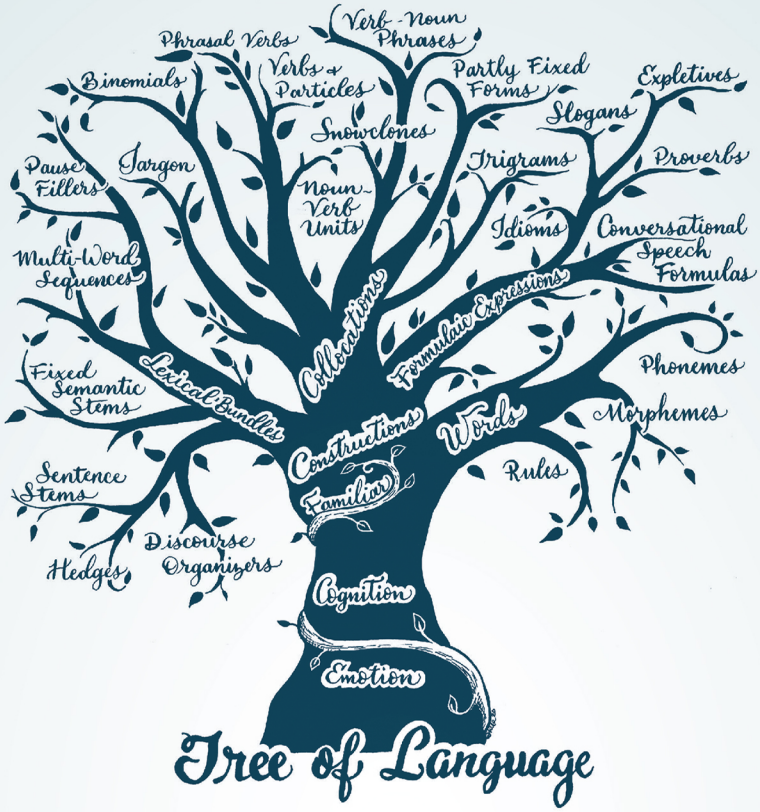

FOUNDATIONS OF FAMILIAR LANGUAGE

FORMULAIC EXPRESSIONS, LEXICAL BUNDLES, AND COLLOCATIONS AT WORK AND PLAY

DIANA SIDTIS

WILEY Blackwell

Foundations of Familiar Language

Foundations of Familiar Language

Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play

Diana Sidtis

WILEY Blackwell

This edition first published 2021

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Sidtis, Diana, author.
Title: Foundations of familiar language : formulaic expressions, lexical bundles, and collocations at work and play / Diana Sidtis.
Description: Hoboken : John Wiley & Sons Inc, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2021009912 (print) | LCCN 2021009913 (ebook) | ISBN 9781119163275 (hardback) | ISBN 9781119163329 (paperback) | ISBN 9781119163282 (adobe pdf) | ISBN 9781119163299 (epub) | ISBN 9781119163305 (ebook)
Subjects: LCSH: Idioms. | Phraseology.
Classification: LCC P326.5.P45 S54 2021 (print) | LCC P326.5.P45 (ebook) | DDC 418--dc23
LC record available at https://lccn.loc.gov/2021009912
LC ebook record available at https://lccn.loc.gov/2021009913

Cover image: © Illustration by Sheila Monahan
Cover design by Wiley

Set in 10/12.5pt GalliardStd by Integra Software Services, Pondicherry, India

To my late father, Thomas Roupas, who taught me to love learning, and to my husband, John Sidtis, who is showing me how to love. Both have slathered me with familiar expressions.

Contents

Acknowledgments

And now I squeeze 50 years of rich opportunities and fertile grounds into a few paragraphs. First, there was the UCLA Phonetics Laboratory, where the director, Peter Ladefoged, although strictly a British phonetician, tolerated all our nonstandard activities. He was often amused over what he referred to as my "dirty words" (dichotic) tape. He read several drafts of the 1975 dissertation, which aired earlier versions of some of the ideas in the present book. Similar forbearance, this time for my preoccupation with ditropic utterances, emerged during a postdoctoral fellowship with mentor Gerald J. Canter at Northwestern University. There followed a decade devoted to voice research, funded by NIH, until a clinical position at the VA Outpatient Clinic in Los Angeles allowed for time, space, study participants and student helpers to explore questions about familiar language. Protocols were developed, testing was done, and results were published. Timely and fabulous was the invitation from Professor Michael Flynn to teach four classes at Carleton College in Northfield, MN. For one of them, I set out to explore "The Neurolinguistics of Nonliteral Language." Student enthusiasm for the lore of familiar expressions, reassuringly always detectible, exploded during my 20 years at New York University, where undergraduate and graduate students mined a huge cache of insightful topics for papers, theses, and dissertations. Their outstanding published work is referenced throughout these pages.

