



Published in final edited form as:

Ment Lex. 2017 January ; 12(2): 234–262. doi:10.1075/ml.16019.ram.

Perception of formulaic and novel expressions under acoustic degradation

C. Sophia Rammell¹, Diana Van Lancker Sidtis^{2,3}, and David B. Pisoni^{1,3}

¹ Indiana University

² Indiana University School of Medicine

³ New York University

Abstract

Background: Formulaic expressions, including idioms and other fixed expressions, comprise a significant proportion of discourse. Although much has been written about this topic, controversy remains about their psychological status. An important claim about formulaic expressions, that they are known to native speakers, has seldom been directly demonstrated. This study tested the hypothesis that formulaic expressions are known and stored as whole unit mental representations by performing three perceptual experiments.

Method: Listeners transcribed two kinds of spectrally-degraded spoken sentences, half formulaic, and half novel, newly created expressions, matched for grammar and length. Two familiarity ratings, usage and exposure, were obtained from listeners for each expression. Text frequency data for the stimuli and their constituent words were obtained using a spoken corpus.

Results: Participants transcribed formulaic more successfully than literal utterances. Usage and familiarity ratings correlated with accuracy, but formulaic utterances with low ratings were also transcribed correctly. Phrase types differed significantly in text frequency, but word frequency counts did not differentiate the two kinds of expressions.

Discussion: These studies provide new converging evidence that formulaic expressions are encoded and processed as whole units, supporting a dual-process model of language processing, which assumes that grammatical and formulaic expressions are differentially processed.

Keywords

formulaic language; corpus analysis; speech perception; dual-process model

Interest in formulaic language has grown in the past few decades. Scholars use an array of terms and have described various categories of formulaic expressions, along with their characteristic properties (Van Lancker & Rallon, 2004; Wray, 2002; Wulff, 2008). Formulaic expressions include conversational speech formulas, idioms, proverbs, pause fillers,

Corresponding address: Diana Van Lancker Sidtis, Department of Communicative Sciences and Disorders, New York University, NY10012, New York, diana.sidtis@nyu.edu.

¹Many of these phrases have a specific prosodic shape and vocal quality, as in the case of the phrase, “whatever,” being pronounced as “what-EV-ah” (Ashby, 2006; Lin, 2010; Van Lancker Sidtis, 2012a; Hallin & Van Lancker Sidtis, 2014).

counting, swearing, and other conventional and multiword units. Some examples are *He's got his head in the clouds*, *I'll get back to you later*, *Cat got your tongue?*, and *Gosh darn it*. Despite differences between these expressions, all have two important characteristics in common: they are not newly created utilizing grammatical rules to combine words and they are known to speakers of a language community. Detailed linguistic analyses of formulaic expressions reveal that formulaic expressions and other multiword expressions form categories along continua (Tannen & Öztekin, 1981; Van Lancker, 1975; Kecskes, 2003, 2007; Wray, 2002; Ellis, 2012) in association with such characteristics as attitudinal and emotional nuance, degree of cohesion, more or less nonliteral meaning, dependency on context, optional or obligatory status in social contexts, and semantic transparency (Jarema, Busson, Nikolova, Tsapkin, & Libben, 1999). Nonetheless, these expressions, despite their differences, have in common that they are not newly created and they are recognized as known by speakers in a community.

Formulaic expressions make up a large part of language use, with estimates of proportions in normal discourse from 25–70% (Foster, 2001; Hill, 2001; Van Lancker Sidtis, 2014) and total counts between 100,000 and 300,000 (Jackendoff, 1995; Kuiper, 2009). They perform an assortment of communicative functions, including conveying nonliteral meanings and cultural memes, humor, interpersonal bonding, attitudinal and emotional expression, sociological group identity, and language play. Understanding how formulaic expressions are processed, stored, and retrieved from memory can contribute to more complete models of language processing.

Formulaic expressions typically have a stereotyped form,¹ conventionalized meaning (usually beyond the direct lexical meaning), and an appropriate context (with requirements for formality and register), all of which are immediately recognizable to native speakers of a language (Fillmore, 1979; Pawley & Syder, 1983; Kuiper, 2006). Second language speakers, in acquiring the form, meaning, and contextual contingencies of formulaic expressions, face a considerable challenge (Paquot & Granger, 2012); producing a formulaic expression with a replaced lexical item or nonstandard prosodic contour is generally taken to be a second language speaker error or a humorous gesture (Kuiper, 2007; Bell, 2012; Millar, 2011). Child language acquisition schedules differ for novel and formulaic expressions (e.g., Gleason & Weintraub, 1976; Gleason, 1980; Kempler, Van Lancker, Marchman & Bates, 1999; Nippold, 1998; Peters, 1983; Locke, 1993; Perkins, 1999). Evaluation and treatment in speech-language pathology are best informed by distinguishing between loss and rehabilitation of formulaic or novel language (Van Lancker Sidtis, 2012b, 2014; Stahl & Van Lancker Sidtis, 2015).

In the linguistic sciences, formulaic expressions, or “formulemes,” have been studied using surveys and sentence completion tasks (Van Lancker & Rallon, 2004; Van Lancker Sidtis, Cameron, Bridges, & Sidtis, 2015), word association (Clark, 1970), interpretation and recognition (Gibbs, 1980; Libben & Titone, 2008; Cutting & Bock, 1997; Osgood & Housain, 1974; Van Lancker Sidtis, 2003), language acquisition schedules (Nippold, 1998; Reuterskiöld & Van Lancker Sidtis, 2012; Pickens & Pollio, 1979), auditory/acoustic measures (Van Lancker, Canter & Terbeek, 1981; Lieberman, 1963; Yang & Van Lancker Sidtis, 2016), and speech errors (Kuiper, Van Egmond, Kempen, & Sprenger, 2007;

Nooteboom, 2011). A variety of psycholinguistic designs, including eye tracking and response times, have aimed at discovering principles that distinguish the two kinds of language, formulaic and newly created (e.g., Siyanova-Chanturia, Conklin, & Schmitt, 2011; Underwood, Schmitt, & Galpin, 2004; Swinney & Cutler, 1979). Corpus linguistics and computational approaches have focused on collocation frequencies (Moon, 1998a,b; Biber, 2009; Conrad & Biber, 2004) and mutual information scores (Lin, 1999; Lin & Adolphs, 2009; Paquot & Granger, 2012; Wulff, 2008). Corpus linguistic approaches have used Latent Semantic Analysis (Schone & Jurafsky, 2001) and semantic similarity measures (Bannard, Baldwin, & Lascarides, 2003).

Many studies have suggested that formulaic expressions have unitary structure. Early studies revealed differences in pronunciation and perception between matched novel and formulaic exemplars (Lieberman, 1963; Van Lancker et al., 1981). As part of their stereotyped form, formulaic expressions have been shown to exhibit phonological coherence, which may be thought of a surrogate indicator of holistic structure (Hallin & Van Lancker Sidtis, 2015). Lin and others (Lin, 2010; Lin & Adolphs, 2009) have proposed that these expressions form a single intonation unit. In similar fashion, previous research has shown that formulaic expressions, under controlled conditions, are uttered faster and more fluently than novel language (Erman, 2007; Lin, 2010; Wray, 2002; Van Lancker et al., 1981; Hallin & Van Lancker Sidtis, 2015; Tabossi, Fanari & Wolf, 2009), again suggesting unitary structure. Other distinguishing characteristics are loudness, distinctive voice quality, and temporal cues such as initial shortening and phrase final lengthening (Yang, Ahn, & Van Lancker Sidtis, 2015; Yang & Van Lancker Sidtis, 2016). These studies address the structure and physical characteristics of formulaic expressions, but do not directly probe knowledge of the expressions and their place in memory or information processing. Numerous studies have probed various constituent and usage properties of one important category of FEs, idioms (Cacciari & Tabossi, 1988; Nunberg, Sag, & Wasow, 1994), referencing their varying properties such as literalness and transparency (Titone & Connine, 1994, 1999). These studies have led to proposals of mental representation that distinguish holistic, word like storage of non-decomposable subtypes from a configurational format (Caillies & Butcher, 2007). However, many of characteristics inhering in idioms are not viable for other kinds of FEs. This study utilized a range of FEs, including idioms and other kinds (see methods below), that differ from novel expressions in only one parameter: they, as a unit, are known to speakers.

