HHS Public Access

Author manuscript

Ampersand (Oxford). Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:

Ampersand (Oxford). 2015 ; 2: 39–48. doi:10.1016/j.amper.2015.02.001.

The formulaic schema in the minds of two generations of native speakers

Diana Van Lancker Sidtis[1,2], **Krista Cameron**[2], **Kelly Bridges**[1,2], and **John J. Sidtis**[2,3]

[1]Communicative Sciences and Disorders, New York University, 665 Broadway, New York, NY 10012

[2]Brain and Behavior Laboratory, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10961

[3]Department of Psychiatry, New York University Langone School of Medicine, 550 First Avenue, New York, NY 10016

Abstract

Schemata are expressions that are fixed except for slots available for novel words (*I'm not a _____ person*). Our goals were to quantify speakers' knowledge, examine semantic flexibility in open slots, and compare performance data in two generations of speakers using cloze procedures in formulaic expressions, schemata open slots, fixed portions of schemata, and novel sentences. Fewer unique words appeared for the schemata-fixed and formulaic exemplars, reflecting speakers' knowledge of these utterances; the most semantic categories appeared for schemata-open responses. Age groups did not differ. Schemata exemplify creative interplay between novel lexical retrieval and fixed formulaic expression.

Introduction

Formulaic language has relevance to many branches of linguistic study and interest arising from many disciplines is increasing rapidly. It is known that formulaic expressions—conversational speech formulas, idioms, proverbs, expletives, and other fixed phrases--are important in processes of language development (Locke, 1993, 1997; Peters, 1977, 1983; Kempler et al., 1999) and that special challenges arise in second language learning (Lieven, 2007; Perkins, 1999; Foster, 2001). Conversational speech formulas have received considerable attention (Pawley and Syder, 1983; Fillmore, 1979; Tannen, 1989; Schegloff, 1988; Kuiper, 2007; Kuiper, 2009). It has been proposed that formulaic expressions played initially important roles in the evolution of human language (Code, 2005). Psycholinguistic studies suggest that formulaic expressions are processed faster or more cohesively than matched novel expressions (Clark, 1970; Swinney and Cutler, 1979; Libben and Titone, 2008; Sprenger, 2003). Further attesting to their holistic nature, constituent parts of idiomatic expressions were not recalled or recognized as well as those in matched novel expressions; (Horowitz and Manelis, 1973; Osgood and Housain, 1974), and participants

Corresponding author: Diana Van Lancker Sidtis, Ph.D., Communicative Sciences and Disorders, New York University, 665 Broadway, Room 936, New York, NY 10012, 917-224-5483, diana.sidtis@nyu.edu.

performed a judgment task more rapidly to the formulaic than the novel expression (Jiang and Nekrasova, 2007; Tabossi, Fanari, and Wolf, 2009). Eye movement studies showed an advantage for formulaic expressions (Conklin and Schmitt, 2008; Underwood et al., 2004; Siyanova-Chanturia, Conklin, and Schmitt, 2011).

More recently, the specific effects of neurological disease on incidence of formulaic language in spontaneous speech (Cappelle, Shtyrov, and Pulvermüller, 2010; Dieguez and Bogousslavsky 2007; Van Lancker Sidtis 2004; Van Lancker Sidtis and Postman 2006; Sidtis, Canterucci, and Katsnelson, 2009; Van Lancker Sidtis, 2012) point to differential cerebral systems underlying these two kinds of linguistic competence, suggesting that differential modes of processing are involved. This proposal is supported by models of brain function, details of which are beyond the scope of this paper (Bever, 1975; Ullman, 2004; Graybiel, 1998, 2008; for a review see Van Lancker Sidtis, 2014).

Despite this considerable scholarly activity, controversies remain about how to identify and quantify formulaic expressions in actual use. Most approaches use intuitions, assuming universal or general knowledge of idioms, speech formulas, proverbs, and so on (Barlovi-Harlig, 2012). One field study focused on use of proverbs in a naturalistic setting (Hain, 1951). Other approaches comb large written and (transcribed) spoken corpora, using automated algorithms of various kinds, identifying formulaic or collocational expressions and their relative incidence in texts (Moon, 1997, 1998a, b, c). In these approaches, extending interest to 3- or 4-word lexical bundles (*in the meantime, all things being equal*), frequency of occurrence in the texts of words in a specified order is a determining parameter (Biber, 2009; Conrad and Biber, 2004; Cowie, 1992; Biber, Conrad and Cortes, 2003).

Observational data regarding speakers' knowledge—an essential property of formulaic expressions—is sparse. Some familiarity rating systems for proverbs have been applied (Hallin and Van Lancker Sidtis, 2015), mainly with children (Nippold, 1991, 1998; Nippold and Rudzinski, 1993). It is now widely agreed that language users have command of a very large set of fixed expressions (along with the phonetic, prosodic, lexical, semantic, and usage characteristics unique to each one) (Kuiper, 2009; Lin, 2010; Lin and Adolphs, 2009; Bybee, 2002; Wray, 2002). Personal knowledge is an important fact, one that crucially differentiates the world of formulaic expressions from newly created language, and one that is implied in any study of proverbs, idioms, or conversational speech formulas. There is considerable evidence that a very large number of formulaic expressions are personally familiar, in the sense of being stored with their structure, meaning, and usage characteristics in the mental grammar of the native speaker (Bolinger, 1976, 1977; Jackendoff, 1995). This study is another in a series from our laboratory that attempts to probe and quantify speakers' knowledge of formulaic expressions and to establish incidence of actual use, using instruments designed for this purpose (Kempler and Van Lancker, 1996; Hall, 1996).

In an early study, it was shown that native speakers of English reliably identified the idiomatic from the literal intended meaning of ditropic (naturally ambiguous, as in *at the end of his rope*) sentences and the acoustic cues underlying these successful contrasts were identified (Van Lancker, Canter, and Terbeek, 1981). Later it was shown that this competence, distinguishing idiomatic from literal utterances, belonged to native speakers

only, in that even highly proficient nonnative speakers were significantly worse or performed at chance on the task (Van Lancker Sidtis, 2003). This ability was replicated using French (Abdelli-Beruh, Ahn, Yang and Sidtis, 2007) and Korean sentences (Yang, Ahn, and Van Lancker Sidtis, 2013), although different acoustic cues were found to form significant contrasts for ditropic utterances in these languages. Rammell and her colleagues (2013) demonstrated that listeners transcribed formulaic expressions presented auditorily in noise with 30% greater accuracy than matched novel expressions. These results support the notion that native speakers know formulaic expressions and can successfully utilize the acoustic cues belonging respectively to them.

The interest in quantifying formulaic language usage led to analysis of a screenplay, *Some Like It Hot*. Examiners' intuitions identified formulaic expressions and established a proportion of 25% in a screenplay (Van Lancker and Rallon, 2004). These utterances were adapted to a recall and recognition survey study, where it was established that formulaic expressions were recognized as formulaic, and missing words were correctly recalled, significantly more often for the formulaic than the matched novel expressions. Incidence data were then acquired from other naturally occurring discourse samples from healthy and brain damaged speakers. It was determined that approximately 25% of natural spontaneous speech is made up of words in formulaic expressions for normal speakers across a range of styles, participants, and topics in conversation (Sidtis, Canterucci, and Katsnelson, 2009; Bridges and Van Lancker Sidtis, 2013). Further, there were clear cut effects of neurological impairment: left hemisphere damage was associated with a significantly greater proportion of words in formulaic language, while right hemisphere damage showed significantly less (Van Lancker and Postman, 2006). Performance data from the vertical dimension of the brain, comparing cortical (Alzheimer's disease) with basal ganglia (Parkinson's disease) impairment, revealed retention of formulaic language in the former and loss in the latter (Bridges and Van Lancker Sidtis, 2013; Bridges, Van Lancker Sidtis and Sidtis, 2013). These studies lead to a model of formulaic language as governed by a right hemisphere-subcortical system (See Van Lancker, 2014, for review).