My deepest gratitude is reserved for my husband, John, who is a true partner in everything I do and think and anything I think of doing. He is a brilliant spontaneous generator of fixed expressions. I write them down. We are endlessly entertained pursuing labyrinthine, energetic discussions about a single phrase. I was thrilled when he contributed his expertise in brain scanning to questions about brain function underlying the dual-process model. As the book progressed, he read and commented on all the pages, and he prepared about half of the figures. And alongside all this high-level intellectual support, he shopped and cooked and vacuumed, while I sat at the computer. These copious blessings are not deserved or earned, but there they are.

Foundations of Familiar Language: Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play, First Edition. Diana Sidtis.

Preface

From hazy beginnings, the study of familiar language has unfurled throughout the speech sciences, touching every one of its subdisciplines. The purpose of this book is to bring together as much understanding of the topic, gleaned from personal observation, home-grown research studies, and from the many published works, as can fit between two covers. The term *familiar language* is used in these pages as a superordinate term to refer to expressions that have two special attributes: (1) the peculiar status of being known, or recognized as familiar, along with their specialized meanings and use conditions, to speakers of a language; and (2) they are fixed, unitary, or cohesive in form. The term "fixed, familiar expressions," identifying the primary differences between these and novel, newly created expressions, best embraces these attributes.

Exemplars of familiar language differ essentially from grammatical, novel expressions in the specific manner mentioned above, and they also differ from each other. Key attributes inhering in greater or lesser degree are nonliteral meaning and nuance. With an eye to using these characteristics as guidelines, exemplars of familiar language fall into three classes. Formulaic expressions, made up of idioms, expletives, conversational speech formulas, and proverbs, trade in nonliteral meanings and strong nuances. The second class is lexical bundles, conveying little or no nonliteral meaning and strictly diminished in nuance. Collocations, a very large body of phrases, adhere to literal meaning processes and carry a cornucopia of nuances, while enjoying a special status as unitary entries in the mental lexicon. Other parameters that distinguish the three classes include the factors contributing most potently to their acquisition, differences in linguistic features, and aspects of psychological and cerebral processing.

Knowledge in the speech community of the status of familiar language is delightfully and obviously revealed in social media. One purpose of the introductory sections is to document and illustrate insight and understanding of the complex characteristics of formulaic expressions, lexical bundles and collocations, as referenced from public

Foundations of Familiar Language: Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play, First Edition. Diana Sidtis.

material. People know a great deal about these linguistic phenomena, so much so that humor, word play, subtle messages, manipulation of the reader, and just plain intelligence for familiar language, as well as fun with it, are widely practiced. It is astonishing as well as rewarding to perceive how comfortable – and creative – the general populace is with this mode of language competence.

The sections proceed with overviews of the current state of knowledge, touching on experimental results, proposals, and models. Much has been learned in the past few decades. It is exciting to see the potency and authority in many writers on the various subtopics of familiar language. Secondary aims in producing this book have been to offer resources for further thought and research and to provide glimpses of the historical record. In service of these goals, examples, lists, and the several appendices, showcasing research protocols and listings of fixed, familiar expressions, current and past, are included.

Exemplars of familiar language are emergent, in that they feature characteristics that particularize them beyond the sum of their constituent parts. Evidence shows that these characteristics are known to native speakers, and that the exemplars appear to be stored in an abstract format. As has been amply examined in research studies, the abstract format can be instantiated in many ways. These kinds of manipulations are reviewed and explored in these pages. To provide a theoretical base for the three classes, the abstract representations are designated as *formuleme* for formulaic expressions, t*emplate* for lexical bundles, and *construction* for collocations. It will also be seen that the parameters of cohesion and use of nonliteral meaning to describe and distinguish types of familiar expressions can occur in degrees, leading to inevitable flexibility in classification. It will also be proposed that frequency of occurrence is only one parameter of importance in acquiring fixed expressions. Others are affect, arousal and attentional mechanisms, and innate mental capacities for the familiarity sense and for chunking.