The properties of the broad constituency of formulaic language are well accounted for by the proposal of a dual-process model of language (Van Lancker Sidtis, 2012a; Wray & Perkins, 2004; Erman & Warren, 2000). In this model, formulaic expressions and newly created, novel expressions differ in how they are learned, processed, and stored. Novel expressions are processed and analyzed in real time using stored lexical and morphological units organized according to grammatical rules. Formulaic expressions, in contrast, at some level of mental representation, may be accessed from stored traces as whole, precompiled units (Horowitz & Manelis, 1973; Osgood & Housain, 1974). Related to the dual-processing model is the “hybrid” model, which suggests that idioms may have at least two kinds of representations, one in holistic profile and another in compositional form (Sprenger, Levelt, & Kempen, 2006). This view accommodates some psycholinguistic results that show

abilities of language users, in experimental studies, to process elements of constituency of formulaic expressions at phonological and lexical levels. Yet in these studies, the status of many kinds of formulaic expressions as relatively unitary in some stage or level of mental representation is attested (Osgood & Housain, 1974; Swinney & Cutler, 1979; Conklin & Schmitt, 2008; Horowitz & Manelis, 1973; Siyanova-Chanturia et al., 2011). While any verbal object can be decomposed in various ways, the hybrid view seems to provide the best characterization for idioms: “idioms are represented and retrieved as units that can interact” with compositionality and other factors (Libben & Titone, 2008; p. 1117). This general perspective forms a foundation also for all the larger class of FE variants utilized in this study.

The hypothesis tested in the present set of experiments is that formulaic expressions are known to the native speaker, and that they are known (stored) as single, holistic units in at least one level or stage of mental representation (Bolinger, 1976, 1977). Specifically, we predicted that, under acoustic degradation conditions, formulaic expressions will be correctly perceived more often than novel expressions because they are familiar and are stored in long-term memory as unitary holistic units. Exposure to degraded and incomplete perceptual information will suffice to elicit the associated, stored unitary form. Further, transcribed responses will fit the original, entire formulaic target more accurately than the matched, original complete novel target. For this report, the formulaic expressions chosen for study are conversational speech formulas, idioms, proverbs, lexical bundles, and other conventional expressions (see Appendix A). Our interest here is to establish empirical evidence for the proposal that native speakers know and process these expressions in a way that makes them fundamentally different in mental representation from newly created, novel utterances. Three experiments were conducted.

Experiment 1

Method

Subjects—Participants were native speakers of English with no known speech or hearing disorders at the time of testing. Twenty-two subjects ($F=9$, $M=13$) completed a two-part experimental protocol. The mean age of the subjects was 18.9 years, with a range of 18–21 years. Participants in all three experiments were recruited using the Indiana University Psychological and Brain Sciences Departmental Volunteer Subject Pool. Subjects were all undergraduate students at Indiana University enrolled in introductory psychology classes.

Stimuli—The stimulus materials consisted of 140 meaningful English spoken sentences, produced with natural expression by a native speaker of American English. Half of the sentences were formulaic and half were novel sentences matched for lexical syllable structure, length and grammatical construction (Appendix A). Forty-four sentences were taken from the Familiar and Novel Language Comprehension task (FANL-C) (Kempler & Van Lancker, 1996). The remaining sentences were selected from a matched idiom-novel expression list compiled for use in a previous study. Of the formulaic set, the stimuli fell into several categories. About half were classical idioms (e.g., *That’s the way the cookie crumbles*, *Straight from the horse’s mouth*). The rest consisted of proverbs (*When the cat’s*

away, the mice will play; Don't burn your bridges behind you), lexical bundles (*None of your business; On the other hand*), conversational speech formulas (*And now for something completely different; I should be so lucky*), and numerous other conventional utterances (*A little of that goes a long way; You never had it so good; That's for me to know and you to find out; No sooner said than done*). This heterogeneous array of FEs have in common that they are not newly created and, in our view, they are known to speakers in their canonical ("formuleme") form.

Analyses were performed to determine the text frequency of the expressions and their constituent words. Spoken corpus data from the Corpus of Contemporary American English (COCA) were used to obtain and compare whole phrase frequency and individual content word frequency from the phrases for a subset of 41 pairs of test expressions (used in Experiment 3; Davies, 2008). The median frequency in COCA of the entire set of formulaic phrases was 3, and the median frequency of novel phrases was 0.073. The mean frequency for formulaic expressions, 102.24, was strongly influenced by one outlier ("On the other hand," frequency = 3832). An analysis was also performed to assess raw frequencies across both types of expressions. For each expression and for the sum of the content word frequencies in each expression, a measure of $\ln(\text{frequency} + 1)$ was calculated (Baayen & Hendrix, 2011; Baayen, Milin, Durdevic, Hendrix, & Marelli, 2011) (See Appendix B).

A repeated-measures ANOVA performed on the natural log measure of the frequency values of the content words and the whole phrase by type of expression revealed that the frequency of occurrence of the individual content words in both formulaic and novel expressions was the same. However, the types of expressions differed in frequency: the formulaic expressions occurred more frequently in the COCA corpus than the novel expressions. A significant interaction of type of expression and expression versus content word frequency was obtained in the ANOVA, $F(1, 40) = 21.681, p < 0.001$. (please see Appendix B).

There are many different ways of degrading a speech signal to reduce performance such as filtering or using white noise or multi-talker babble (Pisoni, 1996). In the present study, an acoustic simulation of a four-channel cochlear implant was used (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). Cochlear-implant simulated speech is an easy way to manipulate the intelligibility of the speech signal over a range from very degraded to less degraded. The .wav files of all of the stimuli were processed using Angel Sim to create a 4-channel cochlear implant simulation of sinewave vocoded speech (TigerCIS 2012). The original speech signal was first band-pass filtered into four frequency bands based on Greenwood's function at a frequency range of 200–7000 Hz and filter slope of 24 dB/oct (Greenwood, 1990). The input and output signals were matched in frequency range and filter slope. Then the amplitude envelope was derived from each filter band using a low-pass filter with a cutoff frequency of 160 Hz and a roll off of 24 dB/oct. Residual spectral information was removed and replaced with either white noise or sine waves (Dorman & Loizou, 1997, 1998). This form of vocoded speech maintains the original speech envelope but removes the temporal fine structure. This approach to speech signal degradation was used in the present studies because it reduced speech intelligibility to levels close to the threshold of identification accuracy (Shannon et al., 1995; Shannon, Fu, & Galvin, 2004).

Procedure—All three experiments consisted of two parts: a sentence recognition task and a familiarity rating task based on usage or exposure. The participants were naïve to the purpose of the study and IRB procedures were followed. Prior to beginning Experiment 1, participants were informed that the sentences they would hear over their headphones had been processed by a computer and the speech would sound degraded. Five practice sentences taken from another test protocol were played first to familiarize the listeners with spectrally degraded speech (Nilsson, Soli, & Sullivan, 1994). Listeners then completed a speech recognition transcription task with the test stimuli under a 4-channel acoustic simulation. Stimuli were presented in random order. Listeners heard each sentence only once and were asked to type what they heard at the end of each presentation using a computer keyboard. Transcription tasks have a venerable history in providing a “window” onto mental knowledge for speech and language and for assessing listeners’ abilities to process speech and language samples.

In the second phase, the participants heard the same set of sentences again in unprocessed form, one time each, in a different random order, for a rating task. Listeners rated each utterance on a scale of 1 to 3 (1 = I never use this sentence, 2 = I sometimes use this sentence, 3 = I often use this sentence). The entire experiment took between 45 minutes and one hour.

Scoring—Two methods of scoring were used. First, the transcription of the whole phrase responses was scored as correct or incorrect by whether all keywords were present in the correct order. Second, analyses of total words correct in each phrase were calculated. Usage ratings were scored on a scale of 1–3.

Results

Overall, participants correctly transcribed entire formulaic expressions more often than novel expressions under acoustic degradation (Figure 1). Formulaic expressions were correctly transcribed in 57.9% of cases; novel expressions were correctly transcribed in only 32.7% of cases. The difference was significant using a paired-samples t-test, $t(21) = -12.95$, $p < 0.001$. Given that the utterances were carefully matched and the constituent words did not differ in frequency, this is a large difference in performance.

Twenty-two out of 22 participants showed the predicted effect. To assess the individual variation on this task, difference scores in percent correct between formulaic and novel expressions were calculated for each subject (Figure 2). The magnitude of the difference scores ranged from 2 to 38. When the phrases were rescored by total number of words correct from each phrase, the difference was also significant using a paired-samples t-test, $t(2652) = -91.817$, $p < 0.001$.

Expressions selected as “never use” had a mean transcription accuracy of 37.3%, “sometimes use” a mean accuracy of 64.3%, and “often use” a mean accuracy of 66.1%. Paired-samples t-tests revealed statistical differences between percent correct averages for stimuli reported as “I never use this sentence” and “I sometimes use this sentence”, $t(17) = -10.31$, $p = 0.000$, and “never use” and “often use”, $t(16) = -3.58$, $p = 0.002$. The difference between “sometimes use” and “often use” was not significant (Figure 3). The lack of a

significant difference between “sometimes use” and “often use” could arise from participants’ reluctance to classify phrases as “often use”. “Never use” was chosen most frequently by subjects, 67.3% of the time. “Sometimes use” was selected for 24.4% of expressions, and “often use” was very infrequently chosen (8.3%). This can be explained by a description of formulaic language of consisting of a very large available repertory, only a small fragment of which is actively used by the individual user. These usage subsets differed across language users. These results suggest that many more formulaic expressions are known and can be recognized than are actively used.

