates strong revulsion about the topic: e.g., *I wouldn’t be caught dead (out late at night, in a dress, eating sushi, on a plane)*. Schemata can be long: *I may not know much about \_\_\_\_\_, but I know what I like*. They can have two or more open slots: *Changing \_\_\_\_\_, one \_\_\_\_\_ at a time*; here the nuance is positive and uplifting, as in *Changing the (world, cities, musical life), one (person, citizen, conductor) at a time*. Another example of multiple open slots is *You can take the \_\_\_\_\_ out of the \_\_\_\_\_, but you can’t take the \_\_\_\_\_ out of the \_\_\_\_\_*. This expression communicates the stalwart nature of some personality traits: *You can take the (boy, man, soldier) out of the (country, office, warzone), but you can’t take the (country, office, warzone) out of the (boy, man, soldier)*. For some schemata



**Figure 2.** Frequency chart of schemata displayed by numbers of words. Numbers of schemata are represented on the ordinate while the numbers of words making up the utterances are on the abscissa

with more than one open slot, the inserted word is repeated, as in *What happens in \_\_\_\_\_ stays in \_\_\_\_\_*; for most multi-slot schemata, different words are inserted, as in *You can say hello to \_\_\_\_\_, goodbye to \_\_\_\_\_*. It is likely that schemata are to be found in other languages; we look to native speakers to discover and report them. Some German examples are *Es war einmal ein \_\_\_\_\_; ran an \_\_\_\_\_ (die Arbeit, ans Werk, den Speck); Hoch die \_\_\_\_\_ (Tassen, Gläser)*. A current listing of American English (See Appendix I) schemata recorded from live communicative contexts reveals a range of word-count lengths from 1–19 words, with a mean utterance length of 8.3 words and an average of 1.31 open slots (Figure 2).

Schemata have the advantage of communicating special nuances and connotations (having a meaning that is more than the sum of the parts, as is the case with formulaic expressions), while allowing for this meaning constellation to be applied to very disparate phenomena. For example, the schema *The mother of all \_\_\_\_\_* carries dense connotative nuances of extreme and over-the-top characteristics, and these nuances can be utilized to communicate an attitude about, for example, the *Airbus 380 airplane, an advertising campaign, a marathon through the Sierras, a car race, a brutal attack, a climb, a building*, and so on. Schemata carry a meaning independent of how the slots are filled. As mentioned above, this independent meaning, which is more than the sum of the words put together (the well-known property of a fixed expression) can then be *conferred* onto the inserted word, which constitutes the topic of the utterance. The lead example, *Down with \_\_\_\_\_*, is an expression with an intense nuance connoting rebellion, strong emotion, turmoil, even violence, and therefore, whatever word (gerund, mass or count common noun, abstract or concrete noun, pronoun, proper noun) is inserted will take on these connotations. Thus, again, the schema combines the characteristics of a speech formula with the flexibility of a novel phrase. Finally, schemata differ significantly from formulemes in the following way; formulemes *allow for* flexible lexical insertion, while for schemata, creative lexical insertion is *mandatory*, because a constituent slot is empty.

## 2 Purpose of study

The purpose of this study is to examine users' knowledge of formulemes and schemata as contrasted with their performance on novel expressions. The questions were:

1. Do native speakers endorse the stereotyped forms of formulaic expression by agreeing on their lexical content? This portion of the study attempts to replicate findings by Van Lancker Sidtis and Rallon (2004), who performed

a survey on formulaic expressions from the screenplay *Some like it hot* (Wilder and Diamond 1959), this time using naturalistic stimuli;

2. Do native speakers endorse the lexical content of the fixed portions of schemata? This extends the findings from the previous study to a related semi-fixed expression, the schema; and
3. To what extent, in contrast, are native speakers able to utilize the creative capacities of schemata, as available in the open slots, and of novel expressions?

Our interest was to obtain objective measures in addressing these questions. It was predicted that subjects' responses in blanks within formulaic expressions and the fixed portion of schemata would be relatively uniform; that is, the responses would form a relatively homogeneous set of lexical items. In contrast, responses written into the blanks in novel sentences and the novel open slots in schemata were predicted to form a more diverse set of lexical items.

### 3 Method

**Stimuli:** We chose 40 formulaic expressions (e.g., *It was a blessing in disguise*), 40 novel sentences (*The two of you are soaked*), and 80 schemata from previously established lists, divided into two subsets of 40 each (see below) for the survey. Schemata had been recorded from conversation and the media over a period of several years and accumulated into a working list (see Appendix I); from this list, only those schemata with one open slot were selected for the survey. Formulaic expressions, made up of conversational speech formulas, idioms, and proverbs, were taken from available published dictionaries and lists, and vetted in previous surveys administered to native speakers of American English. Our criteria for including formulaic expressions were ordinariness, naturalness, and familiarity by native speakers with these expressions. Novel sentences were generated with appropriate English grammar using the criteria of naturalness, plausible meaning, and high- to mid-range lexical frequencies. Each set of phrases was balanced to match on number of words. The 160 test items were then randomized and compiled onto an answer sheet. Each test item had a blank (cloze procedure) for participants to fill in the missing word (see Appendix II).