The role of memory and mental knowledge in acquisition and storage of fixed expressions fits well with the perspectives of the episodic theory of speech perception and with construction grammar.

An all-encompassing explanation for the distinctive presence of familiar and novel language in language use, acquisition, disturbance and loss lies in the dual process model, which describes holistic alongside sequential modes as essentially distinctive mental competences. Extensive evidence for the dual process model comes from linguistic analyses, psychological studies and conditions occurring in neurologic disorders. Evidence from these sources, as it pertains to a differentiation of the three classes and their subsets, is presented throughout this book. These are preliminary proposals about the status and functioning of familiar language in mind and brain. Understanding is in its proverbial infancy and I realize I know only a little. More will constantly be revealed. It may be that the sheer heterogeneity of these phenomena, when better understood, will lead – beyond the dual – to a multiplex concept of language. This vision would accord well with the iconic representation of human language in this book as a majestic tree with many branches.

Appreciation:

Sheila Monahan and Michele Burgevin gave brilliant service to the many aspects of manuscript development.

Introduction

Where foundational stones are put in place, comprising definitions, history, translation challenges, incidence, provenance, and staying power.

You cannot open a book without learning something.
Confucius

As old as Adam

It has been known for a long time that much of communication proceeds with routinized, prefabricated expressions. Robert Louis Stevenson (1882, p. 13) observed that *the business of life is not carried on by words, but in set phrases, each with a special and almost a slang signification*. This prescient comment refers to the highly specialized knowledge that speakers have, knowledge that includes a constellation of detail around every fixed, familiar phrase. John Ciardi made a similarly astute observation in the forward to his 1987 book (p. 1):

> Idiom [i.e., language] is a seemingly sequential illogic (psycho-logic?) to which native speakers of any particular language become conditioned. It is a language convention and encodement, and we become imprinted with it in something like the way a gosling is inner directed to follow the first creature it sees. The gosling asks no questions. It does what seems to be its nature. Like it, we follow our language lead even to the point of absurdity.

The germaneness of this remark – following language conventions to the point of absurdity – pertains to the nature of many fixed, familiar phrases: the meanings are often nonliteral and not predictable from the words themselves; grammatic structure is sometimes distorted; the pronunciation might be idiosyncratic, with specified melody, voice quality, and phonetics; nuances and connotations are strong; and, in

Foundations of Familiar Language: Formulaic Expressions, Lexical Bundles, and Collocations at Work and Play, First Edition. Diana Sidtis.

many cases, only certain social and linguistic contexts allow for appropriate use. These interesting notions are explored from many perspectives in this book.

It is generally accepted by lay persons and linguists alike that human language contains a large proportion of fully formed expressions, varied in shape and meaning, more or less cohesive, and that these expressions are recognized and known to native speakers (Mackin, 1978; MacKenzie & Kayman, 2018; Schmitt, 2004; Tabossi, Fanari, & Wolf, 2009). Casual observation and introspection, as well as numerous studies, bear witness to the fact that the average language user frequently produces and recognizes many expressions that are "known" or familiar[1] (Mitchell, 1971). These recurrent, cohesive expressions in everyday language use are somehow represented in long term memory, as implied by the fact that they are "recognized." The ubiquity of these familiar expressions in all of discourse is examined and celebrated in this book.[2] Their hefty and bracing presence throughout verbal and written communication will be amply covered and described. These have been variously called prefabricated, unitary, routinized, fixed, semi-fixed, frozen, cohesive, collocated, and pre-assembled (Wray, 2002). In these pages, the term *familiar language* will be used as the superordinate category that embraces a large group of language behaviors that, while diverse, share two important characteristics: cohesion and familiarity.

To spare the reader the tedium of lexical repetition as the discussions go forward, other terms for the phenomena are sometimes used interchangeably, and are intended all to designate familiar language: *prefabricated expressions*, *prefabs*, *fixed expressions, unitary utterances*, *prepatterned expressions*, or *known expressions.* These terms inclusively designate the very large domain of expressions that are known in the speech community with the property of cohesion; this is useful because studies have been performed on varied renditions, subsets, or groupings within this domain. In this current treatment, the popular cover terms, *formulaic language* and *formulaic expression*, are reserved for one class of familiar expressions: the most targeted and celebrated within the realm of familiar language, with its colorful members, idioms, expletives, and conversational speech formulas. It is a goal of this book to clarify and establish classes and subsets based on linguistic, psychological, and neurological benchmarks.