Importantly, the usage ratings employed in Experiment 1 differed by type of expression. Formulaic expressions were more frequently rated as “often use” (14.6%) or “sometimes use” (37.3%) than novel expressions, which were rated as “sometimes use” in 11.4% or “often use” in 2.0% of cases. On the other hand, novel expressions were significantly more frequently rated as “never use” at 86.6%, while formulaic expressions were rated as “never use” in 48.1% of cases ($\chi^2 = 428.892$, $df = 2$, $p = 0.000$.) (Figure 4).

Beyond a relationship with usage ratings, formulaic expressions were always transcribed more accurately than novel expressions, regardless of usage rating (Figure 5). A logistic regression analysis was conducted to predict whole phrase transcription accuracy using usage ratings and expression type as predictors. Both predictors were significant: usage rating ($\chi^2 = 89.526$, $df = 1$, $p < 0.001$) and expression type ($\chi^2 = 67.542$, $df = 1$, $p < 0.001$). Figure 5 also shows that 50% of formulaic expressions rated as “never use” were correctly recognized, again implying a large repertory of known expressions independent of usage ratings.

Discussion

As hypothesized, listeners correctly transcribed spectrally-degraded formulaic expressions more often than novel expressions. Subjects also reported higher usage ratings for formulaic than novel expressions. These results are consistent with predictions based on a dual-process model of language (Lounsbury, 1963; Erman & Warren, 2000; Perkins, 1999; Sinclair, 1987; Van Lancker Sidtis, 2012a). According to this model, formulaic language is processed and stored differently than novel language. These results also indicate that subjects transcribed many expressions correctly that were rated as “never use.” This observation led to a restructuring of the familiarity rating task in the next experiment reported below.

Experiment 2

A second experiment was performed to ensure that subjects could transcribe these phrases correctly without any acoustic degradation. We expected that expressions which are not spectrally degraded would be correctly transcribed at ceiling performance levels. We replaced the usage scale of familiarity, used in Experiment 1, with a 7-point exposure rating scale.

Methods

Subjects—Participants were all native speakers of English with no known speech or hearing disorders at the time of testing. They were recruited from the Indiana University

Volunteer Subject Pool. Nineteen new subjects ($F=12$, $M=7$) who did not participate in Experiment 1 completed the task. The average age was 19 years, with a range of 18–22 years.

Stimuli—The stimuli were the same matched, paired 140 spoken sentences used in Experiment 1.

Procedure—Subjects completed a two-phase experiment. The first phase was identical to the first phase of Experiment 1. In the second phase, the participants heard the same 140 sentences again, one time only, in a different random order, and rated each sentence on a scale of 1 to 7 (1 = Never, 7 = Often) based on how often they had heard the expression before in their life.

Scoring—Transcription responses were scored by the whole phrase correct. Familiarity ratings were scored using a scale of 1–7.

Results

Without acoustic degradation, subjects correctly transcribed novel expressions 96.2% and formulaic expressions 97.5% of the time. The range for novel expressions was 52.6%–100%, and the range for formulaic expressions was 78.9%–100%. Subjects differed significantly in their familiarity ratings by type of expression. A paired-samples t-test on the mean scores established that subjects gave significantly higher exposure ratings to formulaic expressions than to novel expressions, $t(2379) = -77.352$, $p < 0.001$ (Figure 6).

Even when a seven-point familiarity scale was used, subjects were reluctant to select a high exposure rating for any phrase, regardless of whether it was formulaic or novel. Subjects selected “never heard” (1) more often than “often heard” (7).

Discussion

As hypothesized, without acoustic degradation listeners correctly transcribed both formulaic and novel expressions at ceiling levels of performance. Participants also gave higher exposure ratings to formulaic expressions than to novel expressions. Nearly all the expressions were endorsed as familiar by all the subjects, and very few sentences were rated as unfamiliar.

Experiment 3

Because the subjects in Experiment 1 performed well on both types of sentences in the sentence recognition task under a four-channel acoustic simulation, a third experiment was carried out to assess transcription accuracy using sentences under more degraded conditions. Previous studies have shown that formulaic and matched novel expressions differ consistently in prosodic cues. This experiment served as a boundary condition for perception of formulaic and novel expressions.

Methods

Subjects—Twelve newly recruited listeners ($F = 7$, $M = 5$) meeting the same selection criteria of the two other studies completed the new transcription task. The average participant age was 19.75 years, with a range of 18–22 years.

Stimuli—The stimuli were 82 spoken sentences taken from the original set of stimulus materials. The subset of sentences was selected based on the transcription accuracy and exposure ratings obtained in Experiment 2. Any sentence which had a mean transcription accuracy without acoustic degradation lower than 94.7% was eliminated. This percentage cutoff value was chosen because it indicated that only one person incorrectly transcribed the sentence. Formulaic expressions with mean familiarity ratings of three or less were also eliminated. Finally, after these formulaic sentences were eliminated, their matched pair novel sentence was also eliminated to equate expression type.

Procedure—Participants completed a two-phase experiment. The first part was identical to the initial phase of Experiments 1 and 2, with 5 practice items preceding a transcription task, except that in this experiment, all stimuli were played using one-channel acoustic degradation. This mode of transformation preserves only some gross prosodic information (durations and amplitude changes) in the signal, while eliminating temporal fine structure. In the second phase, the participants heard the same sentences in unprocessed form a second time in a new random order. Listeners rated the sentences on a scale of 1 to 7 (1 = Never, 7 = Often) based on how often they had heard the expression before in their life.

Scoring—Transcription task responses were scored based on whether the whole phrase was correct. Familiarity ratings were scored using the rating scale of 1–7.

Results

Listeners were unable to transcribe any of the formulaic or novel expressions correctly. As in Experiment 2, familiarity ratings differed by type of expression. A paired-samples t-test established that higher familiarity ratings were given to formulaic expressions than to novel expressions, $t(983) = -48.574$, $p < 0.000$. As in Experiment 2, when participants gave a low familiarity rating, it was more likely that the expression was novel than formulaic. When subjects gave a high familiarity rating, it was more likely that the expression was formulaic than novel (Figure 7).

Discussion

The results of the sentence transcription task under one-channel acoustic degradation conditions demonstrated that listeners were unable to make use of any of the acoustic-phonetic information in the signals to identify words in the sentences. Selected prosodic cues, consisting mainly of durational and amplitude information, which were preserved under these degraded presentation conditions, were not sufficient to support reliable word recognition in sentences. It is possible that degradation preserving sentence intonation characteristics, as well as duration and amplitude, would yield higher recognition performance.

The familiarity rating results from this experiment replicated the findings obtained in Experiment 2. As expected, formulaic expressions were given higher familiarity ratings than novel expressions. The high familiarity ratings reflect prior knowledge of the expressions.

Summary of perceptual results

In summary, subjects correctly identified formulaic expressions more often than novel expressions under four-channel acoustic simulation. However, under one-channel simulation conditions they were unable to identify any of the formulaic or novel expressions. Listeners identified expressions they reported that they used more often with higher accuracy than those they reported to have used less often under four-channel CI-simulation, but 50% of formulaic rated as “never use” were also correctly identified. Subjects also reported they use the formulaic expressions more often than the novel expressions. In support of our original predictions, formulaic expressions were identified more often than novel expressions and usage and exposure ratings differed for the respective expression sets (formulaic and novel).

General discussion

Subjects transcribed formulaic expressions, when scored by entire unit and by correct constituent words, more accurately than novel expressions under acoustic degradation, and they gave higher usage and exposure ratings to formulaic expressions as well. Self-rating of use and exposure, as well as text frequencies, are of interest, but they are only minimally revealing; they are also not good measures of speaker’s potential or veridical knowledge (familiarity with) of FEs. Our data substantiate this perspective, based on the “one-trial” learning theory of FE acquisition, which allows for rapid uptake of FEs into mental representation because of their anomalous nature. Persons were successful at recognizing FEs rated high on use and exposure and also those not rated as high. As would be predicted, corpus analyses of the individual words and sentences provided additional converging support showing the higher frequency of entire formulaic expressions in the language compared to novel (by definition) compositional phrases. Spoken corpus data show high formulaic expression frequency and low novel expression frequency even when individual content word frequency was equated across both types of expressions.

FEs have been found to carry stereotypical prosodic information (Lin, 2010; Hallin & Van Lancker Sidtis, 2015; Yang et al., 2015; Van Lancker et al., 1981; Yang & Van Lancker Sidtis, 2016; Ashby, 1992) which might have provided cues to the listeners under extreme acoustic degradation. The second experiment featured an aggressively masked signal, which apparently did not allow prosodic cues sufficient for recognition of the phrase. Please note that the set of FEs was heterogeneous in durations and in FE classification. A post hoc acoustic analysis of the formulaic and novel stimuli presented in this study using Praat software (Boersma & Weenink, 2007) was conducted, comparing phrase and syllable durations, mean fundamental frequencies (F0), and F0 ranges between the two types of expressions, formulaic and novel. T-tests revealed no significant differences between paired utterance types for syllable durations or the two F0 measures. Phrase durations between the two sets of stimuli did differ significantly ($t = 4.2041$, $df = 69$, $p < 0.0001$), with FEs averaging 2.06 sec, and novel utterances 2.27 secs. However, given the varied durations of the matched pairs utilized in this study (length was not an independent variable) and their

random presentation to the listeners, it could not be expected that duration information would suffice to provide better recognition for one or the other kind of language.