The study reported here continues the pursuit of performance data from healthy language users on formulaic expressions focusing on schemata. The schema is intermediary between formulaic and novel expressions. We probed speakers' knowledge of the linguistic schema by testing a sample of native speakers sorted into two age groups. This was followed by semantic analysis to evaluate the versatility of schemata open slots.

It has been suggested that large sets of formulaic expressions are known primarily to a particular generational age cohort and not to the generation before or after (Brown and Wright-Harp, 2011). This may be true certain instances of slang, which famously follows trends, often recycling to drop out and then appear a generation or two later (cf. *cool*). However, our perusal of very large lists of formulaic expressions spanning several decades does not support a notion of general decay of the larger repertory of formulaic language knowledge with time. For example, the recent survey, reported above, of knowledge by college students of formulaic expressions from *Some Like it Hot*, a film made in 1958 and released in 1959 (Wilder and Diamond, 1959), revealed high recognition of the expressions (Van Lancker Sidtis and Rallon, 2004), even though the story is set in an earlier time.

Contemporary ratings of a list of conversational speech formulas submitted by college students at Berkeley in the 1970s (Fillmore, 1979) revealed that these utterances were familiar and recognizable as formulaic expressions by today's students (Van Lancker Sidtis, 2011).

Schemata carry the characteristics of formulaic expressions: canonical form, specific lexical items in a certain order, stereotyped intonation, signature voice quality, and (often) precise articulatory detail (Van Lancker Sidtis, 2004). Like formulaic expressions, they exhibit connotational and social meanings; and they are known with these properties (form and meaning) to the native speaker. But schemata possess an additional versatility in having one or more free open slots. While formulemes *allow for* optional flexible lexical insertion or movement, for schemata, creative lexical insertion is *mandatory*, because at least one constituent slot is open. The open slot(s), which provide(s) the thematic crux of the utterance, is/are surprisingly versatile, allowing for a variety of lengths and grammatical forms. For example, *I'm not a ____ person* expresses a personal preference that is asserted to make up part of one's identity, as in *I'm not a morning person, I'm not a horror movie person, I'm not an eat and run person, I'm not a kissy kissy person, I'm not a leave someone in the lurch person*. Similarly, *The end of (the) X as we know it* communicates resignation, superior knowledge, and a bit of doom, all of which will color the meaning of X, which can be any word or phrase. This is the value of schemata: they provide the ability to communicate highly specialized nuances, while allowing for this meaning constellation to be applied to very disparate phenomena—the chosen novel words. A schema is a speech formula with the flexibility of novel insertion as part of the phrase itself.

Verbal schemata had received only sparse mention in the scholarly linguistic literature (Lyons, 1968; Crystal, 1995). Recently, linguistic blogs and popular media have became active describing versions of schemata and related phenomena, originally mostly in written contexts. The term *snowclone* was coined (Pullum, 2003, 2004) to refer, first, to journalistic turns of phrase that utilize a prefabricated phrase or familiar, reified concept to introduce a new topic, using the classic example *If Eskimos have **N** words for snow, **X** surely have **Y** words for **Z*** (see also Pullum, 1991). Since then, many hundreds of examples have been submitted to the respective websites along with international commentary. Contributions from the public continue to supply numerous examples that come from spoken language and thus are unlikely to be identified in published corpora: *If that's X, every Y should be so lucky; X gone wild; no rest for the X; A lot of people, when they have a problem, say 'I know, I'll use X'. Now they have two problems; An Xer shade of Y; If it's not the X, it's the Y; That's why they call it X; Once an X, always an X*. The snowclone notion has been picked up by the popular media (e.g., McFedries, 2008).). A German website[1] lists 61 examples, such as *Ein X kommt selten allein* (*A X comes seldom alone*) and *und ewig lockt X* (and X is eternally seductive). It is clear from the enthusiastic responses in blogs, websites and journalistic reports that these phrases have a vibrant presence in native language competence. In all these discussions, it has been assumed that people know the utterances--that they are personally familiar in the sense of being stored with form, meaning, and usage

[1](http://emmanuel.dammerer.at/snowclonery)

principles. Questions arise about how generally and reliably the expressions are known, whether there is an effect of age cohort, and the semantic versatility of the mandatory open slots. This study was designed to address these questions empirically, utilizing schemata gathered from actual usage.

Purpose of study

The purpose of this study was to examine native speakers' knowledge of formulemes (the canonical forms of formulaic expressions) and schemata as contrasted with their performance on novel (newly created) expressions and to investigate such knowledge across two generations of native speakers. Formulaic sentences, such as idioms and conversational speech formulas, are generally fixed in that certain words appear in a certain order. Schemata are similar with the notable exception that there is one (or more than one) "mandatory" open slot, which is filled at the discretion of the speaker, while producing the rest of the expression with the inherent characteristics (i.e., connotations) belonging to formulaic expressions. There is a freedom of lexical choice in schemata which is not a property of standard formulaic expressions. This places schemata, in a sense, midway between novel expressions, which are freely generated according to the grammatical rules of the language, and formulaic expressions, which are unitary in form.

The questions were:

1. Do native speakers agree on the lexical content of formulaic expressions, implying common knowledge of the expressions?
2. Do native speakers agree on the lexical content of the fixed portions of schemata, implying common knowledge of the expressions?
3. To what extent are native speakers able to utilize the creative capacities of schemata, as available in the open slots, and of novel expressions? Does creativity in schemata insertions match or exceed that of novel sentences?
4. Are differences in performance (reflecting knowledge and familiarity of the expressions as well as semantic creativity) to be found between two different age groups? It is assumed that performance data can be interpreted to reveal the status of competence for formulaic expressions in these two age groups.
5. Are there differences in number and types of semantic categories utilized for generating novel words in novel expressions compared to words generated in the mandatory open slots of schemata?

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. In contrast, responses written into the blanks in novel sentences and the novel (mandatory) open slots in schemata were predicted to form a more diverse set of lexical items across a broader range of semantic categories. From our perusal of lists of formulaic and idiomatic expressions accumulated over several decades from various sources, we predicted that there would be no significant effect of age cohort.

Method

Stimuli

Forty formulaic expressions (e.g., *It was a blessing in disguise*) were selected from lists previously evaluated for familiarity by native speakers of American English (Van Lancker Sidtis and Rallon, 2004). Formulaic expressions included conversational speech formulas, idioms, and proverbs. Forty novel (newly created, grammatical) sentences (e.g., *The two of you are soaked*) were created to match the formulaic expressions on number of words (+/− 1 word). Novel sentences contained common lexical items and were plausible in meaning. Eighty schemata (those with only one open slot), each selected from a working list of schemata accumulated and recorded during several years from observed actual usage (see Appendix I), were divided into two subsets of 40. (See Appendix II for a sample of the survey sheet.) The 160 test items, randomized and compiled onto an answer sheet, each featured a blank (cloze procedure) for participants to fill in the missing word (Taylor, 1953). The four groups of stimuli utilized for the slot-filler task (Table 1) are referred to in this study as *formulas* (standard formulaic sentences), *novel sentences* (newly created sentences), *schemata-fixed* (schemata with an open slot in the fixed portion of the expression) and *schemata-open* (schemata with a blank in the mandatory open slot where the novel word belongs). In the formulas and novel items, the locations of the blank (open slot) were balanced across initial, middle, and final position. For the 40 open schemata, a natural open slot was provided in the "mandatory" position (*He eats and breathes* ____). In the second set of 40 schemata, the fixed schema 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, a novel word was provided in the (mandatory) natural open slot position, and an open slot was created in the fixed portion of the schema. The set of fixed schemata was included to probe subjects' knowledge of the schema itself. For a display of sample items and responses, see Table 1.