There are many possible and viable classification systems for fixed expressions. A classification system and criteria for labeling of subsets are presented in these pages. In this treatment, exemplars of familiar language in the current treatment fall into one of three classes, based on their distinctive characteristics. The complex features within classes and subsets will be described and illustrated. As will be amply demonstrated, for a given individual utterance token, there is copious evidence that speakers

[1] The term *familiar* will benefit from a precise definition. The Google dictionary defines *familiar* as *well known from long or close association*. Throughout these pages, the term *familiar* consistently means *known* and *personally relevant*, and *known* implies stored in mental language representation. An expression may qualify as *familiar, i.e. known and stored in mental representation*, from either long or short association.

[2] Formulaic expressions appear to be present in other, perhaps all, languages. This assertion is supported by performing internet searches for "Idioms in ____," inserting *Arabic, Swahili, Finnish, Chinese, French, Afrikaans, Turkish, Athabascan*, and so on. Lists are provided in every case; similar results arise from searching on proverbs.

have knowledge of complex, layered nuances and connotations inherent in fixed expressions.

Exposure to the intrinsic features of familiar language provides comfort and humor, as will be amply demonstrated in the pages to come, referencing material from a large range of sources. All fixed, familiar expressions offer potential for creativity, in the eternal dance between theme and variation. These expressions offer benefits to the speaker, buying time, organizing the discourse, and arranging packets of information, and to the listener, by giving a rest in processing effort and allowing a better grasp of the speaker's world (Wray, 2002).

Definitions of "Familiar Language"

Back to square one

The rubrics *fixed*, *familiar phrase* and *fixed*, *familiar expression* are employed here to refer to expressions that have unitary structure and are known to a speech community. Alternate terms are *prefabricated* (sometimes abbreviated to *prefab*) and *fixed expressions*, to be used interchangeably with familiar expression and familiar phrase.[3] *Multiword expressions*, another useful and current term, is used less here, because many single-word expressions fall into our purview: *Hello, shoot!, well, sorry, heck, bullocks, right, righto, pardon*, to mention a few.

Given the heterogeneity of these expressions, the most inclusive and the best operational definition for formulaic expressions is an exclusionary one: exemplars of familiar language have in common that they are not newly created, in the moment, from the operation of grammatical rules on lexical items. It follows that they are known – stored in memory (Bybee, 2006). They are familiar in this special sense. They are stored in the mind in a cohesive form. Familiar expressions are known in stereotyped, unitary form, with conventional meaning, and pragmatic contingencies – the social circumstances appropriate to their use – by members of a language community. Cohesive form usually includes certain words in a certain order (while allowing for flexibility of form), often with a prosodic and phonetic signature. With respect to the three classes proposed here, most familiar expressions can be comfortably classified by family resemblance, while many fit well into more than one class as will be demonstrated below. Important in this treatment is consideration of the properties inhering in the different types of expressions.

A familiar expression is generally defined here as a word or multiword sequence that is known to the speaker-hearers of a language as having a special status in the language community (Alexander, 1978; Altenberg, 1998; Bobrow & Bell, 1973). The expression is known to be recognizable not only to oneself but also to other speakers of the language (Kecskés, 2000). These expressions have unique characteristics and specialized functions in verbal and written discourse. Detailed descriptions of the use of fixed expressions in written and spoken discourse reveal that all and any of these characteristics can be transmuted in various ways to achieve a particular

[3] I avoid acronyms and abbreviations of terms.

communicative purpose, providing that Kuiper's "Law[4]" is upheld: the known, canonical form is retrievable (Kuiper, 2009). Flexibility of form varies with classes and subsets of familiar exemplars (Fellbaum, 2015). In many items, words can be inserted, morphemes (minimal units of meaning, such as suffixes) and syntactic operations may adjust the original grammatical structure.

This considerable scholarship has revealed new realms of known, familiar expressions and it has led to awareness of their fundamental differences. It is evident, for example, that the expression *Shut your mouth!* differs in several identifiable linguistic features from *All things being equal.* Similarly, *I also had an ax to grind* is not to be comfortably classified with *bride and groom*. Yet all these expressions are familiar and they are cohesive in form. Adjusting to developments in the past few decades, the current treatment proposes three classes of fixed, familiar expressions: formulaic expressions (e.g., *Shut your mouth!*), lexical bundles (e.g., *All things being equal*), and collocations (e.g., *bride and groom*).