Results from the three experiments and corpus analyses showed that formulaic expressions in language differ in fundamental ways from novel expressions. We propose that formulaic expressions are stored and processed holistically in perception as single units in at least one level of representation in memory. Our findings support the notion “multiword sequences leave memory traces in the brain” (Tremblay, Derwing, Libben, & Westbury, 2011, p. 569).

Following James (1885), Bousfield (1953) and Miller (1956), Simon’s (1974) article demonstrated the important role of “chunks” as units of processing in human memory. When people are asked to remember and recall a spoken list of items, the number of recallable items – or chunks – is nearly the same for syllables, words, compound words, and idiomatic expressions, a finding replicated for Chinese (Simon, Zhang, Zang, & Peng, 1989). These results are supported in the visual mode by studies revealing significant perceptual effects of known, holistic configurations in contrast to constituent details (Pomeranz, Sager, & Stoeber, 1977; Poljac, de-Wit, & Wagemans, 2012).

Given the support for a model of memory traces of an extensive repertory of FEs, the findings in this study have implications for other areas of language research. They underscore an important role of episodic memory in speech processing, supporting the proposal that numerous short and long utterances are stored in memory with their lexical and prosodic characteristics. These findings also provide an impetus for better understanding of language development in children, yielding the notion that FEs are acquired differently from grammatical sentences, likely following disparate brain maturational schedules. Further, treatment in acquired language disorders will benefit from the information that FEs are stored in unitary form in the native language speaker; these utterances can be exploited and utilized in the process of restoring language use following stroke and other neurological impairment. Studies of FE processing in persons with cochlear implants are already planned.

The perceptual data reported in this paper can serve to document a time course of acquisition in second language learners. There is an extensive literature on the special role of formulaic expressions in second language learning (e.g., Cowie, 1992; Ellis, 1996, 2012; Conklin & Schmitt, 2008; Groom, 2009; Jiang & Nekrasova, 2007; Meunier, 2012). Later acquisition of formulaic language is known to be a challenge. We have begun collecting data on L2 speakers of English using operationally defined classifications of language competence, based on experience and exposure to L1 and L2 (Rammell, Van Lancker Sidtis, & Pisoni, 2016). While early results reveal that sentence transcription scores are very low under acoustic degradation conditions, some L2 listeners were able to identify individual words but not whole phrases. We expect to find large differential effects of linguistic experience in this population. Knowledge of the forms and subtle meanings of formulaic expressions may not develop fully even after many years of exposure and immersion in the L2 environment (e.g., Groom, 2009), possibly due to brain maturational factors. Better understanding of the time course of these developmental trajectories will enhance our understanding of language development and the effects of early linguistic experience on speech perception and production, especially under degraded listening conditions.

In summary, subjects correctly identified formulaic expressions more often than novel expressions under spectrally-degraded four-channel acoustic simulation. However, under one-channel simulation that preserved only gross durations and amplitude changes in the prosodic material, listeners were unable to identify any of the expressions correctly. It is possible that the presence of more detailed prosodic information, including intonation contour, sentence accent, and lexical stress patterns would yield higher recognition scores for formulaic expressions, which typically have stereotyped prosodic cues (Van Lancker et al., 1981; Ashby, 2006; Lin, 2010; Hallin & Van Lancker Sidtis, 2015; Yang et al., 2015; Yang & Van Lancker Sidtis, 2016). Further studies using acoustically altered formulaic and novel utterances may help to elucidate the prosodic cues that are utilized under these challenging listening conditions.

Under degraded four-channel acoustic simulation, subjects also correctly identified expressions to which they gave higher familiarity scores more often than those which they gave lower usage and familiarity scores. Subjects rated formulaic expressions as more familiar than novel expressions. Corpus analyses of the test sentences also revealed that formulaic expressions occur more frequently in spoken language corpora than novel expressions (by definition of “novel”), providing converging support for the patterns observed in the familiarity rating data obtained in the three behavioral experiments. However, frequency data alone do not suffice to account for the high recognition scores for the formulaic expressions, many of which were recognized despite low usage, exposure, or frequency of occurrence scores. It is not known, and possibly not knowable, even given the belief that frequency plays a major role, how frequently an FE must be heard in order to be acquired. Further, frequency counts for FEs are especially problematic, because FEs to a large extent occur as vehicles of spoken language; therefore many corpora designed for frequency analyses will not be appropriate. We submit that in addition to the effects of frequency of exposure, formulaic expressions may be acquired very rapidly in language development (Reuterskiöld & Van Lancker Sidtis, 2012), analogously to imprinting (Rauschecker & Marler, 1987), “flashbulb” memorial processes (Christianson, 1992), or fast mapping (Carey, 1978). Taken together, the perceptual and computational results reported in this paper support a dual-process model of language processing, in which formulaic and novel expressions are differentially acquired, stored, and processed.

Acknowledgements

This work was supported by grants from the National Institutes of Health to Indiana University (NIH-NIDCD: T32 Training Grant DC00012 and Research Grant 1R01 DC000064); and Research Grant R01 DC007658 to the Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY. We thank Luis Hernandez for assistance with programming, D. Kempler for contribution to stimulus development, M. Burgevin for editorial assistance, and John J. Sidtis for comments on a previous draft. Special thanks are extended to Michael N. Jones for his help in computational analyses of corpus data.

Appendix

Appendix A.: List of stimuli by type of expression, novel or formulaic: Matched pairs

Novel expression	Formulaic expression
1. A pitcher of this makes a fine drink.	1. A little of that goes a long way.
2. My van's the place.	2. The sky's the limit.
3. I always feed them out back.	3. You never had it so good.
4. Someone yelled, no one came.	4. Nothing ventured, nothing gained.
5. Little to feel good over	5. Nothing to write home about.
6. Twice in the morning	6. None of your business
7. For my little boy	7. On the other hand
8. Your uncle works with no one.	8. My record speaks for itself.
9. He might leave here quickly.	9. I should be so lucky
10. Out of the mountain's view	10. Straight from the horse's mouth
11. Like stealing his car out of a lot	11. Like beating your head against a wall
12. That's the trip my cousin offers.	12. That's the way the cookie crumbles.
13. From school or church	13. Through thick or thin
14. He made a certain move toward his left.	14. She took a sudden turn for the worse.
15. Those girls can't be in school.	15. Two wrongs don't make a right.
16. Wash her dresses in the stream	16. Wear my fingers to the bone.
17. My home and my people	17. Your money or your life
18. The smaller of my problems	18. The lesser of two evils
19. When they can't buy them, they can make them.	19. If you don't like it, you can lump it.
20. It ought to be nice at work.	20. There's going to be hell to pay.
21. Everyone is going out happy.	21. Everything is coming up roses.
22. It's like her to try and him to give up.	22. That's for me to know and you to find out.
23. He shouldn't have gambled with a clever girl.	23. It couldn't have happened to a nicer guy.
24. While the dog is around, the boy must leave.	24. When the cat is away, the mice will play.
25. Don't leave those cookies around him.	25. Don't burn your bridges behind you.
26. It's like riding my car over a rough street.	26. It's like beating your head against a brick wall.
27. But this they can all certainly understand.	27. And now for something completely different.
28. He won't hear the music of the bands.	28. I can't see the forest for the trees.
29. Drop those on the ground and break them.	29. Put that in your pipe and smoke it.
30. Lots cheaper old and used.	30. No sooner said than done.
31. It's awfully nice when they give up.	31. There's plenty more where that came from.
32. My son went to a better school.	32. The shoe is on the other foot.
33. As I drive, I sing.	33. When it rains, it pours.
34. This stops the leaks.	34. That takes the cake.
35. Stop your stupid dogs	35. Thank my lucky stars
36. With that I am convinced	36. To whom it may concern
37. She's running around in her yard.	37. He's turning over in his grave.