Task

Raters—Forty native speakers of English with normal vision formed two age groups of rater participants. "Native speaker" is defined in our study as born and educated in the United States and speaking English in the home since infancy. The younger age group included four males and 16 females with mean age of 25.05 years and an age range of 21 to 33 years (*SD* = 3.47). The older age group included seven males and 13 females with a mean age of 59.80 years, ranging from 47 to 89 years (*SD* = 10.23). The younger age group had an average of 17.73 (Range = 12 – 20; *SD* = 2.45) years of education and the older age group completed an average of 15.90 years of education (Range = 14 – 21; *SD* = 2.29). An independent-samples *t*-test revealed that the younger group of raters had significantly more years of education than the older raters [$t(38) = 2.44$, $p = 0.02$].

Procedure—After completing the written informed consent form, raters were given the test protocol, for which they were instructed to write down one word at each open slot provided. Subjects were requested to write down a single word that seemed to fit in the slot, and to guess responses if they were not sure. (The few two-word responses were discarded.)

All one-word responses were recorded and numbers of unique word types produced in each utterance category were calculated, followed by classification of responses into semantic categories. The procedures for the unique word and semantic category analyses are described below.

Analysis procedure

Target-word matches—For the formulaic and schemata-fixed utterances, as mentioned, the blanks were designed to elicit a target word from responders. In the majority of cases, there was one acceptable target word that was considered a correct match. However, for several of the utterances, there were two (or three) possible correct matches for the target word. For the formulaic utterances, the following sentences had two alternative acceptable target words: "The _____ have turned" (*TIDES* or *TABLES*), and "I've got to _____ it to you" (*HAND* or *GIVE*). For the schemata-fixed utterances, the following sentences had two or three alternative acceptable target words: "Where in the _____ is the car?" (*WORLD*, *HECK*, or *HELL*), "I'm not a glitter _____ " (*PERSON* or *FAN*). These alternate words were determined to be acceptable target-matches after consultation between two native English-speaking persons trained in language analysis.

Unique word analysis—The first measure focused on the number of unique words by raters for each stimulus set. The number of unique words (out of a possible forty) was recorded for each sentence. Means were calculated for each of the four sentence types (formulaic, novel, schemata-fixed, schemata-open). Words entered in the survey sheet by younger raters (n = 20) and those given by older raters (n = 20) were recorded separately for each sentence. Means were calculated for each sentence type separately for younger and older raters.

Semantic category analysis—Fields' (2013) conceptualization of semantic fields was used as a guideline to categorize rater responses into semantic categories. Using this framework, words were categorized into one of twenty-two distinct semantic categories. (Please refer to Table 2 for a list of these semantic categories (#1–22) and examples). As some words provided during this task did not fit into one of Fields' (2013) defined categories, an additional set of seven semantic categories was developed. Table 2 contains the remaining seven categories and examples of word types in each category.

The total number of semantic categories overall was calculated and recorded for each sentence. Means were calculated for the number of semantic categories used in each sentence type (formulaic, novel, schemata-fixed, schemata-open). In an additional analysis, the number of semantic categories was again counted for each target item. Quantification was completed separately for younger and older raters. Means for younger and older raters for each of the four subtypes of sentences were calculated and used for additional comparisons.

Results

Subjects were successful in identifying the target words in formulaic and schemata-fixed stimuli (Table 3). Mean numbers of unique words entered for each stimulus type were compared (Table 4). An independent-groups analysis of variance (ANOVA) assessed possible differences in the numbers of unique words in the open slot across the four sentence types and two age groups. There was a significant effect of sentence type [$F(3, 312) = 86.162$; $p < 0.001$] on the number of unique words but there was no significant effect of age group. Further, sentence type and age group did not interact. There were fewer unique words in the formulaic sentences compared to the novel [$t(158) = -15.195$; $p < 0.001$], schemata-fixed [$t(158) = -5.114$; $p < 0.001$], and schemata-open [$t(158) = -13.503$; $p < 0.001$] sentences. Comparing the novel sentences to the two types of schemata, there were more unique words in the novel sentence open slots than in the schemata-fixed open slots [$t(158) = 8.177$; $p < 0.001$], but there was no difference in the number of unique words produced for the novel and schemata-open sentences. As one would expect, it was also the case that there were more novel words produced for the schemata-open than the schemata-fixed sentences [$t(158) = -7.477$; $p < 0.001$]. These results are depicted in Figure 1.

Semantic Category Analysis

Comparisons were made between the utterance types (formulaic, novel, schemata-fixed, schemata-open) for the mean number of semantic categories represented in responses by raters (total group, $N = 40$) (See Table 5). An independent groups ANOVA revealed a significant effect of sentence-type on the number of semantic categories represented [$F(3, 312) = 46.975$; $p < 0.001$] but there was no significant effect of age group, nor was there a significant interaction between age group and sentence type. There were fewer semantic categories in the formulaic sentences compared to the novel [$t(158) = -8.723$; $p < 0.001$], schemata-fixed [$t(158) = -4.305$; $p < 0.001$] and schematic-open [$t(158) = -11.150$; $p < 0.001$]. Comparing novel sentences to the other two sentence types, there were significantly more semantic categories in represented in novel than schematic fixed utterances [$t(158) = -4.289$; $p < 0.001$] and schemata-open utterances [$t(158) = -6.842$; $p < 0.001$]. However, comparisons between novel and schematic open sentences did not reach statistical significance; mean numbers of categories for responses to these sentence types were 5.2 (novel) and 6.1 (schematic open), compared to semantic-fixed (3.8) and formulaic (2.6). These results are depicted in Figure 2.

Discussion

The main purpose of this study was to examine speakers' knowledge of formulaic expressions and schemata in two age groups. Schemata are special types of formulaic expressions with one or more mandatory open slot(s) for insertion of a novel lexical item. Using a survey that provided open slots for speech formulas, novel sentences, the fixed portion of a schema (schemata-fixed) and the open-slot portion of a schema (schemata-open), participants showed knowledge of the formulas and the fixed portions of schemata, and they entered a range of novel words in novel expressions and the open slots of schemata. In the semantic analysis, it was seen that participants' entries for novel and schematic-open

trials differed significantly from entries in the formulas and the fixed portions of the schemata, indicating enhanced creativity for newly created sentences and the open slot of a schema.

These findings support the view that native speakers know formulaic expressions in their canonical lexical form: native speakers indicated knowledge of classical formulaic expressions (idioms and conversational speech formulas) as well as the fixed portions of schemata. Further, they gave evidence of implicit knowledge of the large range of lexical choices available to them in schemata-open forms. In fact, the mean number of semantic categories was (nonsignicantly) higher for entries in schemata-open slots than in novel sentence slots. This indicates that schemata are well positioned to utilize the advantages of formulaic expressions simultaneously with retrieval from the novel word lexicon. A possible explanation for the increased semantic range seen in entries in open slots for schemata in comparison with novel expressions may lie in transitional probabilities and semantic coherence (Schwanenflugel & LaCount, 1988). The novel expressions in this study were meaningful, literally interpretable, and semantically well-formed. In contrast, schemata can carry nonliteral meanings, not requiring the usual linguistic relationships between words preceding and following the open slot. This provides greater freedom in selecting the inserted lexical item.

These results, showing that speakers of a language perform differently for novel and formulaic expressions, lend some support to a model of language processing that posits an interplay of two processing modes, novel and formulaic (Lounsbury, 1963; Van Lancker Sidtis, 2004, 2014; Wray and Perkins, 2000). There are numerous online studies that have suggested such a proposal (Swinney and Cutler, 1979; Tabossi, Fanari, and Wolf. 2009; Katz and Ferretti, 2001; Lin 2010; Reuterskiöld and Van Lancker Sidtis, 2013). Further, it has been proposed, as mentioned previously, that these two kinds of language are modulated by differing cerebral processes. These facts have strong implications for models of language as well as for language rehabilitation following brain damage (Van Lancker Sidtis, 2012).

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, nonliteral 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; Lounsbury, 1963; Van Lancker Sidtis, 2004; Wray and Perkins, 2000). 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. Formulaic schemata illustrate the dual mode process in linguistic competence, in which these two distinct modes coexist in continuous interplay.