The term *formulaicity* has been used to refer to the tendency for words to appear together into recognizable phrases; these groupings have been tied to frequency counts in corpora. A major contributory role of text frequency, as nearly fully accountable for familiar expressions, has been put forward (Arnon & Snider, 2010; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Kapatsinski & Radicke, 2009). Noteworthy questions arise here. Does high frequency correspond to and account for all kinds of prefabs in natural language? Can utterances be highly frequent but not a member of a familiar language class? And, conversely, can infrequent expressions be classified as familiar? We will explore the extent to which frequency of many fixed, familiar expressions in speech is an effect rather than a cause of usage. Further, how are we to distinguish fixed, familiar expressions from concatenations of words that closely reflect the transitional probabilities of meaning and form in a language, as routinely generated in auto-prediction processes in texting and emailing? Statistical properties of language also predict specific sequences. In this book, one of the tasks is to propose principles that distinguish these different influences on the acquisition and processing of familiar phrases.

Sketchy Reputation and Revival

Down through the corridors of time

The value of familiar expressions in communication has not always been fully appreciated in all realms. The aura around the term *cliché* reveals something about their reputation; the connotations resonate with shallow, superficial, insubstantial, insignificant (Miller & Villarreal, 1945; Lindauer, 1968; Redfern, 1989). This common sentiment is encountered in humorous material. A tell-tale cartoon in the *New Yorker* by Donald Reilly depicts men, women, and children standing on a cliff overlooking a scenic canyon, near a signpost displaying the words "Inspiration Point," each person with a thought balloon consisting of a "cliché." This trope is schematically represented as

[4] The term *Kuiper's Law* was originated in these pages.

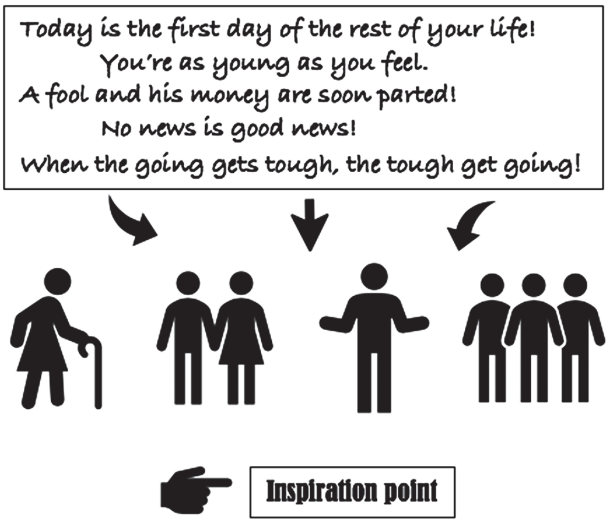

Figure 1.1 Comment on the reputation of clichés, such as *you're as young as you feel.*

Figure 1.1, providing an ironic display of superficial thoughts, in the form of familiar expressions, as people stand at an scenic view traditionally regarded as inspiring. This cartoon tells us several things. It says that the cartoonist, Mr. Reilly, knows these expressions and that he expects the readers to know them. The artist is poking fun at the average person's lack of creativity, and he is using prefabricated language to make this point – exposing tourists for being superficial.

Many kinds of familiar expressions have long been burdened by a reputation that they are shallow, simplistic, irrelevant to real cognition, and on the periphery of human language ability.

Major credit cards welcome;
The more it snows, tiddely pum;
Never stiff, never greasy;
Vroom! Vroom!;
Did I leave home without it?
Hold the mayo.

Another *New Yorker* cartoon by Henry Martin shows a banner at the top with the words "What's on the mind of America?" The thought balloons over people walking amongst the crowd contain these familiar expressions, arising from varied sources , as shown in the box above. Here, again, their alleged trivial content is brought forth to indict mundane thinking in the populace. In a similar vein, Jack Ziegler[5] produced a

[5] In most examples cited in this book, dates of cartoons are not included; cartoons sometimes appear in more than one publication, and they are present and available on the internet in *New Yorker* Cartoon compilation.

cartoon in the *New Yorker* magazine depicting motorcycle riders, each with a slogan on the back of their shirts: *Cliché masters.* Presumably the presence of this slogan, entailing a small-minded, light-hearted content, is ironically at odds with the reputation of motorcyclists as tough, intense, and manly.