Four groups of stimuli were utilized for the slot-filler task (Table 1). These are referred to in this study as formulas (standard formulaic sentences), novel sentences (newly created sentences), schemata-fixed (schemata with a open slot in the fixed portion of the expression) and schemata-novel (schemata with an open slot where the novel word belongs). In the formulaic expressions and the novel

40 Formulaic Expressions	40 Novel Expressions	40 Schemata-novel	40 Schemata-fixed
Open slot	Open slot	Natural open slot	Natural open slot filled in, open slot in fixed portion
Throw in the _____	I'm allergic to _____	He eats and breathes _____	I can do <u>math</u> in my _____

**Table 1:** Format of survey protocol. Examples are given in the third row. For the first column, “towel” is the expected word in the formulaic expression. No expectations are made for examples in columns two and three, as these slots take novel lexical items. “Sleep” is an expected response for the fixed portion of the schema in the 4th column.

sentences, the blank (open slot) occurred anywhere in the sentence. For the 40 schemata-novel, a natural open slot was provided (*He eats and breathes \_\_\_\_\_*). This category was intended to elicit novel responses from subjects, thus probing their creativity in the natural open slot position. In the second set of 40 schemata, the schemata-fixed set, items had blanks in the fixed portion of the utterance and a novel word was included in the natural open slot: *You can take your report and \_\_\_\_\_ it*, where “shove” belongs in the fixed portion of the schema, and “report” is the novel word in the schema. That is, in this set of 40 schemata-fixed items, a novel word was provided in the natural open slot position, and an open slot was created in the fixed portion of the schema. This set of schemata was intended to probe subjects’ knowledge of the schema itself. One open slot was chosen for each item. To the extent possible, open slots were matched for grammatical form across sets and placed equally often toward the beginning, middle or the end of the items.

**Subjects:** Ten native speakers of English with normal vision completed the survey after signing a consent form according to IRB guidelines. The participants had an average age of 22.4 with a range of 20–28 years. Their average number of years of education was 16.8 years with a range of 16–22 years. All were born and educated in the United States and eight reported speaking American English since infancy; two spoke English since preschool. None had history of neurological or psychiatric disorder.

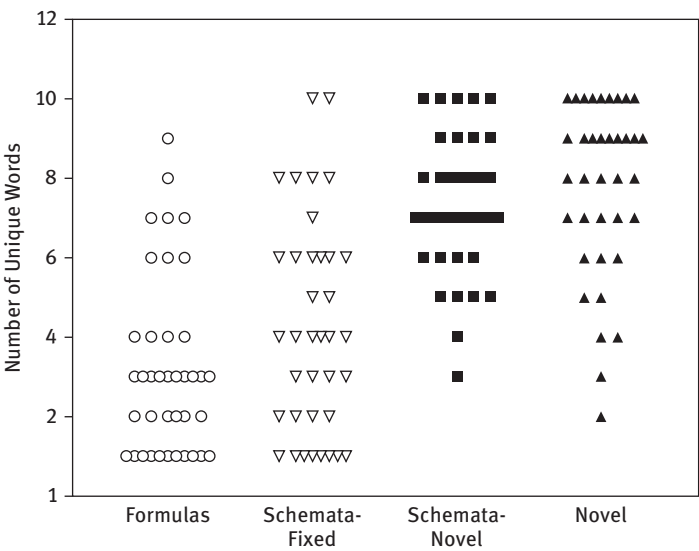
**Procedure:** Subjects were briefed generally that the purpose of the study was to learn more about different kinds of expressions. After completing the written informed consent form, subjects were given a survey form and asked to write down one word for each missing word (blank or open slot). Instructions to subjects were:

*Thank you for agreeing to take our word survey. Please fill in the blanks below with a single word. Some of the items will seem “familiar,” and some not. This is to be expected. Just write in the word that seems to work best in the item.*

For scoring, all the responses were catalogued. To test the hypotheses that subjects have knowledge of formulas and the fixed portions of schemata, in contrast to their creative responses for slots in novel sentences and the natural open slots of schemata, numbers of unique word types produced in each test category were calculated.