Novel expression	Formulaic expression
38. It's cloudy over our house.	38. That's water under the bridge.
39. Such a big car	39. What a small world.
40. What has he done most of the day?	40. Where have you been all of my life?
41. Give her a present she won't return.	41. Make him an offer he can't refuse.
42. We wrote brilliantly even before.	42. They lived happily ever after.
43. Grab a big dinner plate.	43. Keep a stiff upper lip.
44. Was that a letter from his girl?	44. Is there a doctor in the house?
45. He thinks he's still representing somebody.	45. I hope I'm not interrupting anything.
46. Couldn't she give us something better?	46. Haven't I met you someplace before?
47. The new facts were kept from him.	47. A good time was had by all.
48. He may take her in but he won't keep her.	48. You can dish it out but you can't take it.
49. He takes his pets in the car.	49. He's got his head in the clouds.
50. The cook is angry.	50. The coast is clear.
51. The dog's trying to give her a ride on the wagon.	51. I'd like to give you a piece of my mind.
52. The nails are under the square and the hammer is in the circle.	52. Sticks and stones will break my bones but words will never hurt me.
53. He's racing a truck against a horse.	53. She took a sudden turn for the worse.
54. Almost to the bottom of the mountain.	54. Just in the nick of time
55. She tried jumping over the striped cat.	55. It's like talking to a brick wall.
56. He's got a picture of show of her.	56. I've got a bone to pick with you.
57. He jumped up to her suddenly.	57. I'll get back to you later.
58. The clown, the small clown, and not the one near a girl.	58. The truth, the whole truth, and nothing but the truth.
59. She's looking down at her black cat.	59. He's saving up for a rainy day.
60. Then her dog walks in.	60. When our ship comes in
61. He sees her drinking from a bowl.	61. She's got him eating out of her hand.
62. It's easy to teach a dog to swim.	62. That's enough to drive a man to drink
63. Whenever the sun sets, the dog barks.	63. While the cat's away, the mice will play.
64. Follow your sister to the dinner table.	64. Keep your nose to the grindstone.
65. He kisses the thin lady.	65. It seems like just yesterday.
66. Where are they not showing each other their own hats?	66. Why don't you pick on somebody your own size?
67. He's sitting deep in the bubbles.	67. He's living high on the hog.

Appendix B.: Frequency counts for expressions and words

Pair	Sentences	Formulaic expressions frequency				Novel expressions frequency				
		Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3
1	He's got his head in the clouds He takes his pets in the car	5	19453	732		0	726	19739		

Pair	Sentences	Formulaic expressions frequency					Novel expressions frequency				
		Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
2	The coast is clear The cook is angry	5	6363	20197			0	2748	6633		
3	I'd like to give you a piece of my mind The dog's trying to give her a ride on the wagon	5	11047	19223			0	3526	365		
4	Just in the nick of time Almost to the bottom of the mountain	33	1954	175915			0	6821	3013		
5	It's like talking to a brick wall She tried jumping over the striped cat	0	52921	512	10785		0	1220	78	2012	
6	I've got a bone to pick with you He's got a picture of show of her	15	1692	9833			0	12693	59623		
7	I'll get back to you later He jumped up to her suddenly	7	28340				0	5556			
8	He's saving up for a rainy day She's looking down at her black cat	0	2008	326	83120		0	35937	22506	2012	
9	She's got him eating out of her hand He sees her	0	3716	17476			0	3265	1973		

Pair Sentences	Formulaic expressions frequency					Novel expressions frequency				
	Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
drinking from a bowl										
10 That's enough to drive a man to drink	0	6873	76912	3623		0	3905	6838	777	
It's easy to teach a dog to swim										
11 It seems like just yesterday	3	16317				0	7111			
He kisses the thin lady										
12 Cat got your tongue?	1	2012	623			0	6993	6838		
Who's following the dog?										
13 None of your business	104	7050	30834			2	4463	56564		
Twice in the morning										
14 A little of that goes a long way	0	91673	51013	125584		0	421	10739	3623	
A pitcher of this makes a fine drink										
15 The sky's the limit	31	2120	2760			0	5429	38553		
My van's the place										
16 You never had it so good	4	66606	139583			0	46812	2141		
I always feed them out back										
17 On the other hand	3832	132285	17476			1	91673	13139		
For my little boy										
18 My record speaks for itself	2	15279	2298			0	1697	10811		
Your uncle works										

Pair Sentences	Formulaic expressions frequency					Novel expressions frequency				
	Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
with no one										
19 I should be so lucky	3	3819				0	18948			
He might leave here quickly										
20 Like beating your head against a wall	0	2480	19453	10785		0	1075	19739	119703	
Like stealing his car out of a lot										
21 Through thick or thin	3	1074	1786			0	36549	11040		
From school or church										
22 She took a sudden turn for the worse	1	4264	19980	8051		0	16187	21395	31285	
He made a certain move toward his left										
23 Two wrongs don't make a right	18	248300				0	8881	36549		
Those girls can't be in school										
24 The lesser of two evils	21	560	125			0	3385	23353		
The smaller of my problems										
25 Don't burn your bridges behind you	0	1801	1043			0	18949	707		
Don't leave those cookies around him										

Pair	Sentences	Formulaic expressions frequency					Novel expressions frequency				
		Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
26	No sooner said than done Lots cheaper old and used	0	1668	170304	50112	0	1061	33382	34304		
27	There's plenty more where that came from It's awfully nice when they give up	0	3577	205073	46750	0	876	18839	233529		
28	When it rains it pours As I drive, I sing	2	495	75		0	6873	3946			
29	That takes the cake This stops the leaks	28	13475	1368		0	1685	829			
30	Thank my lucky stars Stop your stupid dogs	3	75516	3819	4802	0	22049	3380			
31	That's water under the bridge It's cloudy over our house	15	17760	2686		0	164	71618			
32	What a small world Such a big car	1	19437	68832		0	57877	19739			
33	Keep a stiff upper lip Grab a big dinner plate	7	520	552		0	4346	1505			
34	Is there a doctor in the house? Was that a letter	7	11584	71618		0	7470	13404			

Pair Sentences	Formulaic expressions frequency					Novel expressions frequency				
	Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
from his girl?										
35 Haven't I met you someplace before?	0	912	74879			0	95075	39479		
Couldn't she give us something better?										
36 Nothing ventured, nothing gained.	2	29632	101	1859		0	419	46737		
Someone yelled, no one came.										
37 It couldn't have happened to a nicer guy.	0	39094	282	29620		0	739	13397		
He shouldn't have gambled with a clever girl.										
38 It's like beating your head against a brick wall.	0	2480	19447	512	10783	0	1543	19721	2091	18458
It's like riding my car over a rough street										
39 Put that in your pipe and smoke it.	4	65620	764	3635		0	4312	14130	37002	
Drop those on the ground and break them.										
40 To whom it may concern	21	5036	8471			0	546531	4244		
With that I am convinced										
41 You can dish it out but you can't take it	9	832	100571			0	100571	35560		

Pair Sentences	Formulaic expressions frequency				Novel expressions frequency					
	Whole Phrase	CW1	CW2	CW3	CW4	Whole Phrase	CW1	CW2	CW3	CW4
He may take her in but he won't keep her										

Note: The first sentence of each pair is a formulaic expression, the second a novel expression. Content words are boldfaced in the sentence. Content word frequencies are displayed in the same order that they appear in the sentence. CW = Content Word.

References

- Ashby M (2006). Prosody and idioms in English. *Journal of Pragmatics*, 38, 1580–1597. doi: 10.1016/j.pragma.2005.03.018
- Baayen RH, & Hendrix P (2011, 1). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning In *Empirically examining parsimony and redundancy in usage-based models*, LSA workshop.
- Baayen RH, Milin P, Durdevic DF, Hendrix P, & Marelli M (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, 118(3), 438–481. doi: 10.1037/a0023851 [PubMed: 21744979]
- Bannard C, Baldwin T, & Lascarides A (2003). A statistical approach to the semantics of verb-particles. *Proceedings of the ACL-Workshop on Multiword Expressions: Analysis, Acquisition, and Treatment*, 65–72.
- Bell N (2012). Formulaic language, creativity, and language play in a second language. *Annual Review of Applied Linguistics*, 32, 189–205. doi: 10.1017/S0267190512000013
- Biber D (2009). A corpus-driven approach to formulaic language. *International Journal of Corpus Linguistics*, 14, 275–311. doi: 10.1075/ijcl.14.3.08bib
- Bolinger D (1976). Meaning and memory. *Forum Linguisticum*, 1, 1–14.
- Bolinger D (1977). Idioms have relations. *Forum Linguisticum*, 2, 157–169.
- Boersma P, & Weenink D (2007). Praat: Doing phonetics by computer (V. 4.6. 34) (Computer program). Retrieved October 19 2007.
- Bousfield WA (1953). The occurrence of clustering in recall of randomly arranged associates. *Journal of General Psychology*, 49, 229–240. doi: 10.1080/00221309.1953.9710088
- Caillies S, & Butcher K (2007). Processing of idiomatic expressions: Evidence from a new hybrid view. *Metaphor and Symbol*, 22(1), 79–108. doi: 10.1080/10926480709336754
- Cacciari C, & Tabossi P (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683. doi: 10.1016/0749-596X(88)90014-9
- Carey S (1978). Less may never mean more In Campbell R & Smith P (Eds.), *Recent advances in the psychology of language* (pp. 109–132). New York: Plenum Press.
- Christianson S-Å. (1992). Do flashbulb memories differ from other types of emotional memories? In Winograd E & Neisser U (Eds.), *Affect and accuracy in recall: Studies of “flashbulb memories”* (pp. 191–211). Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511664069.010
- Clark HH (1970). Word associations and linguistic theory In Lyons J (Ed.), *New horizons in linguistics* (pp. 271–286). Penguin Books: Baltimore.
- Conklin K, & Schmitt N (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72–89. doi: 10.1093/applin/amm022
- Conrad S, & Biber D (2004). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographica*, 20, 56–71.