Despite a certain whiff of low-brow repute that drifts around the topic from time to time, the presence and persistence of familiar language – our cover term for prefabs of all kinds – cannot help but engage interest and imagination from many quarters. Early signs of interest in the linguistic community, in the past few decades, have burgeoned into a full-fledged movement. The subject matter does well in the classroom. As this author can attest, students love the topic of familiar language. They ramp up quickly to understanding and enthusiasm given even the briefest introduction and a few examples. Smiling and hearty participation ensues. Given the opportunity, creative ideas emerge in the form of projects and papers, many arising from personal interest of some sort of other. Below is a list of topics generated by the class members in Nonliteral Language Course offered at Carleton College, 1999, in Northfield, Minnesota, all resulting in outstanding research papers (Sidelight 1.1).

Sidelight 1.1 Students' self-generated research topics in class at Carleton College, 1999

1. Survey of origins and familiar ratings of familiar expressions
2. Familiar expressions in Old and New Russia
3. Translations into German of familiar expressions in Calvin and Hobbes cartoons
4. Phonology versus lexicology in idiom recognition
5. Use of familiar expressions compared in TV vs radio transmission
6. Grammaticality judgments of transformed idiomatic expressions
7. How geographic location affects speech patterns in familiar expressions
8. On-line processing differences in propositional and formulaic sentences
9. Preserved automatic speech in transcortical sensory aphasia
10. Differences in perception of spoken versus written sarcasm
11. Advertising and nonliteral language: use and recognition studies
12. Role of popularity in familiar expression usage in comic strips
13. Minnesotan familiar expressions by natives and others
14. Idiom decomposability reconsidered
15. The role of familiar expressions in rap music: the language of rap
16. Familiar expressions as used in taboo topics
17. Familiar expressions in Presidential Addresses from a historical perspective
18. Gender differences in familiar expression usage in pop TV shows
19. Familiar expressions in Middle English Chaucer and Modern English translations
20. Familiar expressions in chat rooms: role of experience of topic

21. Comparing nonliteral language in sports and news
22. Nonliteral language in James Bond movies: use and alterations
23. Familiar expressions in speeches versus interviews of B. Clinton and R. Reagan
24. Familiar expressions in the media
25. Familiar expressions in *Rolling Stone* versus *New Yorker* magazines
26. The gendering of familiar phrases
27. Role of familiar expressions in ratings of movie dialogues, good versus bad

I urged some of the undergraduate authors of these studies to pursue publication, but they were more interested in graduating.[6] Since then, a variety of research questions of interest, imagined and conducted by students, have found their way into print, many of which are cited in these pages.

Early and Current Commentary

Give it your full attention

Varied perspectives on fixed, familiar expressions can be found post-onset of the generative grammar movement (from 1957). Some early linguistic commentary about idioms, a prime subset of the formulaic expression class, revealed uncertainty about how they were to be accommodated in the contemporary model of language. Chafe (1968, p. 109) and Fillmore, Kay, and O'Connor (1988) described idiomaticity as an anomaly in generative grammar. Similarly, in his chapter "Problems in the analysis of idioms," Weinreich[7] apologized for taking up "so unfashionable a topic," "idioms," that for many "would surely smack of the most outlandish romanticism" (1969, p. 23).

In other cases, the presence of fixed, familiar expressions in human language was acknowledged with interest. In the early 20th century, de Saussure (1916[8]) famously brought to attention *locutions toute faites* (*fully formed expressions*). Jespersen (1933) stated that formulaic characteristics pervade English grammar and participate in diachronic processes. Jerry Sadock, in 1972, noted that idioms and literal utterances involve two very different sorts of semantic structures, and Zwicky (1978) referred to an idiom as a combination of words with a meaning associated with it as a whole rather than by compositional principles: his examples were *a fat chance, trip the light fantastic, that cat is out of the bag*.

[6] For class syllabus and activities, see Van Lancker Sidtis, 2011.

[7] This article by Professor Weinreich, who taught linguistics at Columbia University, was published two years after his demise in 1967.

[8] De Saussure's lectures were printed posthumously in later editions.

In his one-time popular text book, John Lyons (1968) advocated separate analysis of "ready-made utterances," describing them as learned as a whole and passed on from one generation to the next. He noted that they are "not profitably regarded as sentences… Their internal structure…is not accounted for by means of (grammatical) rules" (p. 177). Otto Jesperson, in *Essentials of English Grammar*, described the "important distinction between formulas or formular units and free expressions… pervading all parts of the grammar" (1933, p. 18). According to Jespersen, formulas are "felt and handled as a unit" and therefore may involve different kinds of mental activity (pp. 19–21). With his usual prescience, Bolinger's (1976b) review of Makkai (1972) includes this comment (p. 238):

> We are just beginning to realize how much of the competence we hear so much about is carried in our heads as prefabs, with or without the interior vision of the assembled parts.