## 4 Results

As predicted, more unique words were generated for the novel sentence and schemata-novel slots than for the formulas or the schemata-fixed slots. In Figure 3, the number of unique words across the ten subjects is on the ordinate and data points on the graph show the distribution across utterances in each grouping: formulas, schemata-fixed, schemata-novel, and novel. For the first two categories (formulas and schemata-fixed), responses to a large number of stimuli

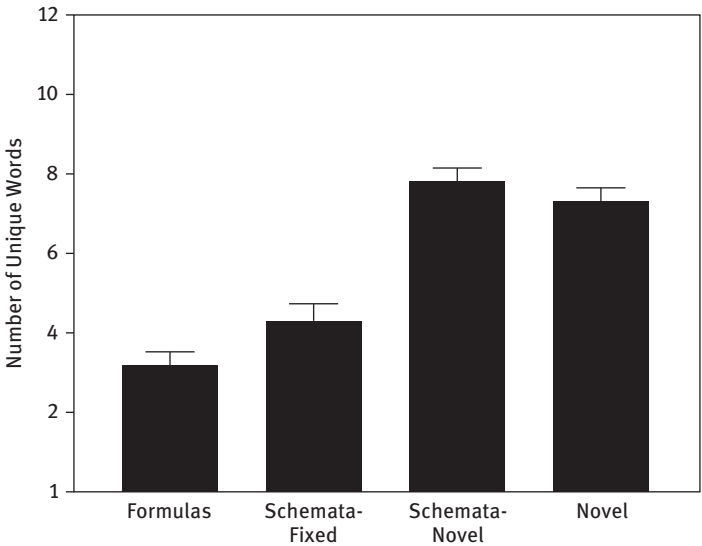


**Figure 3.** Distribution of individual survey items showing numbers of unique words in each stimulus set



consisted of only 1–4 unique words. For formulas, 11 of the 40 utterances contained the same word; 6 of the utterances received 2 unique words, and 10 utterances showed a concordance of 10 words. The schemata-fixed also showed relative concordance, with the majority of the responses toward the lower portion of the graph, representing fewer unique words. In contrast, for schemata-novel and novel sentences, the bulk of the responses per utterance consisted of 7–10 unique words, appearing toward the top of the graph, indicating a greater number of unique words. The differences for each stimulus set are graphically displayed in Figure 4, where the mean number of unique words (plus standard error of the mean) for each stimulus category is given. Numerical values of the medians for each condition are above the error bars.

The numbers of unique words in each sentence type were compared using sentences as observations for within-subject comparisons across conditions. There were significantly more unique words in the novel sentences (mean  $\pm$  SD; median:  $7.4 \pm 1.7$ ; 7.0) than in the formulas ( $3.2 \pm 2.2$ ; 3.0) [Wilcoxon Signed Ranks Test:  $z = -5.03$ ;  $p < 0.001$ ]. Similarly, there were significantly more unique words in the schemata with novel words ( $7.8 \pm 2.1$ ; 8.5) than in the fixed schemata ( $4.3 \pm 2.6$ ; 4.0) [ $z = -4.38$ ;  $p < 0.001$ ]. The difference between unique words in formulas and fixed schemata approached significance [ $z = -1.86$ ;  $p = 0.06$ ], but there was no difference in the number of unique words between the novel sentences and the schemata with novel words.



**Figure 4.** Mean numbers of unique items for each stimulus set. Median values are given above the error bars



formulaic portions of schemata, and they entered a range of novel words in novel expressions and the open slots of schemata.

In a dual-process model of language, schemata occupy an intermediate position between formulaic and novel expressions. In this depiction, schemata enjoy the interplay of two processing modes, novel and formulaic. For schemata, a known unitary form mandates one or more specific flexible lexical choices, and thus, without distorting the fixed expression, allows naturally for highly flexible application to novel meanings. This study has shown that subjects perform well and with implicit knowledge of this large range choices available to them in schemata. In our survey, subjects endorsed the stereotyped forms of formulaic expressions and the fixed portions of schemata by agreeing in their lexical choices; and, as expected, they revealed diversity of lexical choice in the natural open slots of schemata and in the open slots of novel sentences.

Schemata allow speakers to benefit from the conversational advantages of formulaic expressions, which include establishing bonding by using a mutually known expression, exploiting the humorous nuance, conveying an indirect, non-literal meaning, and often introducing a playful note (Tannen 1989); at the same time, the availability of the open slot allows for applying the phrase specifically and distinctly – and literally – to the topic at hand.

A model of language use that accommodates these three utterance types (formulaic expression, schema, and novel sentence) is the dual process model of language use, which proposes two modes of processing, variously designated by speech scientists as analytic and holistic, novel and idiomatic or formulaic, and as governed by principles of open choice and idiom (Fillmore 1979; Erman and Warren 2000; Van Lancker 2004; Wray and Perkins, 2000) (see Figure 5). It is well known that human language allows for potentially infinitely new combinations of words governed by grammatical rules. In addition, and not less important, formulaic language has a vivid presence in all of human verbal communication. Schemata illustrate the dual mode process in linguistic performance, in which these two distinct modes coexist in continuous interplay.