- Cowie AP (1992). Multiword lexical units and communicative language teaching In Arnaud P & Bejoint H (Eds.), *Vocabulary and applied linguistics* (pp. 1–12). London: Macmillan. doi: 10.1007/978-1-349-12396-4_1
- Cutting JC, & Bock K (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory-Cognition*, 25(1), 57–71. doi: 10.3758/BF03197285 [PubMed: 9046870]
- Davies M (2008). The corpus of contemporary American English: 450 million words, 1990–present. Available online at <http://corpus.byu.edu/coca/>.
- Dorman MF, & Loizou PC (1997). Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants. *American Journal of Otolaryngology*, 18, S13–S114.
- Dorman MF, & Loizou PC (1998). Identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels. *Ear Hear*, 19, 162–166. doi: 10.1097/00003446-199804000-00008 [PubMed: 9562538]
- Ellis NC (1996). Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition*, 18, 91–126. doi: 10.1017/S0272263100014698
- Ellis NC (2012). Formulaic language and second language acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics*, 32 (1), 17–44. doi: 10.1017/S0267190512000025
- Erman B (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25–53. doi: 10.1075/ijcl.12.1.04erm
- Erman B, & Warren B (2000). The idiom principle and the open choice principle. *Text – International Journal for the Study of Discourse*, 20(1), 29–62.
- Fillmore C (1979). On fluency In Fillmore CJ, Kempler D, & S-Y Wang W (Eds.), *Individual differences in language ability and language behavior* (pp. 85–102). London: Academic Press.
- Foster P (2001). 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 (Eds.), *Researching pedagogic tasks: second language learning, teaching, and testing* (pp. 75–93). Harlow, UK: Longman.
- Gibbs R (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8, 149–156. doi: 10.3758/BF03213418 [PubMed: 7382816]
- Gleason J, Berko, & Weintraub S (1976). The acquisition of routines in child language. *Language in Society*, 5, 129–136. doi: 10.1017/S0047404500006977
- Gleason J Berko. (1980). The acquisition of social speech and politeness formulae In Giles H, Robinson WP, & Smith PM (Eds.), *Language: social psychological perspectives* (pp. 21–27). Oxford and New York: Pergamon Press.
- Greenwood DD (1990). A cochlear frequency-position function for several species – 29 years later. *The Journal of the Acoustical Society of America*, 87, 2592–2605. doi: 10.1121/1.399052 [PubMed: 2373794]
- Groom N (2009). Effects of second language immersion on second language collocational development In Barfield A & Gyllstad H (Eds.), *Researching collocations in another language* (pp. 21–33). Basingstoke, UK: Palgrave MacMillan.
- Hallin A, & Van Lancker Sidtis D (2015). A closer look at formulaic language: Prosodic patterns in Swedish proverbs. *Applied Linguistics*. In press.
- Hill J (2001). Revising priorities: From grammatical failure to collocational success In Lewis M (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–69). Hove, UK: Language Teaching Publications.
- Horowitz LM, & Manelis L (1973). Recognition and cued recall of idioms and phrases. *Journal of Experimental Psychology*, 100, 291–296. doi: 10.1037/h0035468
- Jackendoff R (1995). The boundaries of the lexicon In Martin Everaert Erik-Jan, van der Linden Andre Schenk, & Rob Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp.133–166). Hillsdale NJ: Lawrence Erlbaum Associates.
- James W (1895). The knowing of things together. *The Psychological Review*, II(2), 105–124. doi: 10.1037/h0073221

- Jarema G, Busson C, Nikolova R, Tsapkin K, & Libben G (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, 68 (1–2), 362–369. doi: 10.1006/brln.1999.2088 [PubMed: 10433782]
- Jiang N, & Nekrasova TM (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91(3), 433–445. doi: 10.1111/j.1540-4781.2007.00589.x
- Kecskes I (2003). *Situation-bound utterances in L1 and L2*. Berlin: Mouton de Gruyter. doi: 10.1515/9783110894035
- Kecskes I (2007). Formulaic language in English Lingua Franca In Kecskes I & Horn LR (Eds.), *Explorations in pragmatics*. Berlin: Walter de Gruyter.
- Kempler D, & Van Lancker D (1996). *The Formulaic and Novel Language Comprehension Test (FANL-C)*. Copyright. (For complete test materials, see http://blog.emerson.edu/daniel_kempler/fanlc.html).
- Kempler D, Van Lancker D, Marchman V, & Bates E (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology*, 15, 327–349. doi: 10.1080/87565649909540753
- Kuiper K (2006). Knowledge of language and phrasal vocabulary acquisition. *Behavioral and Brain Sciences*, 29, 291–92. doi: 10.1017/S0140525X06359067
- Kuiper K (2007). Cathy Wilcox meets the phrasal lexicon: Creative deformation of phrasal lexical items for humorous effect In Munat J (Ed.), *Lexical creativity, texts and context* (pp. 93–112). Amsterdam: John Benjamins. doi: 10.1075/sfsl.58.14kui
- Kuiper K (2009). *Formulaic genres*. UK: Palgrave Macmillan. doi: 10.1057/9780230241657
- Kuiper K, Van Egmond M, Kempen G, & Sprenger S (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2(3), 313–357. doi: 10.1075/ml.2.3.03kui
- Libben MR, & Titone D (2008). The multidetermined nature of idiom processing. *Memory and Cognition*, 36, 1103–1121. doi: 10.3758/MC.36.6.1103 [PubMed: 18927029]
- Lieberman P (1963). Some effects of semantic and grammatic context on the production and perception of speech. *Language and Speech*, 6, 172–187.
- Lin PMS (1999). Automatic identification of noncompositional phrases. *Proceedings of the 37th Annual Meeting of the ACL* (pp. 317–324). College Park: USA.
- Lin PMS, & Adolphs S (2009). Sound evidence: Phraseological units in spoken corpora In Barfield A and Gyllstad H (Eds.), *Collocating in another language: multiple interpretations* (pp. 34–48). Basingstoke, England: Palgrave Macmillan.
- Lin PMS (2010). The phonology of formulaic sequences: A review In Wood D (Ed.), *Perspectives on formulaic language: acquisition and communication* (pp. 174–193). London, UK: Continuum.
- Locke JL (1993). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.
- Lounsbury FG (1963). Linguistics and psychology In Koch S (Ed.), *Psychology: a study of a science* (pp. 552–582). NY: McGraw-Hill, Inc.
- Meunier F (2012). Formulaic language and language teaching. *Annual Review of Applied Linguistics*, 32(1), 111–129. doi: 10.1017/S0267190512000128
- Millar N (2011). The processing of malformed formulaic language. *Applied Linguistics*, 32, 129–148. doi: 10.1093/applin/amq035
- Miller GA (1956). The magic number seven plus or minus two: Some limits to our perspective for processing information. *Psychological Review*, 63, 81–87. doi: 10.1037/h0043158 [PubMed: 13310704]
- Moon RE (1998a). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford, UK: Clarendon Press.
- Moon RE (1998b). Frequencies and forms of phrasal lexemes in English In Cowie AP (Ed.), *Phraseology* (pp. 79–100). Oxford: Clarendon Press.
- Nilsson M, Soli SD, & Sullivan JA (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, 95, 1085–1099. doi: 10.1121/1.408469 [PubMed: 8132902]
- Nippold MA (1998). *Later language development: The school-age and adolescent years*. 2nd edition, Austin, TX.