Qualitative discussion of schemata: status in language competence

Examination into the provenance of individual schemata reveals that their origins, when traceable, are highly heterogeneous (titles of books or films, quotes, song lyrics, lines from poetry, slogans, dialogue in plays, etc.) and many are unknown. This has also been shown for a smaller set of German schemata (see Footnote [1]). A robust presence of this constituent of language—fixed constructions that invite a fecundity of variation-- can be seen in the many linguistic phenomena in the world around us: bumper stickers, newspaper headlines, and advertising copy all utilize the trope of a known phrasal structure treated with different lexical items. A compelling example is seen in the advertising slogan developed by the telephone company AT&T in the 1980s, “reach out and touch somebody” (Ramey, 2008). Playing on that slogan, a New Yorker cartoon by R. Reilly depicted a jungle with thought balloons, presumably generated by fauna in the area, emanating from sky, trees, bushes, and underbrush in the scene, all playing on the original slogan, which itself does not appear (See Table 6)

Formulaic expressions make up a very heterogeneous set, having only in common that they are not newly created, as are novel expressions. They have been usefully represented along a continuum (Penttilä, 2010) governed by such parameters as category of expression, social role, attitudinal and affective content, degrees of coherence, and frequency of exposure (Van Lancker, 1975, 1988), ranging from fixed to novel (see also Barkema, 1996). Fixed formulaic expressions, as in the idiom or conversational speech formula, schemata, indirect requests and sentence stems take their places along the continuum depending on criteria important for classifying these types into categories. More generalized, structured constructions, pairing form and function, that are posited in construction grammar, may take a place in this configuration (Goldberg, 1995, 2006).

Recent studies document verbatim retention of spoken propositional (Gurevich, Johnson, and Goldberg, 2010) and idiomatic utterances following single exposure (Reuterskiöld and Van Lancker Sidtis, 2013), while Goldinger (1996) and others have shown that phonetic and voice characteristics heard in lists of words are retained by listeners and strongly influence speech perception. These studies indicate that the physical characteristics of utterances may be retained *in toto* in memory, following acquisition procedures described for other kinds of learning (e.g., Horn, 1985; see Kreiman and Sidtis, 2011, pp. 224–228). Our studies propose that learning following a single exposure is even more likely when the utterances have strong attitudinal nuances and nonliteral meanings, such as conversational speech formulas, idioms, and schemata. The ability of formulaic expressions to be acquired in a single exposure can account for the proposed storage and processing of a very large repertory. Subtle contingencies involving the prodigious capacity of human memory for linguistic phenomena contrasted with opportunities for frequent exposure remain to be understood.

Acknowledgements

This work was supported in part by NIH R01 DC007658.

References

Abdelli-Beruh, N.; Ahn, J.; Yang, S.; Sidtis, D. Acoustic cues differentiating idiomatic from literal expressions across languages; American Speech and Hearing Association Convention; Boston, M.A.. November 15–17; 2007.

Altenberg, B. On the phraseology of spoken English: the evidence of recurrent word-combinations. In: Cowie, AP., editor. Phraseology. Oxford: Clarendon Press; 1998. p. 101-124.

Bardovi-Harlig K. Formulas, routines, and conventional expressions in pragmatics research. Annual Review of Applied Linguistics. 2012; 32:206–227.

Barkema H. Idiomaticity and terminology: A multi-dimensional descriptive model. Studia Linguistica. 1996; 50(2):125–160.

Bever TG. Cerebral asymmetries in humans are due to the differentiation of two incompatible processes: Holistic and analytic. Annals of the New York Academy of Science. 1975; 263:251–262.

Biber D. A corpus-driven approach to formulaic language in English. International Journal of Corpus Linguistics. 2009; 14:275–311.

Biber, D.; Conrad, S.; Cortes, V. Lexical bundles in speech and writing: An initial taxonomy. In: Wilson, A.; Rayson, P.; McEnery, T., editors. Corpus linguistics by the Lune: A festschrift for Geoffrey Leech. Frankfurt: Peter Lang; 2003. p. 71-93.

Bolinger D. Meaning and memory. Forum Linguisticum. 1976; 1:1–14.

Bolinger D. Idioms have relations. Forum Linguisticum. 1977; 2(2):157–169.

Bridges K, Van Lancker Sidtis D. Formulaic language in Alzheimer's disease. Aphasiology. 2013; 27(7):799–810.

Bridges K, Van Lancker Sidtis D, Sidtis JJ. The role of subcortical structures in recited speech: Studies in Parkinson's disease. Journal of Neurolinguistics. 2013; 26(6):591–601.

Brown J, Wright-Harp W. Cultural and generational factors influencing proverb recognition. Contemporary Issues in Communication Sciences and Disorders (CICSD). 2011; 38:111–122.

Bybee J. Phonological evidence for exemplar storage of multiword sequences. Studies in Second Language Acquisition. 2002; 24:215–221.

Cappelle B, Shtyrov Y, Pulvermüller F. Heating up or cooling up the brain: MEG evidence that phrasal verbs are lexical units. Brain and Language. 2010; 115:189–201. [PubMed: 21030069]

Clark, HH. Word associations and linguistic theory. In: Lyons, J., editor. New horizons in linguistics. Baltimore: Penguin Books; 1970. p. 271-286.

Code C. First in, last out? The evolution of aphasic lexical speech automatisms to agrammatism and the evolution of human communication. Interaction Studies. 2005; 6:311–334.

Conklin K, Schmitt N. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? Applied Linguistics. 2008; 29(1):72–89.

Conrad S, Biber D. The frequency and use of lexical bundles in conversation and academic prose. Lexicographica. 2004; 20:56–71.

Cowie, AP. Multiword lexical units and communicative language teaching. In: Arnaud, PJL.; Bejoint, H., editors. Vocabulary and applied linguistics. London: Macmillan; 1992. p. 1-12.

Crystal, D. The Cambridge Encyclopedia of the English Language. Cambridge: Cambridge University Press; 1995.

Dieguez, S.; Bogousslavsky, J. Baudelaire's aphasia: From poetry to cursing. In: Bogousslavsky, Julien; Hennerici, MG., editors. Neurological disorders in famous artists, Part 2, Vol. 22. Frontiers of Neurology and Neuroscience. Basel: Karger; 2007. p. 121-149.

Erman B, Warren B. The idiom principle and the open choice principle. Text-International Journal for the Study of Discourse. 2000; 20(1):29–62.

Fields HC. Indo-European linguistics: Semantic fields. Linguistics Research Center in The College of Liberal Arts at UT Austin. 2013 from http://www.utexas.edu/cola/centers/lrc/iedocctr/ie-ling/ie-sem/.

Fillmore, C. On fluency. In: Fillmore, CJ.; Kempler, D.; Wang, WS-Y., editors. Individual differences in language ability and language behavior. London: Academic Press; 1979. p. 85-102.

Foster, P. Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. In: Bygate, M.; Skehan, P.; Swain, M., editors. Researching pedagogic tasks: Second language learning, teaching, and testing. Harlow: Longman; 2001. p. 75-93.

Goldberg, A. Constructions at work: The nature of generalization in language. Oxford, UK: Oxford University Press; 2006.

Goldberg, A. Constructions: A construction grammar approach to argument structure. Chicago: The University of Chicago Press; 1995.

Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1996; 22(5): 1166–1183.

Graybiel AM. Habits, rituals, and the evaluative brain. Annual Review of Neuroscience. 2008; 31:359–387.

Graybiel AM. The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory. 1998; 70:119–136. [PubMed: 9753592]

Gurevich O, Johnson MA, Goldberg AE. Incidental verbatim memory for language. Language and Cognition. 2010; 2(1):45–78.

Hain, M. Sprichwort und Volkssprache. Sidtis, D.; Mohr, S., translators and editors. Giessen: Wilhelm Schmitz Verlag. English trans; 1951. *Formulaic language in the field*, Anja Tachler, translator. Copyright.