## **6 Qualitative analysis of schemata: provenance and status in language competence**

Perusing the list of schemata in Appendix I provokes questions about their origins. Provenance is highly varied and may not be fully knowable in most cases. A full description calls for a separate study. Many suggestive derivations can be found on the internet, with the expected variable reliability. Many, such as \_\_\_\_

*and counting*, appear to be so frequently encountered as to not be traceable to a source. Some schemata come from titles or lines from film or literature that became popular: \_\_\_\_\_ *are people, too* may have started with as the name of television series that ran from 1978–1982, *Kids are people, too*, and *One in a \_\_\_\_\_* from another series *One in a million* (1980); *shut up and \_\_\_\_\_* from *Shut up and dance*, Aerosmith lyrics; *Tell it to the \_\_\_\_\_* echoes the title *Tell it to the marines* (1927); Another was popularized in a song lyric *What part of no don't you understand* by Lorrie Morgan, country music singer, in 1992; and many people know that *Yes, Virginia, there is a \_\_\_\_\_* originated in an editorial appearing in September 21, 1897 of the *New York Sun*. Another potent lyric was *You can take your job and shove it* from a hit single by country music singer David Allan Coe. Others that were dormant may have been brought into awareness by popular television vehicles, as \_\_\_\_\_ *is not just a pretty face*, which is associated with the Mary Tyler Moore television show (1970–1977). In many cases it is not clear whether the citation is the source or merely the more frequent transmission of the expression. Some have a first instance that is fairly certain, such as *Ask not what \_\_\_\_\_ can do for you, ask what you can do for \_\_\_\_\_*, spoken about one's country by JFK on January 20, 1961. *Too \_\_\_\_\_ by half* may have started as *Too clever by half* and then morphed to allow novel words in the adjective slot. According to a report, *if not \_\_\_\_\_, \_\_\_\_\_?* Arose from a title of a novel (Primo Levi, 1984) which itself is taken from a well-known rabbinical saying, *If not now, when?* *In \_\_\_\_\_ we trust* likely originated in the USA motto. *It was a \_\_\_\_\_ from hell* allegedly originates in an 1888 letter by a man claiming to be Jack the Ripper, an unidentified serial killer in London. *Keep your eyes on the \_\_\_\_\_* may have come from, or been popularized by, a folk song in the 1950s. *The mother of all \_\_\_\_\_* could have sprung from Saddam Hussein in a 1991 speech, referring to a war. *One man's \_\_\_\_\_ is another man's \_\_\_\_\_* is probably a schematic morphing of the proverb *One man's meat is another man's poison*. *The \_\_\_\_\_ is the enemy of the \_\_\_\_\_* may have first been generated by the 18th century French writer, Voltaire, in 1772.

The variety of these sources and the vagueness of their origins are a further testimony to the fact that the linguistic schema holds an honored place in the native speaker's competence. A viral productivity of this constituent of language can be seen in various sets of bumper stickers, such as *Honk if you \_\_\_\_\_*, and *If you can read this, \_\_\_\_\_; \_\_\_\_\_ do it \_\_\_\_\_*; e.g., *plumbers do it deeper*, and tee shirt sayings, such as *Save the \_\_\_\_\_*; I ♥ \_\_\_\_\_; *with a \_\_\_\_\_ like this, who needs \_\_\_\_\_? I'm the Michael Jordan of \_\_\_\_\_* (asserting excellence in a field, whatever is inserted in the blank). Language users know schemata (and that a slot is open for their use) in the way that they know formulaic expressions. Native speakers know the schema's stereotyped form (including its prosody), conventional meaning, and the guidelines of pragmatic use. Like formulemes,

schemata are likely to enter quickly into the speaker's repertory (Reuterskiold and Sidtis 2012) due to their unique status with respect to meaning and form. Thus their specific provenance is not germane to the process of acquiring and using them. Of interest to students of language is the fact that linguistic schemata, as modified versions of formulaic expressions, form a natural part of human language competence.

A limitation of the current study arises from the number of subjects tested. Nonetheless, the differences in performance on utterance types were statistically significant, reflecting the robustness of these effects. In addition, by design, all the subjects belonged to a younger age group. We are pursuing a study designed to replicate these findings using larger groups of subjects from two different age groups representing different demographics in the form of language users separated by at least one generation.

## 7 Clinical relevance

Studies show that persons with language disorder, or aphasia, following left hemisphere damage utilize significantly more formulaic language in their conversational speech, probably due to a demonstrated contribution of the right hemisphere in processing of formulaic expressions (Van Lancker Sidtis and Postman 2006; Sidtis, Canterucci, and Katsnelson 2009). Use of schemata in language rehabilitation for persons with aphasia could advantageously exploit a preserved knowledge of formulaic expressions while offering recursive opportunities to access novel lexical material. Further, schemata have special qualities of familiarity and, often, clever and provocative nuances. Our experience is that people smile and nod in recognition and amusement when hearing any of these schemata. This added entertainment value may be beneficial to new learning in the therapy setting.

a: *New York University*  
b: *Nathan Kline Institute for Psychiatric Research*  
c: *New York Langone Medical Center*

## Acknowledgements

This work was supported in part by NIH R01 DC007658. Contributions at various stages of this study were provided by Elana Winters, Lisa Bonura, Silke Mohr and our de-identified subjects.