In a chapter called "Collocation and commonplace knowledge," Tyler (1978) refers to "whole units which we do not, on every occasion of their use, assemble from their component words" (p. 230). The psychologist Lounsbury (1963) outlined crucial differences between novel and fixed, familiar expressions succinctly (p. 561):

> Of two constructions made according to the same pattern, one may be an ad hoc construction of the moment and the other may be a repetition or use of one coined long ago, often heard, and much employed as a whole unit, e.g., as an idiom, a cliché… It is apparent that as behavioral events they are quite different and that in some sense their psychological statuses and in the actual speaking behaviour may be quite different…Some of this is old, familiar, and quite automatic at any given time—some of it is new as of the moment and may even be hesitatingly put together.

Despite these tentative signs of a certain appreciation of fixed expressions, linguistic interest during those years focused overwhelmingly on syntax and ways to generate new sentences. Despite the environment of linguistic thought at the time, two bombshell publications, appearing at about the same time in different edited volumes (Fillmore, 1979; Pawley & Syder, 1983) stimulated interest far and wide. These two key contributions to the field made similar points: familiar, fixed expressions occur throughout natural speech, and failure to incorporate these expressions renders a speaker less fluent or sounding nonnative. These ideas are echoed by Kecskés (2010). The arguments and the examples were compelling. Fillmore designed a course on this theme and accumulated a list of "formulaic expressions" from contributions by students at UC Berkeley (see Appendix I).

These early adumbrations have radiated and expanded throughout linguistic studies into a fully elaborated field of research. Although skepticism remains in some quarters, familiar language of various kinds has recently gained a modicum of appreciation in current scholarship, with the strongest presence in second language learning, considerable interest in other branches of the linguistic sciences, and creditable activity in psychology and neurolinguistics (see Pawley, 2007 for a review). The topic remains a challenging one. Stress and controversy accompany definitions, reliable identification of exemplars, consistent classifications, as well as numerous theoretical points, such as

the role of text frequency (frequency of exposure and use) and the durability of phrase-internal cohesion.

Along with these controversies, study of familiar language now takes a prominent place in many branches of language studies (Čermák, 1994; Heringer, 1976; Osman, 2009; Pawley, 1985; 1986; Jackendoff, 1995). It is broadly recognized that known, unitary expressions thrive within many cultural entities: country, region, town, family, work place, literature, communication medium, and profession. Any of these domains can and will establish a repertory of holistic expressions known to members of the speech community.

While human language, for several decades, was described as a homogeneous system comprised of morpholexical items and rules for grammatically combining them, there is, in addition to this competence, a natural capacity to acquire holistically established material, recognizable to the cohort as "known" and cohesively engendered. Two modes of speaking have been identified: speakers either originate expressions or produce expressions that are accessed from a pre-established inventory (Bolinger, 1961b, p. 381; Sinclair, 1987; 1991; Erman & Warren, 2000). It is a goal of this book to present the state of the art in these fields of study, as they examine familiar language, and to more fully describe the holistic mode of processing in mind and brain.

Relevance to Models of Language

There's more to this than meets the eye

The field of linguistics currently has benefit of numerous schools of thought, leading to various approaches to modeling grammar and language use. A few will be mentioned here. Sociolinguistic studies and the related endeavor, pragmatics, advance excellent potential to contribute to the understanding of fixed, familiar expressions. Early profiles of frame semantics yielded useful insights (Fillmore, 1977). Numerous linguistic treatments outside of the generative model (Lakoff, 1987; Fillmore, Kay, & O'Connor, 1988) cast light on various shades of the topic. Cognitive grammar grasps the complex meaning element (Langacker, 1987b; Pagán Cánovas & Antović, 2016). One depiction of this approach posits that grammar constitutes the cognitive organization of experience (Bybee, 2006). These models, themselves, are closely allied with construction grammar (Bybee, 2013; Ellis, 2008c).

W. N. Francis stated that the old axiom from Euclid, *the whole is equal to the sum of all its parts*, does not apply to organized wholes.