Hall, E. Northridge Evaluation of Formulas, Idioms and Proverbs in Social Situations NEFIPPS). Northridge, CA: 1996. Copyright.

Hallin A, Van Lancker Sidtis D. A closer look at formulaic language: Prosodic patterns in Swedish proverbs. Applied Linguistics. 2015 in press.

Horn, G. Oxford Psychology Series No. 10. Oxford: Clarendon Press; 1985. Memory, imprinting and the brain.

Horowitz LM, Manelis L. Recognition and cued recall of idioms and phrases. Journal of Experimental Psychology. 1973; 100:291–296.

Jackendoff, R. The boundaries of the lexicon. In: Everaert, M.; van der Linden, E-J.; Schenk, A.; Schreuder, R., editors. Idioms: Structural and psychological perspectives. Hillsdale, NJ: Lawrence Erlbaum Associates; 1995. p. 133-166.

Jiang N, Nekrsova TM. The processing of formulaic sequences by second language speakers. The Modern Language Journal. 2007; 91(3):433–445.

Katz AN, Ferretti TR. Moment-by-moment reading of proverbs in literal and nonliteral contexts. Metaphor and Symbol. 2001; 16(3–4):193–221.

Kempler D, Van Lancker D. The Formulaic and Novel Language Comprehension Test (FANL-C). 1996 Copyright. http://blog.emerson.edu/daniel_kempler/fanlc.html.

Kempler D, Van Lancker D, Marchman V, Bates E. Idiom comprehension in children and adults with unilateral brain damage. Developmental Neuropsychology. 1999; 15(3):327–349.

Kreiman, J.; Sidtis, D. Foundations of voice studies: An interdisciplinary approach to voice production and perception. Hoboken, NJ: Wiley-Blackwell; 2011.

Kuiper, K. Formulaic genres. Basingstoke: Palgrave Macmillan; 2009.

Kuiper, K. Cathy Wilcox meets the phrasal lexicon: Creative deformation of phrasal lexical items for humorous effect. In: Munat, Judith, editor. Lexical creativity, texts and contexts. Amsterdam: John Benjamins; 2007. p. 93-112.

Libben MR, Titone D. The multidetermined nature of idiom processing. Memory and Cognition. 2008; 36:1103–1121. [PubMed: 18927029]

Lieven, E. Producing multiword utterances. In: Kelly, B.; Clark, EV., editors. Constructions in acquisition. Stanford, CA: CSLI Publications; 2007.

Lin, PMS.; Adolphs, S. Sound evidence: Phraseological units in spoken corpora. In: Barfield, A.; Gyllstad, H., editors. Collocating in another language: Multiple interpretations. Basingstoke, England: Palgrave Macmillan; 2009. p. 34-48.

Lin, PMS. The phonology of formulaic sequences: A review. In: Wood, D., editor. Perspectives on formulaic language: Acquisition and communication. London, UK: Continuum; 2010. p. 174-193.

Locke JL. A theory of neurolinguistic development. Brain and Language. 1997; 58:265–326. [PubMed: 9182750]

Locke, JL. The child's path to spoken language. Cambridge, MA: Harvard University Press; 1993.

Lounsbury, FG. Linguistics and psychology. In: Koch, S., editor. Psychology: Study of a science. New York: McGraw-Hill; 1963. p. 553-582.

Lyons, J. Introduction to theoretical linguistics. Cambridge, U.K.: Cambridge University Press; 1968.

McFedries P. Snowclone Is The New Cliché. Spectrum. IEEE. 2008

Moon, R. Fixed expressions and idioms in English: A corpus-based approach. Oxford, UK: Clarendon Press; 1998a.

Moon, R. Vocabulary connections: Multi-word items in English. In: McCarthy, M., editor. Vocabulary: Description, acquisition and pedagogy. New York: Cambridge University Press; 1997. p. 40-63.

Moon, RE. Frequencies and forms of phrasal lexemes in English. In: Cowie, AP., editor. Phraseology. Oxford: Clarenden Press; 1998b. p. 79-100.

Moon, RE. Oxford Studies in Lexicology and Lexicography. Oxford: Clarendon Press; 1998c. Fixed expressions and text: a study of the distribution and textual behaviour of fixed expressions in English.

Nippold M. Evaluating and enhancing idiom comprehension in language-disordered students. Language, Speech and Hearing Services in Schools. 1991; 22:100–106.

Nippold M, Rudzinski M. Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. Journal of Speech and Hearing Research. 1993; 36:728–737. [PubMed: 8377485]

Nippold MA. Later language development: The school-age and adolescent years (second edition). 1998 Pro Ed.

Osgood CE, Housain R. Salience of the word as a unit in the perception of language. Perception and Psychophysics. 1974; 15:168–192.

Pawley, A.; Syder, FH. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In: Richard, JC.; Schmidt, R., editors. Language and Communication. London: Longman; 1983. p. 191-226.

Penttilä, E. A prototype-based taxonomy of idiomatic expressions. In: Kristiansen; Ruiz de Mendoza Ibáñez; Tabakowska; Choi´nski; Wiraszka, editors. Applications of Cognitive Linguistics 14. Cognitive Linguistics in Action. Berlin: Walter de Gruyter; 2010. p. 145-162.

Perkins, MR. Productivity and formulaicity in language development. In: Garman, M.; Letts, C.; Richards, R.; Schelletter, C.; Edwards, S., editors. Issues in normal and disordered child language: From phonology to narrative. Reading: University of Reading; 1999. p. 51-67.Special Issue of The New Bulmershe Papers

Peters, A. The units of language. Cambridge, U.K.: Cambridge University Press; 1983.

Peters AM. Language learning strategies: Does the whole equal the sum of the parts? Language. 1977; 53(3):560–573.

Pullam G. Snowclones: lexicographical dating to the second. Language Log. 2004 Posted January 16, 01:37PM.

Pullum, G. The Great Eskimo Vocabulary Hoax. In: Pullum, G., editor. The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language. Chicago, IL: University of Chicago Press; 1991. p. 159-171.

Pullum GK. Phrases for lazy writers in kit form. Language Log. 2003

Ramey CH. The metaphorical mind: When AT & T asked us to 'reach out and touch someone', did they mean that literally? Psychology Today. 2008 psychologytoday.com/blog/the-metaphorical-mind/200807.

Rammell, CS.; Pisoni, DB.; Van Lancker Sidtis, D. Perception of formulaic and novel utterances under acoustic degradation: Evidence for a unitary memory trace. Paper presented at Midwestern Cognitive Science Conference; Columbus, Ohio. May 18; 2013.

Reuterskiöld C, Van Lancker Sidtis D. Retention of idioms following one-time exposure. Child Language Teaching and Therapy. 2012; 29(2):216–228.

Schegloff, E. Discourse as an interactional achievement: An exercise in conversation analysis. In: Tannen, D., editor. Linguistics in context: Connecting observation and understanding. Ablex, Norwood, NY: 1988. p. 135-158.

Schwanenflugel P, LaCount K. Semantic relatedness and the scope of facilitation for upcoming words in sentences. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1988; 14(2):344–354.

Sidtis D, Canterucci G, Katsnelson D. Effects of neurological damage on production of formulaic language. Clinical Linguistics and Phonetics. 2009; 23(4):270–284. [PubMed: 19382014]

Sinclair, JM. Corpus concordance collocation. Oxford: Oxford University Press; 1991.

Siyanova-Chanturia A, Conklin K, Schmitt N. Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. Second Language Research. 2011; 27(2): 251–272.

Sprenger, Simone A. MPI Series in Psycholinguistics. Vol. 21. Wageningen/Niederlande: Ponsen and Looijen BV.; 2003. Fixed expressions and the production of idioms.

Stephens LL. The role of memory in the relationship between affect and familiarity. Cognition and Emotion. 1988; 2:333–349.

Swinney D, Cutler A. The access and processing of idiomatic expressions. Journal of Verbal Learning and Verbal Behavior. 1979; 18:523–534.