## References

- Barkema, Henk. 1996. Idiomaticity and terminology: A multi-dimensional descriptive model. *Studia Linguistica* 50(2). 125–160.
- Bartlett, Frederic. 1932. *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.
- Clark, Herbert H. 1970. Word associations and linguistic theory. In John Lyons (ed.), *New horizons in linguistics*, 271–286. Baltimore: Penguin Books.
- Code, Chris. 2005. First in, last out? The evolution of aphasic lexical speech automatisms to agrammatism and the evolution of human communication. *Interaction Studies* 6. 311–334.
- Cutler, Anne. 1982. Idioms: ‘the colder the older.’ *Linguistic Inquiry* 13. 3178–3320.
- Cutting, John C. & J. Kathryn Bock. 1997. ‘That’s the way the cookie bounces’: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition* 25. 57–71.
- Dieguez, Sebastian & Julien Bogousslavsky. 2007. Baudelaire’s aphasia: From poetry to cursing. In Julien Bogousslavsky & M. G. Hennerici (eds.), *Neurological disorders in famous artists, Part 2*, Vol. 22. *Frontiers of Neurology and Neuroscience*, 121–149. Basel: Karger.
- Erman, Britt. & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Interdisciplinary Journal for the Study of Discourse* 20(1). 29–62.
- Fillmore, Charles. 1979. On fluency. In Charles. J. Fillmore, Daniel Kempler & William S-Y Wang (eds.), *Individual differences in language ability and language behavior*, 85–102. London: Academic Press.
- Fillmore, Charles. 2007. Personal communication. The Symposium for Formulaic Language, at UWM, in Milwaukee, Wisconsin, April.
- Jackendoff, Ray. 1995. The boundaries of the lexicon. In Martin Everaert, Erik-Jan van der Linden, Andre Schenk & Rob Schreuder (eds.), *Idioms: Structural and psychological perspectives*, 133–166. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kempler, Daniel, Diana Van Lancker, Virginia Marchman & Elizabeth Bates. 1999. Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology* 15(3). 327–349.
- Kuiper, Koenraad. 2009. *Formulaic genres*. Basingstoke: Palgrave Macmillan.
- Kuiper, Koenraad. 2007. Cathy Wilcox meets the phrasal lexicon: Creative deformation of phrasal lexical items for humorous effect. In Judith Munat (ed.), *Lexical creativity, texts and contexts*, 93–112. Amsterdam: John Benjamins.
- Lieven, Elena. 2007. Producing multiword utterances. In Barbara Kelly & Eve Vivienne Clark (eds.) *Constructions in acquisition*. Stanford, CA: CSLI Publications.
- Locke, John L. 1993. *The child’s path to spoken language*. Cambridge, MA: Harvard University Press.
- Lyons, John. 1968. *Introduction to theoretical linguistics*. Cambridge, U.K.: Cambridge University Press.
- Moon, Rosamund E. 1998. *Fixed expressions and text: A study of the distribution and textual behaviour of fixed expressions in English*. Oxford: Clarendon Press.
- Naciscione, Anita. 2006. Sustainability of phraseological image in discourse. *International Journal of English Studies* 6(1). 43–56.