- Nooteboom S (2011). Self-monitoring for speech errors in novel phrases and phrasal lexical items. *Yearbook of Phraseology*, 1, 1–16.
- Nunberg G, Sag IA, & Wasow T (1994). Idioms. *Language*, 70 (3), 491–538. doi: 10.1353/lan.1994.0007
- Osgood CE, & Housain R (1974). Salience of the word as a unit in the perception of language. *Perception and Psychophysics*, 15, 168–192. doi: 10.3758/BF03205845
- Paquot M, & Granger S (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130–149. doi: 10.1017/S0267190512000098
- Pawley A, & Syder FH (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency In Richards JC & Schmidt RW (Eds.), *Language and communication* (pp. 191–225). London: Longman.
- Perkins MR (1999). Productivity and formulaicity in language development In Garman M, Letts C, Richards B, Schelletter C, & Edwards S (Eds.), *Issues in normal and disordered child language: from phonology to narrative*. Special Issue of *The New Bulmershe Papers* (pp. 51–67). Reading: University of Reading.
- Peters A (1983). *The units of language*. Cambridge: Cambridge University Press.
- Pickens JD, & Pollio HR (1979). Patterns of figurative language in adult speakers. *Psychological Research*, 40, 299–313. doi: 10.1007/BF00309157
- Pisoni DB (1996). Word identification in noise. *Language and Cognitive Processes*, 11, 681–687. doi: 10.1080/016909696387097 [PubMed: 21687807]
- Poljac E, de-Wit L, & Wagemans J (2012). Perceptual wholes can reduce the conscious accessibility of their parts. *Cognition*, 123, 308–312. doi: 10.1016/j.cognition.2012.01.001 [PubMed: 22306190]
- Pomeranz JR, Sager LC, & Stoever RJ (1977). Perception of wholes and their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, 3(3), 422–435. [PubMed: 886276]
- Rammell CS, Van Lancker Sidtis D, & Pisoni D (2016). Perception of formulaic and novel expressions under acoustic degradation by native, non-native, and heritage speakers. Paper presented at the Biennial High Desert Linguistics Society Conference Albuquerque, New Mexico, November 12–14.
- Rauschecker JP, & Marler P, Eds. (1987). *Imprinting and cortical plasticity*. New York: John Wiley & Sons.
- Reuterskiöld C, & Van Lancker Sidtis D (2012). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy*, 29(2), 216–228.
- Schone P, & Jurafsky D (2001). Is knowledge-free induction of multiword unit dictionary headwords a solved problem? *Proceedings of the 6th Conference on Empirical Methods in Natural Language Processing* (pp. 100–108).
- Shannon RV, Zeng FG, Kamath V, Wygonski J, & Ekelid M (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234), 303–304. doi: 10.1126/science.270.5234.303 [PubMed: 7569981]
- Shannon R, Fu QJ, & Galvin JJ III. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica*, 124 (Supplement 552), 50–54. doi: 10.1080/03655230410017562
- Simon HA, Zhang W, Zang W, & Peng R (1989). STM capacity for Chinese words and idioms with visual and auditory presentations In *Models of thought, II*. (pp. 68–75). Yale University Press: New Haven and London.
- Simon HA (1974). How big is a chunk? *Science*, 183, 482–488. doi: 10.1126/science.183.4124.482 [PubMed: 17773029]
- Sinclair JM (1987). Collocation: A progress report In Steele R & Threadgold T (Eds.), *Language topics: Essays in honor of Michael Halliday, II* (pp. 319–331). Amsterdam: John Benjamins.
- Sivanova-Chanturia A, Conklin K, & Schmitt N (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251–272. doi: 10.1177/0267658310382068
- Sprenger SA, Levelt WJM, & Kempen G (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–184. doi: 10.1016/j.jml.2005.11.001

- Stahl B, & Van Lancker Sidtis D (2015). Tapping into neural resources of communication: Formulaic language in aphasia therapy. *Frontiers in Psychology*, 6, Article 1526. doi: 10.3389/fpsyg.2015.01526
- Swinney D, & Cutler A (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18, 523–534. doi: 10.1016/S0022-5371(79)90284-6
- Tabossi P, Fanari R, & Wolf K (2009). Why are idioms recognized fast? *Memory & Cognition*, 37(4), 529–540. doi: 10.3758/MC.37.4.529 [PubMed: 19460959]
- Tannen D, & Öztekin FC (1981). Health to our mouths In Coulmas F (Ed.), *Conversational routine* (pp. 516–534). The Hague: Mouton.
- Tiger CIS (2012). AngeliSIM [Software]. Available from <http://angelsim.emilyfufoundation.org/>.
- Titone DA, & Connine CM (1994). Comprehension of idiomatic expressions: effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 1126–1138.
- Titone DA, & Connine CM (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, 31, 1655–1674. doi: 10.1016/S0378-2166(99)00008-9
- Tremblay A, Derwing B, Libben G, & Westbury C (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61 (2), 569–613. doi: 10.1111/j.1467-9922.2010.00622.x
- Underwood G, Schmitt N, & Galpin A (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences In Schmitt N (Ed.), *Formulaic sequences* (pp. 153–172). Amsterdam: John Benjamins. doi: 10.1075/llt.9.09und
- Van Lancker Sidtis D (2012a). Two-track mind: Formulaic and novel language support a dual-process model In Faust M (Ed.), *The handbook of the neuropsychology of language* (pp. 342–367). Boston: Blackwell Publishing. doi: 10.1002/9781118432501.ch17
- Van Lancker Sidtis D (2012b). Formulaic language and language disorders. *The Annual Review of Applied Linguistics*, 32, 62–80. doi: 10.1017/S0267190512000104
- Van Lancker Sidtis D (2014). Formulaic language in an emergentist framework In MacWhinney M and O’Grady W (Eds.), *Handbook of language emergence* (578–599). Hoboken: Wiley-Blackwell.
- Van Lancker Sidtis D (2003). Auditory recognition of idioms by first and second speakers of English: It takes one to know one. *Applied Psycholinguistics*, 24, 45–57.
- Van Lancker D (1975). Heterogeneity in Language and Speech: Neurolinguistic Studies Working Papers in Phonetics 29, UCLA, Los Angeles, CA Available on line at: <http://escholarship.org/uc/item/8zw4z7ch>
- Van Lancker D, & Rallon G (2004). Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language and Communication*, 24, 207–240. doi: 10.1016/j.langcom.2004.02.003
- Van Lancker Sidtis D, Cameron K, Bridges K, & Sidtis JJ (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand*, 2, 39–48. PMID: . doi: 10.1016/j.amper.2015.02.001 [PubMed: 26392923]
- Van Lancker D, Canter GJ, & Terbeek D (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24, 330–335. doi: 10.1044/jshr.2403.330 [PubMed: 7300273]
- Wray A (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511519772
- Wray A, & Perkins MR (2004). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1, 1–28. doi: 10.1016/S0271-5309(99)00015-4
- Wulff S (2008). *Rethinking idiomaticity: a usage-based approach*. New York: Continuum.
- Yang S-Y, Ahn J-S, & Van Lancker Sidtis D (2015). Listening and acoustic studies of idiomatic- literal contrastive sentences in Korean Speech, Language, and Hearing. In press. doi: 10.1179/2050572814Y.0000000061
- Yang S-Y, & Van Lancker Sidtis D (2016). Production of Korean idiomatic utterances following left- and right-hemisphere damage: Acoustic studies. *Journal of Speech, Language, and Hearing Research*, 59(2), 267–280. PMID: . [PubMed: 26556625]

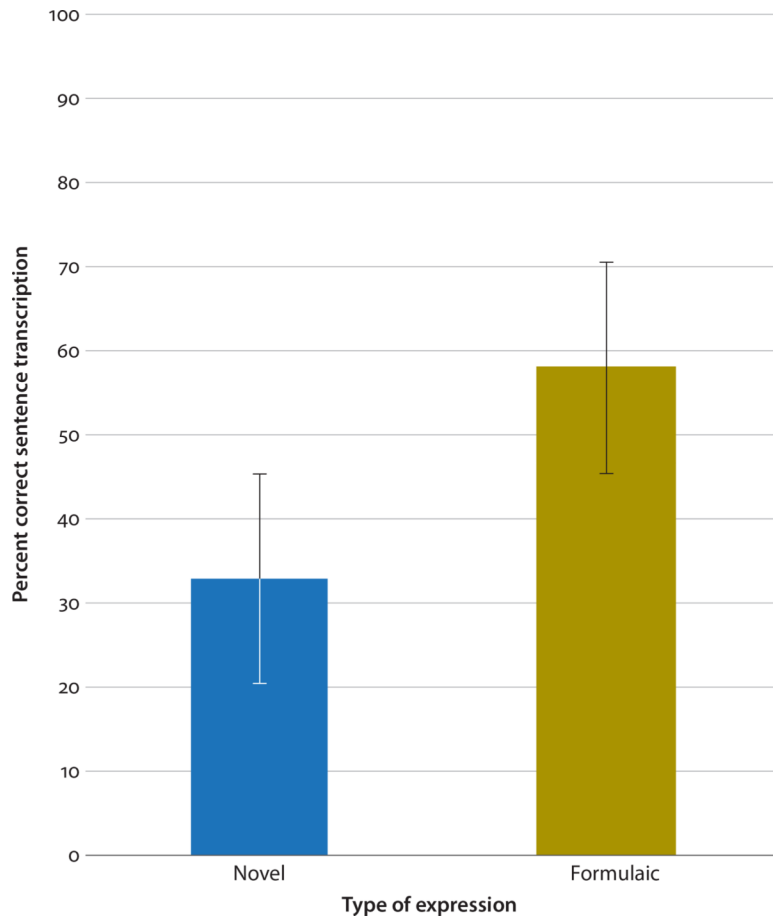


Figure 1. Percent correct sentence transcription by type of expression. Novel expressions are shown on the left, formulaic expressions on the right. Error bars represent standard error of the mean

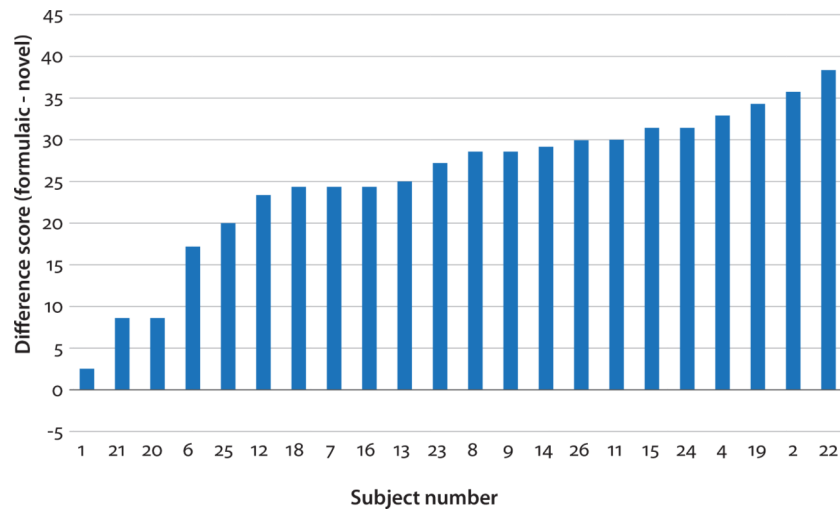


Figure 2. Difference scores between formulaic and novel expression accuracy by participant. Scores are listed in ascending order. X-axis numbers represent subject numbers