Tabossi P, Fanari R, Wolf K. Why are idioms recognized fast? Memory and Cognition. 2009; 37(4): 529–540. [PubMed: 19460959]

Tannen, D. Talking voices: Repetition, dialogue, and imagery in conversational discourse. Cambridge: Cambridge University Press; 1989.

Taylor WL. “Cloze procedure.”: a new tool for readability. Journal Quarterly. 1953; 30:415–433.

Ullman MT. Contributions of memory circuits to language: the declarative/procedural model. Cognition. 2004; 92:231–270. [PubMed: 15037131]

Underwood, G.; Schmitt, N.; Galpin, A. The eyes have it: An eye-movement study into the processing of formulaic sequences. In: Schmitt, N., editor. Formulaic sequence. Amsterdam: John Benjamins; 2004. p. 155-172.

Van Lancker Sidtis D. When novel sentences spoken or heard for the first time in the history of the universe are not enough[2]: Toward a dual-process model of language. International Journal of Language and Communication Disorders. 2004; 39(1):1–44. [PubMed: 14660185]

Van Lancker Sidtis D, Rallon G. Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. Language and Communication. 2004; 24:207–240.

Van Lancker Sidtis D. Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. Applied Psycholinguistics. 2003; 24(1):45–57.

Van Lancker Sidtis, D. Linguistic approaches to nonliteral language: We really knew how to have fun. In: Kuiper, K., editor. Teaching Linguistics. England: Equinox; 2011. p. 110-136.

Van Lancker Sidtis D. Formulaic language and language disorders. The Annual Review of Applied Linguistics. 2012; 32:62–80.

Van Lancker Sidtis, D. Formulaic language in an emergentist framework. In: MacWhinney, B.; O’Grady, W., editors. Handbook of language emergence. Wiley-Blackwell: 2014. in press.

Van Lancker D, Canter GJ, Terbeek D. Disambiguation of ditropic sentences: acoustic and phonetic cues. Journal of Speech and Hearing Research. 1981; 24:330–335. [PubMed: 7300273]

Van Lancker Sidtis D, Postman WA. Formulaic expressions in spontaneous speech of left- and right-hemisphere damaged subjects. Aphasiology. 2006; 20(5):411–426.

Wilder, B.; Diamond, IAL. Some Like It Hot. Screenplay reprinted. In: Thomas, S., editor. Best American Screenplays 2. 1st ed. Vol. 1990. New York: Crown Publishers; 1959. p. 80-146.

[2]Cf. Pinker, 1995, p. 22

Van Lancker D, Canter GJ, Terbeek D. Disambiguation of ditropic sentences: acoustic and phonetic cues. Journal of Speech and Hearing Research. 1981; 24:330–335. [PubMed: 7300273]

Wolf R, Van Lancker Sidtis D, Sidtis JJ. The ear craves the familiar: Pragmatic repetition in left and right cerebral damage. Aphasiology. 2014; 28(5):596–615.

Wray, A. Formulaic language and the lexicon. Cambridge: Cambridge University Press; 2002.

Wray A, Perkins M. The functions of formulaic language: An integrated model. Language and Communication. 2000; 20:1–28.

Wulff, S. Words and idioms. In: Trousdale, G.; Hoffman, T., editors. The Oxford handbook of construction grammar. Oxford, UK: Oxford University Press; 2013. p. 274-289.

Yang SY, Ahn J-S, Van Lancker, Sidtis D. Listening and acoustic studies of idiomatic-literal contrastive sentences in Korean. Speech, Language, and Hearing. 2015 In press.

Appendix I

List of schemata observed and recorded from ordinary language use. From this list, eighty schemata (those with one open slot) were chosen for this study, serving as stimuli for the schemata-open and the schemata-fixed conditions

1. ____'sville
2. ____ city
3. ____ days
4. ____ fool.
5. ____ galore
6. ____ happy
7. ____ hunting.
8. ____ much?
9. ____ power
10. ____ shm _____
11. ____ thinking
12. ____ this.
13. ____ time
14. ____ wars
15. ____crazy.
16. ____wars
17. Dead _____
18. Everything ____
19. Fuck _____
20. Get ____
21. Go, ____!

22. Got ____?

23. nice ____

24. Perfect _____

25. Screw ____

26. That's ____

27. Think ____

28. You: ____

29. ____ and counting

30. ____ and proud

31. ____ are us

32. ____ be us

33. ____ is overrated.

34. ____ loves ____ (written)

35. ____ to ____ (A,Z Mon,Fri, soup, nuts)

36. ____ to death

37. ____ under fire

38. A _____'s ____ (word repeated)

39. A royal ____

40. A walking____

41. All things ____

42. Call me ____

43. Color me ____

44. Do not ____

45. Down with _____

46. For the ____

47. Giant among ____

48. Go and ____

49. Hit the _____

50. I breathe _____

51. It's a ____! (limited list: boy, girl)

52. lose the ____

53. Million dollar _____

54. most _____ ____

55. Move over, _____.

56. Next stop _____

57. Only on _____

58. Sons of ____

59. That's so ____

60. The ____ effect

61. The ____ guy

62. The ____ thing

63. The ____ way

64. The forgotten _____

65. The whole ____

66. Those wacky ____

67. You need _____

68. you _____, you

69. ____ and then some

70. ____ are people, too.

71. ____ as a _____

72. ____ but not ____.

73. ____ do it (with) _____

74. ____ is not pretty

75. ____ like nobody's business.

76. ____ on a mission

77. ____ will be ____

78. ____ working for (you, us)

79. A day of ____

80. A whole nother ____

81. A____among _____

82. All eyes on ____

83. Aren't you a _____

84. Bad news for _____

85. Get your _____ on

86. Goodbye ____, hello _____

87. Have enough _____ there?

88. How _____ is that?

89. I don't do _____

90. I'm a ____ing fool

91. I'm all ____ed out.

92. I'm the ____ king

93. If _____ could talk.

94. If not ____, ____

95. In _____ we trust

96. In case of____

97. It's all about _____.

98. like____, like _____

99. most likely to ____

100. mother of all ____

101. My ____, my _____

102. no _____ee, no _____ ee

103. now that's a _____

104. One in a ____

105. Send us your ____

106. Shut up and ____

107. The ____ are coming.

108. The ____ that roared

109. The ____ type thing.

110. The hell with ____

111. The____ are taking over.

112. Think outside the ____

113. Using the _____ word

114. What am I? ____

115. What's up with _____

116. When ____ goes bad

117. When the _____ comes

118. Why Johnny can't _____

119. You dog of ____

120. You want a _____?

121. ____ as _____does.

122. ____ is my middle name.

123. ____ out and _____ somebody

124. ____ to end all ____

125. _____ is the new _____

126. All those ____look alike.

127. All_____all the time

128. And that man's a _____.

129. Friends don't let friends ____

130. He's a _____ among _____

131. I (he) eat(s) and breathe(s) _____

132. I eat _____ for breakfast.

133. I wouldn't be caught dead _____

134. I'll give you a ____

135. I'm (not) a ____ person

136. If you believe that, ____

137. It's not ____, it's _____

138. It's nothing if not _____

139. Leave the ____ at home

140. my ____ right or wrong.

141. My middle name is _____

142. No one teaches me ____

143. None of this ____ business

144. Not the way I ____.

145. Tell it to (the) ____

146. The ____ behind the ____

147. The ____ de tutti ____

148. There's _____ and there's _____

149. When _____ is not enough

150. You call that a ____?

151. and I do mean _____

152. He makes a mean _____

153. ____ gives you a bad name.

154. A _____ walked into a bar.

155. Do I look like a ____ ?

156. He is too ____ by half

157. I'm not a big ____ person

158. If you _____ they will come.

159. Is that (a)_____ or what?

160. It was (a)____ from hell.

161. Keep your eye(s) on the _____

162. Make like a _____ and ____.

163. So many ____, so little ____

164. So you think you can ____

165. That gives _____ a bad name

166. That was voted the most _____

167. The proof is in the ____

168. There's nothing _____ about it.

169. Wadda I look like, a ____ ?

170. Where in the _____ is ____.

171. Yes, Virginia, there is a ____

172. You're like a ____ to me.

173. You've got to love the ____

174. This is the sound of _____

175. ___,here, ___there, ____everywhere

176. A ___ to end all ____

177. One more ____ than the other

178. ____ is not just another pretty face.