- Penttilä, Esa. 2010. A prototype-based taxonomy of idiomatic expressions. In Gitte Kristiansen & Francisco J. Ruiz de Mendoza Ibáñez (eds.), *Applications of Cognitive Linguistics* 14. *Cognitive linguistics in action*, E. Tabakowska, M. Choiński & L. Wiraszka, (eds.), 145–162. Berlin: Walter de Gruyter.
- Perkins, Michael R. 1999. Productivity and formulaicity in language development. In M. Garman, C. Letts, B. Richards, C. Schellletter & S. Edwards (eds.), *Issues in normal and disordered child language: From phonology to narrative*. Special Issue of The New Bulmershe Papers, 51–67. Reading: University of Reading.
- Reuterskiöld, Christina & Diana Sidtis. 2012. Incidental learning of formulaic expressions. *Child Language and Therapy*, in Press.
- Sidtis, Diana, Gina Canterucci & Dora Katsnelson. 2009. Effects of neurological damage on production of formulaic language. *Clinical Linguistics and Phonetics* 23(15). 270–284.
- Sprenger, Simone A. 2003. Fixed expressions and the production of idioms. *MPI Series in Psycholinguistics*, Vol. 21. Wageningen/The Netherlands: Ponsen and Looijen BV.
- Tannen, Deborah. 1989. *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge University Press, Cambridge.
- Van Lancker Sidtis, Diana & Whitney A. Postman. 2006. Formulaic expressions in spontaneous speech of left- and right-hemisphere damaged subjects. *Aphasiology* 20(5). 411–426.
- Van Lancker Sidtis, Diana & Gail Rallou. 2004. Incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and Communication* 24. 207–240.
- Van Lancker Sidtis, Diana. 2004. When novel sentences spoken or heard for the first time in the history of the universe are not enough: Toward a dual-process model of language. *International Journal of Language and Communication Disorders* 39(1). 1–44.
- Van Lancker Sidtis, Diana. 2006. Where in the brain is nonliteral language? *Metaphor and Symbol* 21(4). 213–244.
- Van Lancker, Diana. 1988. Nonpropositional speech: Neurolinguistic studies. In Andrew Ellis (ed.), *Progress in the psychology of language* (Vol. III), 49–118. Hillsdale, NJ: Lawrence Erlbaum.
- Van Lancker Sidtis, Diana. 2010. Formulaic and novel expressions in mind and brain: Empirical studies and a dual process model of language competence. In Jackie Guendouzi, Filip Loncke & Mandy Williams (eds.), *The handbook of psycholinguistic & cognitive processes: Perspectives in communication disorders*, 247–272. London: Taylor & Francis.
- Wilder, Billy, and I.A.L. Diamond, 1959. *Some like it hot*. Screenplay reprinted in S. Thomas (ed.), *Best American Screenplays* 2. (1st ed.), 1990, 80–146. New York: Crown Publishers.
- Wray, Alison & Mick Perkins. 2000. The functions of formulaic language: An integrated model. *Language and Communication* 20, 1–28.
- Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

# Appendices

## Appendix I.

\_\_\_\_\_ 'sville  
 \_\_\_\_\_ city  
 \_\_\_\_\_ days  
 \_\_\_\_\_ fool.  
 \_\_\_\_\_ galore  
 \_\_\_\_\_ happy  
 \_\_\_\_\_ hunting.  
 \_\_\_\_\_ much?  
 \_\_\_\_\_ power  
 \_\_\_\_\_ shm \_\_\_\_\_  
 \_\_\_\_\_ thinking  
 \_\_\_\_\_ this.  
 \_\_\_\_\_ time  
 \_\_\_\_\_ wars  
 \_\_\_\_\_ crazy.  
 \_\_\_\_\_ wars  
 Dead \_\_\_\_\_  
 Everything \_\_\_\_\_  
 Fuck \_\_\_\_\_  
 Get \_\_\_\_\_  
 Go, \_\_\_\_\_!  
 Got \_\_\_\_\_?  
 nice \_\_\_\_\_  
 Perfect \_\_\_\_\_  
 Screw \_\_\_\_\_  
 That's \_\_\_\_\_  
 Think \_\_\_\_\_  
 You: \_\_\_\_\_  
 \_\_\_\_\_ and counting  
 \_\_\_\_\_ and proud  
 \_\_\_\_\_ are us  
 \_\_\_\_\_ be us  
 \_\_\_\_\_ is overrated.  
 \_\_\_\_\_ loves \_\_\_\_\_ (written)  
 \_\_\_\_\_ to \_\_\_\_\_ (A, Z Mon, Fri, soup, nuts)  
 \_\_\_\_\_ to death



\_\_\_\_\_ under fire

A \_\_\_\_\_'s \_\_\_\_\_ (word repeated)

A royal \_\_\_\_\_

A walking \_\_\_\_\_

All things \_\_\_\_\_

Call me \_\_\_\_\_

Color me \_\_\_\_\_

Do not \_\_\_\_\_

Down with \_\_\_\_\_

For the \_\_\_\_\_

Giant among \_\_\_\_\_

Go and \_\_\_\_\_

Hit the \_\_\_\_\_

I breathe \_\_\_\_\_

It's a \_\_\_\_\_! (limited list: boy, girl)

lose the \_\_\_\_\_

Million dollar \_\_\_\_\_

most \_\_\_\_\_

Move over, \_\_\_\_\_.

Next stop \_\_\_\_\_

Only on \_\_\_\_\_

Sons of \_\_\_\_\_

That's so \_\_\_\_\_

The \_\_\_\_\_ effect

The \_\_\_\_\_ guy

The \_\_\_\_\_ thing

The \_\_\_\_\_ way

The forgotten \_\_\_\_\_

The whole \_\_\_\_\_

Those wacky \_\_\_\_\_

You need \_\_\_\_\_

you \_\_\_\_\_, you

\_\_\_\_\_ and then some

\_\_\_\_\_ are people, too.

\_\_\_\_\_ as a \_\_\_\_\_

\_\_\_\_\_ but not \_\_\_\_\_.