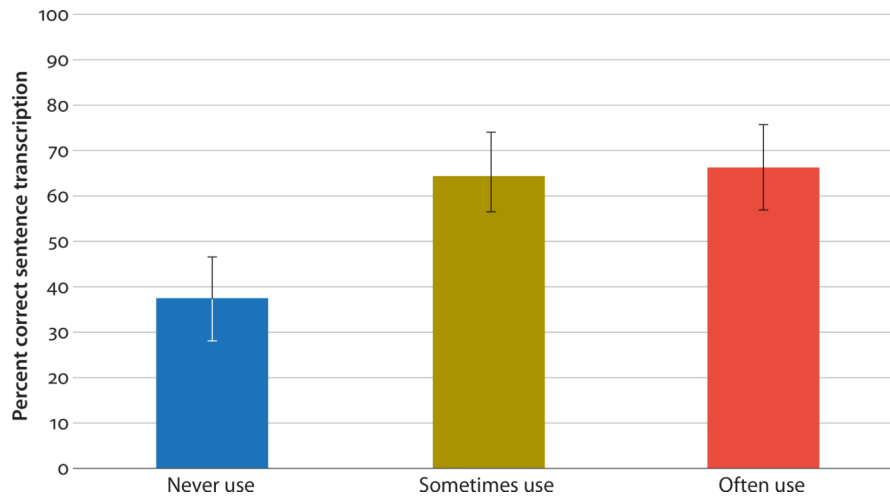


Figure 3. Percent correct transcription by familiarity rating for both expression types

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

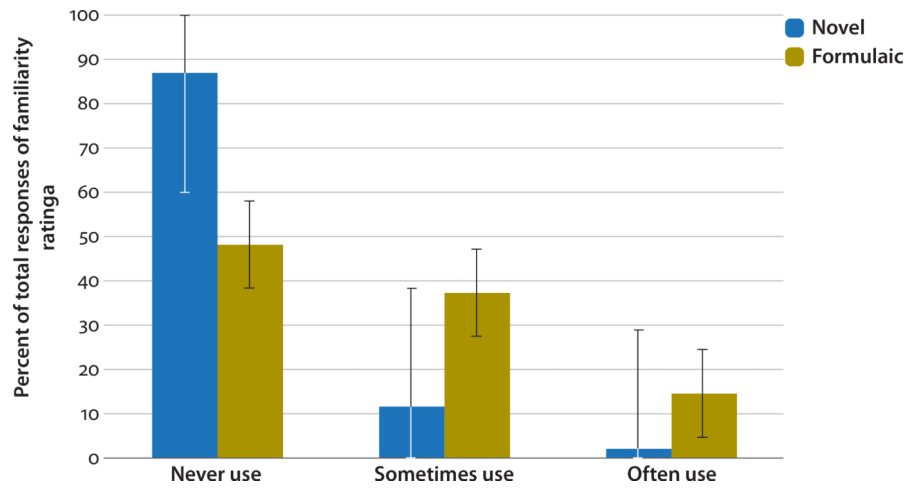


Figure 4.
Familiarity rating by type of expression

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

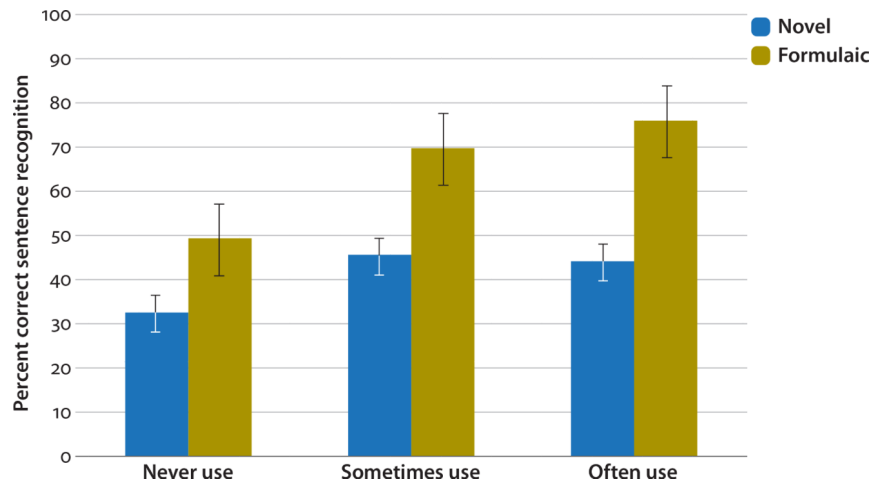


Figure 5. Accuracy as a function of usage familiarity rating and type of expression. Error bars represent standard error of the mean

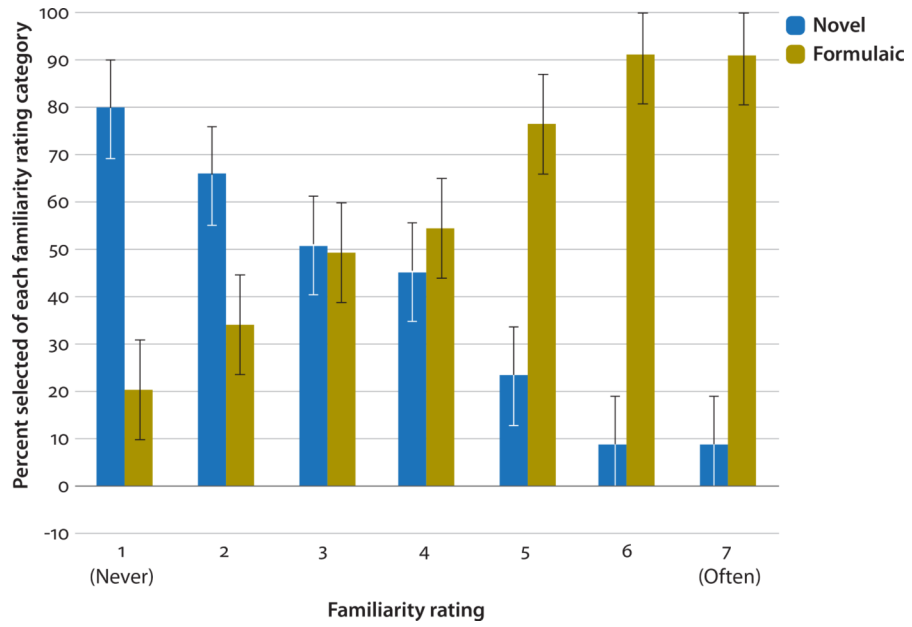


Figure 6. Overall familiarity rating by type of expression from Experiment 2. For each rating, the sum of expression type selected equals 100%. Error bars represent standard error of the mean

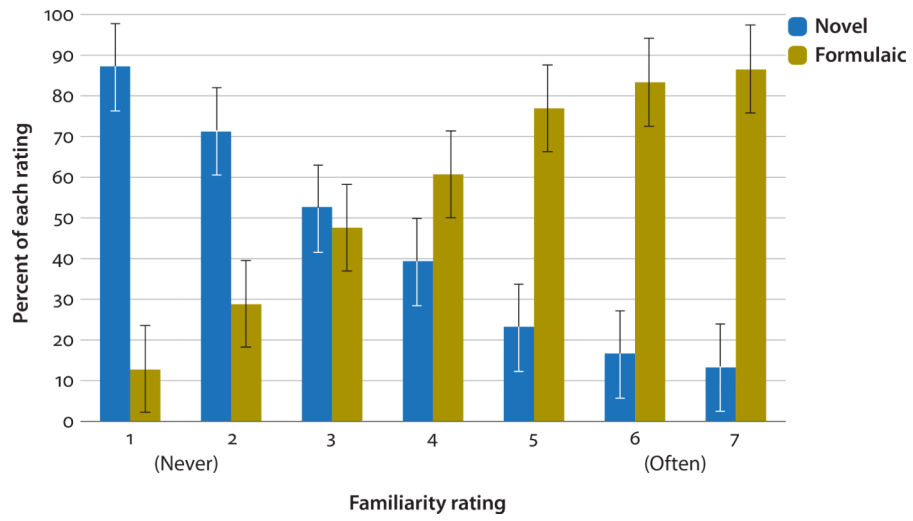


Figure 7. Overall familiarity ratings by type of expression from Experiment 3. For each rating, the sum of expression type selected equals 100%. Error bars represent the standard error of the mean