179. ____ isn't just another _____ for ____

180. _____ is just another word for _____

181. A _____ does not a _____ make.

182. Changing ____ one _____ at a time.

183. Do you know where your_____ is (are)?

184. Have you ever seen a ____ _____ing

185. I can do _____ in my sleep.

186. I'm on that like ____on ____

187. It's(he's, she's) a little too _____ by half

188. One man's _____ is another man's _____

189. Some of my best friends are ____

190. That ___ isn't going to ___ itself.

191. That was a _____ and a half

192. To think I was once (a) ____

193. We know _____ when we hear (see) it

194. What happens in _____ stays in _____.

195. What part of _____ don't you understand?

196. Who (what) do I look like? A ____?

197. With ____ like these, who needs ____

198. He's not the _____ _____ in the ____.

199. I can do _____ with my eyes closed.

200. I wouldn't give you _____ for his ____

201. That's a _____ only a _____ could love

202. The ____ is the enemy of the _____

203. What do you take me for? A _____?

204. What if _____ is what it's all about?

205. You can take (your)_____ and shove it.

206. You've seen one ____, you've seen them all.

207. ____ is my name and _____ is my game.

208. ____ is not the ____est ____in the _____

209. I know ____ like the back of my hand.

210. If you had his/my _____, you'd be ____(-ing) too.

211. What? Do I look like a ____ to you ?

212. You can say hello to_____, goodbye to ____

213. ____ is a few ____short of a full ____

214. A ___ without____ is like a _____ without ____

215. A funny thing happened on the way to the _____

216. It's not just about (the) ____; it's about (the) ____

217. This is your brain. This is your brain on ____

218. I can do ____ with one hand tied behind my back.

219. You (I) must have been absent when they handed out the ____

220. _____: You can't live with them (it), and you can't live without them (it).

221. I may not know much about ____, but I know what I like.

222. Ask not what ____ can do for you, ask what you can do for _____.

223. 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

1. All _____ trees look alike.
2. My bag is _____.
3. If you want the _____, just ask.
4. The players are _____!
5. I can _____ with my eyes closed.
6. I missed the _____.
7. There is a _____ waiting for you.
8. _____ is my middle name.
9. A stitch in time _____ nine.
10. It takes two to _____.

Highlights

- Formulaic expressions form a significant portion of everyday verbal communication
- Schemata are fixed formulaic expressions with a mandatory open slot for novel words
- Schemata show interplay of fixed phrases and novel words
- Formulaic expressions are recognized by native speakers across two generations

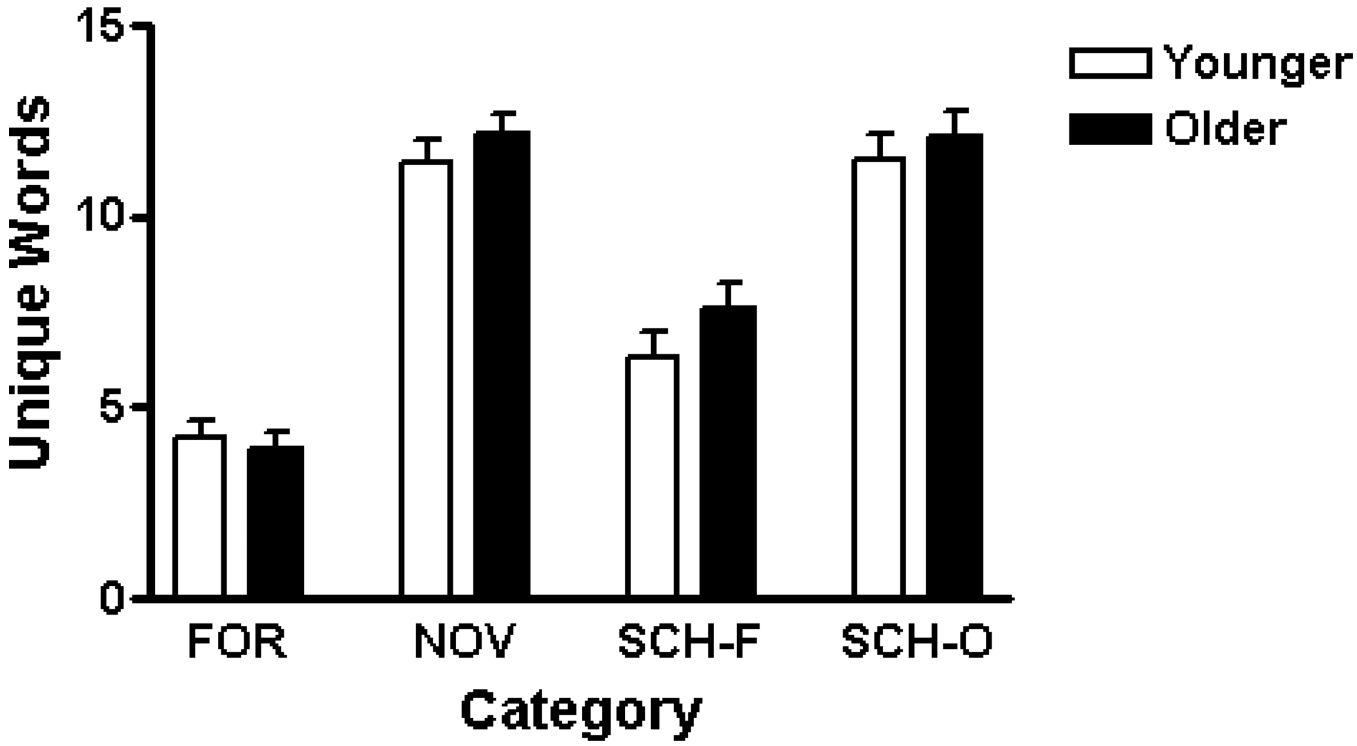

Figure 1.

Mean number and standard error of unique words produced by category (FOR = formulaic, NOV = novel, SCH-F = schemata fixed, SCH-O = schemata open) by younger and older native speakers of American English.