\_\_\_\_\_ do it (with) \_\_\_\_\_

\_\_\_\_\_ is not pretty

\_\_\_\_\_ like nobody's business.

\_\_\_\_\_ on a mission

\_\_\_\_\_ will be \_\_\_\_\_  
 \_\_\_\_\_ working for (you, us)  
 A day of \_\_\_\_\_  
 A whole nother \_\_\_\_\_  
 A \_\_\_\_\_ among \_\_\_\_\_  
 All eyes on \_\_\_\_\_  
 Aren't you a \_\_\_\_\_  
 Bad news for \_\_\_\_\_  
 Get your \_\_\_\_\_ on  
 Goodbye \_\_\_\_\_, hello \_\_\_\_\_  
 Have enough \_\_\_\_\_ there?  
 How \_\_\_\_\_ is that?  
 I don't do \_\_\_\_\_  
 I'm a \_\_\_\_\_ing fool  
 I'm all \_\_\_\_\_ed out.  
 I'm the \_\_\_\_\_ king  
 If \_\_\_\_\_ could talk.  
 If not \_\_\_\_\_, \_\_\_\_\_  
 In \_\_\_\_\_ we trust  
 In case of \_\_\_\_\_  
 It's all about \_\_\_\_\_.  
 like \_\_\_\_\_, like \_\_\_\_\_  
 most likely to \_\_\_\_\_  
 mother of all \_\_\_\_\_  
 My \_\_\_\_\_, my \_\_\_\_\_  
 no \_\_\_\_\_ee, no \_\_\_\_\_ ee  
 now that's a \_\_\_\_\_  
 One in a \_\_\_\_\_  
 Send us your \_\_\_\_\_  
 Shut up and \_\_\_\_\_  
 The \_\_\_\_\_ are coming.  
 The \_\_\_\_\_ that roared  
 The \_\_\_\_\_ type thing.  
 The hell with \_\_\_\_\_  
 The \_\_\_\_\_ are taking over.  
 Think outside the \_\_\_\_\_  
 Using the \_\_\_\_\_ word  
 What am I? \_\_\_\_\_  
 What's up with \_\_\_\_\_  
 When \_\_\_\_\_ goes bad  
 When the \_\_\_\_\_ comes  
 Why Johnny can't \_\_\_\_\_

You dog of \_\_\_\_\_  
 You want a \_\_\_\_\_?  
 \_\_\_\_\_ as \_\_\_\_\_ does.  
 \_\_\_\_\_ is my middle name.  
 \_\_\_\_\_ out and \_\_\_\_\_ somebody  
 \_\_\_\_\_ to end all \_\_\_\_\_  
 \_\_\_\_\_ is the new \_\_\_\_\_  
 All those \_\_\_\_\_ look alike.  
 All \_\_\_\_\_ all the time  
 And that man's a \_\_\_\_\_.  
 Friends don't let friends \_\_\_\_\_  
 He's a \_\_\_\_\_ among \_\_\_\_\_  
 I (he) eat(s) and breathe(s) \_\_\_\_\_  
 I eat \_\_\_\_\_ for breakfast.  
 I wouldn't be caught dead \_\_\_\_\_  
 I'll give you a \_\_\_\_\_  
 I'm (not) a \_\_\_\_\_ person  
 If you believe that, \_\_\_\_\_  
 It's not \_\_\_\_\_, it's \_\_\_\_\_  
 It's nothing if not \_\_\_\_\_  
 Leave the \_\_\_\_\_ at home  
 my \_\_\_\_\_ right or wrong.  
 My middle name is \_\_\_\_\_  
 No one teaches me \_\_\_\_\_  
 None of this \_\_\_\_\_ business  
 Not the way I \_\_\_\_\_.  
 it to (the) \_\_\_\_\_  
 The \_\_\_\_\_ behind the \_\_\_\_\_  
 The \_\_\_\_\_ de tutti \_\_\_\_\_  
 There's \_\_\_\_\_ and there's \_\_\_\_\_  
 When \_\_\_\_\_ is not enough  
 You call that a \_\_\_\_\_?  
 and I do mean \_\_\_\_\_  
 He makes a mean \_\_\_\_\_  
 \_\_\_\_\_ gives you a bad name.  
 A \_\_\_\_\_ walked into a bar.  
 Do I look like a \_\_\_\_\_?  
 He is too \_\_\_\_\_ by half  
 I'm not a big \_\_\_\_\_ person  
 If you \_\_\_\_\_ they will come.  
 Is that (a) \_\_\_\_\_ or what?  
 It was (a) \_\_\_\_\_ from hell.

