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

Diana Van Lancker Sidtis [a,b,*], Krista Cameron [b], Kelly Bridges [a,b], John J. Sidtis [b,c]

[a] Communicative Sciences and Disorders, New York University, 665 Broadway, New York, NY 10012, United States
[b] Brain and Behavior Laboratory, Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10961, United States
[c] Department of Psychiatry, New York University Langone School of Medicine, 550 First Avenue, New York, NY 10016, United States

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.

ARTICLE INFO

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.

1. Introduction

Formulaic language has relevance to many branches of linguistic study and interest arising from many disciplines is increasing rapidly (eg., Wulff, 2013). 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, 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 performed a judgment task more rapidly to the formulaic than the novel expression (Jiang and Nekrsova, 2007; Tabossi et al., 2009). Eye movement studies showed an advantage for formulaic expressions (Conklin and Schmitt, 2008; Underwood et al., 2004; Siyanova-Chanturia et al., 2011).

More recently, the specific effects of neurological disease on incidence of formulaic language in spontaneous speech (Cappelle et al., 2010; Dieguez and Bogousslavsky, 2007; Van Lancker Sidtis, 2004; Van Lancker Sidtis and Postman, 2006; Sidtis et al., 2009; Van Lancker Sidtis, 2012) point to differential cerebral systems underlying these two kinds of linguistic competence, suggesting that

---

* Correspondence to: Communicative Sciences and Disorders, New York University, 665 Broadway, Room 936, New York, NY 10012, United States. Tel.: +1 917 224 5483.

E-mail address: diana.sidtis@nyu.edu (D. Van Lancker Sidtis).

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, in press).

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 (Bardovi-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 expression and their relative incidence in texts (Altenberg, 1998; Sinclair, 1991; 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 et al., 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, in press), 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 et al., 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 et al., 2007) and Korean sentences (Yang and Van Lancker Sidtis, 2015), although different acoustic cues were found to form significant contrasts for ditropic utterances in these languages. Rammell et al. (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 et al., 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 Sidtis 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 et al., 2013; Wolf et al., 2014). These studies lead to a model of formulaic language as governed by a right hemisphere-subcortical system (See Van Lancker Sidtis, in press, for review).

The study reported here continues the pursuit of performance data from healthy language users on expressions focusing on the schema as 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 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