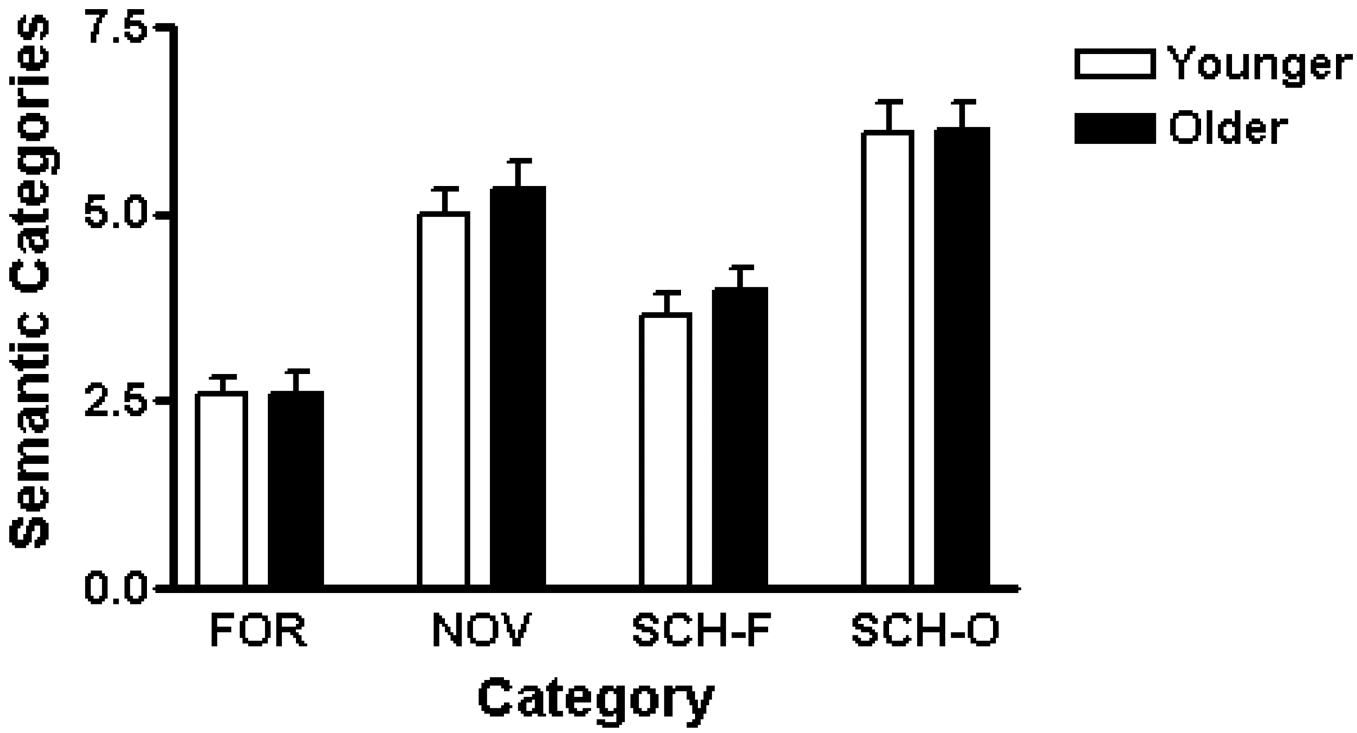

Figure 2.

Mean number and standard error of semantic categories produced by category (FOR = formulaic, NOV = novel, SCH-F = schemata fixed, SCH-O = schemata open) by younger and older native speakers of American English.

Table 1

Example sentences and answers for formulaic, novel, schemata-fixed, schemata-open stimuli3.

Type	Example sentence	Examples of Responses
Formulaic	It takes two to _____.	*tango* (Target), *dance*
	A _____ in disguise.	*blessing* (Target), *angel*, *fool*, *man*
	_____ and let live.	*live* (Target), *love*, *eat*, *go*
	It's way over my _____.	*head* (Target), *budget*, *headache*
Novel	My bag is _____.	*full*, *heavy*, *black*, *big*
	That _____ was very helpful.	*lesson*, *person*, *map*, *advice*
	_____ make a mess.	*they*, *birds*, *don't*, *go*
	The _____ covered my face.	*beard*, *chocolate*, *scarf*, *mud*
Schemata-fixed	That was a workout and a _____.	*half* (Target), *pleasure*, *challenge*, *joy*
	How awesome is _____.	*that* (Target), love, life, pizza
	_____ that's a party.	*now* (Target), *well*, *man*, *like*
	There's _____ loud about it.	*nothing* (Target), *something*, *no*, *some*
Schemata-open	I don't do _____.	*sarcasm*, *apologies*, *heights*, *Mondays*
	It's nothing if not _____.	*everything*, *sincere*, *critical*, *old*
	_____ like nobody's business.	*party*, *cook*, *stinks*, *boogie*
	I eat _____ for breakfast.	*chumps*, *eggs*, *success*, *danger*

Table 2

Semantic categories and examples.

Semantic category	Word examples
1. Physical world	*world, water, clouds, stars*
2. Mankind	*man, woman, she, children, father, Ann*
3. Animals	*bear, dog, cat, bird, beetle, spider*
4. Body parts and functions	*arm, leg, life, death, sickness, medicine, head*
5. Food and drink	*apple, burger, drink, dinner, McDonalds, fork*
6. Clothing and adornment	*shirt, dress, suit, shoes, makeup, shave*
7. Dwellings and furniture	*couch, bed, home, chair, lamp, rug*
8. Agriculture and vegetation	*crops, flowers, grass, rose, leaves*
9. Physical acts and materials	*art, break, glass, bricks, rope, machine, don't*
10. Motion and transportation	*run, walk, car, drive, bus, train, traffic light*
11. Possession and trade	*prize, mine, yours, package, send, receive, give*
12. Spatial relations	*up, thicker, in, long, tall, here, somewhere*
13. Quantity and number	*one, many, all, most, ten, half*
14. Time	*fast, slow, noon, Monday, year, September*
15. Sense perception	*look, soft, hot, blue, red, color, cool*
16. Emotion	*happy, sad, angry, smile, kiss, love*
17. Mind and thought	*think, reason, knowledge, plan, attention*
18. Language and music	*talk, write, book, music, jazz, literature, sing*
19. Social relations	*king, employer, waiter, master, boss, princess*
20. Warfare and hunting	*battle, war, fight, trap*
21. Law and judgment	*voted, jury, judge, law*
22. Religion and beliefs	*heaven, hell, God, prayer, angels, witches, ghosts*
23. Nonhuman pronouns, nonspecific pronouns, indefinite pronouns	*this, that, those, something, nothing, everything, it, some*
24. Function words	articles (*the, a*), auxiliary verb (*is*), infinitive (*to*), conjunctions (*and, if*), copula (*is*)
25. Leisure: Entertainment, sports, games	*play, toy, win, lose, game, basketball, ball, zoo, museum, party*
26. Electronics/technology	*computer, remote, microphone, TV, phone, cell*
27. Expletives (if used as exclamation—not literally)	*hell, damn, freak, darn*
28. Discourse elements	*well, like, just*
29. Negations	*no, not*

Note: Semantic categories 1–22 were taken from Fields (2013) Indo-European Semantic Fields. Semantic categories 23–29 were developed as part of this work.

Table 3

Mean number of raters who correctly identified the target word for formulaic and schemata-fixed utterances.

	Younger (*n* = 20)	**Older (*n* = 20)**	**Total (*N* = 40)**
Formulaic	15.65 (*SD* = 4.61)	16.23 (*SD* = 4.81)	31.93 (*SD* = 8.85)
Schemata-fixed	10.33 (*SD* = 7.08)	9.42 (*SD* = 6.71)	19.75 (*SD* = 13.51)

Table 4

Mean number of unique words provided for each utterance type by younger, older, and the total rater group.

	Younger (*n* = 20)	**Older (*n* = 20)**	**Total (*N* = 40)**
Formulaic	4.18 (*SD* = 3.10)	3.88 (*SD* = 3.15)	6.20 (*SD* = 4.82)
Novel	11.45 (*SD* = 3.43)	12.15 (*SD* = 3.29)	19.10 (*SD* = 5.77)
Schemata-fixed	6.35 (*SD* = 3.86)	7.57 (*SD* = 4.27)	11.05 (*SD* = 6.64)
Schemata-open	11.50 (*SD* = 4.03)	12.13 (*SD* = 4.23)	19.80 (*SD* = 7.55)

Table 5

Mean number of semantic categories represented in each utterance type by younger, older, and the total rater group.

	Younger (*n* = 20)	**Older (*n* = 20)**	**Total (*N* = 40)**
Formulaic	2.60 (*SD* = 1.50)	2.60 (*SD* = 1.75)	3.40 (*SD* = 2.10)
Novel	5.00 (*SD* = 2.00)	5.35 (*SD* = 2.18)	6.90 (*SD* = 2.60)
Schemata-fixed	3.65 (*SD* = 1.83)	3.98 (*SD* = 2.03)	5.18 (*SD* = 2.62)
Schemata-open	6.10 (*SD* = 2.45)	6.13 (*SD* = 2.19)	8.13 (*SD* = 2.77)

Table 6

Examples of creative proliferation of phrases based on the advertising slogan “Reach out and touch somebody” identified with different species of animals

leap out and grab somebody
lumber out and charge somebody
gallop out and kick somebody
pounce out and eat somebody
buzz out and sting somebody
lurch out and squash somebody
crawl out and bite somebody
slither out and wrap somebody
bounce out and bash somebody
sweep down and seize somebody
plunge out and ram somebody