Keep your eye(s) on the \_\_\_\_\_  
 Make like a \_\_\_\_\_ and \_\_\_\_\_.  
 So many \_\_\_\_\_, so little \_\_\_\_\_  
 So you think you can \_\_\_\_\_  
 That gives \_\_\_\_\_ a bad name  
 That was voted the most \_\_\_\_\_  
 The proof is in the \_\_\_\_\_  
 There's nothing \_\_\_\_\_ about it.  
 Wadda I look like, a \_\_\_\_\_ ?  
 Where in the \_\_\_\_\_ is \_\_\_\_\_.  
 Yes, Virginia, there is a \_\_\_\_\_  
 You're like a \_\_\_\_\_ to me.  
 You've got to love the \_\_\_\_\_  
 This is the sound of \_\_\_\_\_  
 \_\_\_\_\_, here, \_\_\_\_\_ there, \_\_\_\_\_ everywhere  
 A \_\_\_\_\_ to end all \_\_\_\_\_  
 One more \_\_\_\_\_ than the other  
 \_\_\_\_\_ is not just another pretty face.  
 \_\_\_\_\_ isn't just another \_\_\_\_\_ for \_\_\_\_\_  
 \_\_\_\_\_ is just another word for \_\_\_\_\_  
 A \_\_\_\_\_ does not a \_\_\_\_\_ make.  
 Changing \_\_\_\_\_ one \_\_\_\_\_ at a time.  
 Do you know where your \_\_\_\_\_ is (are)?  
 Have you ever seen a \_\_\_\_\_ ing  
 I can do \_\_\_\_\_ in my sleep.  
 I'm on that like \_\_\_\_\_ on \_\_\_\_\_  
 It's (he's, she's) a little too \_\_\_\_\_ by half  
 One man's \_\_\_\_\_ is another man's \_\_\_\_\_  
 Some of my best friends are \_\_\_\_\_  
 That \_\_\_\_\_ isn't going to \_\_\_\_\_ itself.  
 That was a \_\_\_\_\_ and a half  
 To think I was once (a) \_\_\_\_\_  
 We know \_\_\_\_\_ when we hear (see) it  
 What happens in \_\_\_\_\_ stays in \_\_\_\_\_.  
 What part of \_\_\_\_\_ don't you understand?  
 Who (what) do I look like? A \_\_\_\_\_?  
 With \_\_\_\_\_ like these, who needs \_\_\_\_\_  
 He's not the \_\_\_\_\_ in the \_\_\_\_\_.  
 I can do \_\_\_\_\_ with my eyes closed.  
 I wouldn't give you \_\_\_\_\_ for his \_\_\_\_\_  
 That's a \_\_\_\_\_ only a \_\_\_\_\_ could love

The \_\_\_\_\_ is the enemy of the \_\_\_\_\_  
 What do you take me for? A \_\_\_\_\_?  
 What if \_\_\_\_\_ is what it's all about?  
 You can take (your) \_\_\_\_\_ and shove it.  
 You've seen one \_\_\_\_\_, you've seen them all.  
 \_\_\_\_\_ is my name and \_\_\_\_\_ is my game.  
 \_\_\_\_\_ is not the \_\_\_\_\_est \_\_\_\_\_ in the \_\_\_\_\_  
 I know \_\_\_\_\_ like the back of my hand.  
 If you had his/my \_\_\_\_\_, you'd be \_\_\_\_\_(-ing) too.  
 What? Do I look like a \_\_\_\_\_ to you?  
 You can say hello to \_\_\_\_\_, goodbye to \_\_\_\_\_  
 \_\_\_\_\_ is a few \_\_\_\_\_ short of a full \_\_\_\_\_  
 A \_\_\_\_\_ without \_\_\_\_\_ is like a \_\_\_\_\_ without \_\_\_\_\_  
 A funny thing happened on the way to the \_\_\_\_\_  
 It's not just about (the) \_\_\_\_\_; it's about (the) \_\_\_\_\_  
 This is your brain. This is your brain on \_\_\_\_\_  
 I can do \_\_\_\_\_ with one hand tied behind my back.  
 You (I) must have been absent when they handed out the \_\_\_\_\_  
 \_\_\_\_\_: You can't live with them (it), and you can't live without them (it).  
 I may not know much about \_\_\_\_\_, but I know what I like.  
 Ask not what \_\_\_\_\_ can do for you, ask what you can do for \_\_\_\_\_.  
 You can take the \_\_\_\_\_ out of the \_\_\_\_\_, but you can't take the \_\_\_\_\_ out of the \_\_\_\_\_.

## Appendix II. Samples: the first ten items from the language survey.

All \_\_\_\_\_ trees look alike.  
 My bag is \_\_\_\_\_.  
 If you want the \_\_\_\_\_, just ask.  
 The players are \_\_\_\_\_!  
 I can \_\_\_\_\_ with my eyes closed.  
 I missed the \_\_\_\_\_.  
 There is a \_\_\_\_\_ waiting for you.  
 \_\_\_\_\_ is my middle name.  
 A stitch in time \_\_\_\_\_ nine.  
 It takes two to \_\_\_\_\_.