> An organized whole is always *greater* than the sum of all its parts, because it is equal to the sum of its parts *plus the way they are organized*. (1958, p. 28)

Formulaic expressions (e.g., idioms, conversational speech formulas) take on this kind of extra meaning. The extra meaning is nonliteral and often fraught with connotations. For lexical bundles, meaning arises from the familiarity and stability of the cohesive form and its structural role in written and spoken discourse. In the remaining group, collocations, meanings tend toward literal and utilize metaphoric

rather than nonliteral processes, with connotations arising from the constituent lexical items. All of these forms, formulaic expressions, lexical bundles, and collocations, are well accommodated by construction grammar. This model of language builds on an earlier, similar approach developed in the days of structural linguistics, dedicated to

> "…discovering and describing as concisely and as accurately as possible the interrelationships and patterns which make up the intricate structures of language". (Francis, 1958, p. 26)

The body of theory evolving in construction grammar differs from the earlier descriptivist approach (Francis, 1958) above in being based in a holistic and usage-based framework that merges meaning and syntax (Goldberg, 2006, 2009). Construction grammar, which arose out of linguistic studies in the 1980s, now has the form of several models, and is better thought of as a grouping of theories. Any type of construction, described as a pairing of sound and meaning, contributes to standardized grammatical patterns, and linguistic knowledge is presumed to emerge from usage (Table 1.1).

This approach is highly compatible with the characterization in these pages of familiar language (Buerki, 2016). Constructions can occur on various levels of abstraction, from fully abstract scaffolding (grammatical categories and structure) to containing one or more actual words or with all the words in place (Michaelis, 2017a, 2017b; Blumenthal-Dramé, 2012). Many different words can appear at the various levels of the phrasal scaffolding. Constructions can be said to have schematic slots (Bybee, 2013). As constructions, idioms are the least abstract, because key words in the idiom must be present. Idiomatic meanings, being nonliteral, are nonstandard and conventionalized, and must be acquired idiosyncratically. Lexical bundles are less flexible than idioms: *in the meantime, as I was saying, at this point in time* are rarely manipulated for communicative purpose (except humor). Collocations vary widely in flexibility and manipulability. Familiar phrases can be viewed on a continuum where

Table 1.1 Examples of a possible range of constructions.

Linguistic unit	*Examples*
Word	*tentacle, the*
Word (partially filled)	*VERB+ing*
Complex word	*textbook, drive-in*
Idiom (filled)	*like a bat out of hell*
Idiom (partially filled)	*believe (one's) ears/eyes*
Covariational	*The more you watch the less you know*
Ditransitive	*She gave him a kiss; he fixed her some fish tacos.*
Passive	*The cell phone tower was struck by lightning.*

Adapted from Goldberg, 2006.

degree of cohesion is a parameter running from high to low. The most cohesion appears in lexical bundles. The role of frequency and degree of connotative nuance also vary with these classes of familiar language. These points are elaborated and illustrated below.

Translating Familiar Expressions

Flying by the seat of their pants

The perils of translation take a dazzlingly disorganized turn in a 19th century publication offering English versions of the Portuguese language, apparently achieved via benefit of French–Portuguese and French–English phrase books (Carolino, 1984). The French collocation *vous m' avez fait trop attendre* (*you made me wait too long*) is rendered as *you have me done to expect too.* Countless such examples, mutilating English phrases almost beyond recognition but using recognizable words, arouse hilarity in readers of this scandalously ill-conceived work.[9] When portions of language are described as a collage of stored verbal Gestalts rather than a system of rules and discrete objects, translation between languages requires an especially thoughtful strategy. Fixed, familiar expressions present special challenges to persons translating between two languages. How do the three classes of fixed, familiar expressions fare in the translation process? Given their similar but differentiating features, members of the three classes can be expected to differ in the challenge to translation (Weinreich, 1969; Gulay, 2018; Fernández-Parra, 2008). Of the three classes, formulaic expressions probably require the greatest depth of knowledge and expertise. Bar-Hillel (1953) named idioms as one of four major challenges in translating between languages. The meaning process of idioms is nonliteral, and so rather than a translation, an equivalent match must be found in the target language. Speaking of situation-bound, routine formulas, Coulmas (1979, p. 339) notes the need to bring entire "systems of belief, wants, wishes, preferences, norms, and values" to bear on their translation. To understand competence for these expressions, a cognitive-pragmatic approach is recommended. Extralinguistic factors interact with the linguistic contingencies of situation-bound formulas, leading to the view that specific kinds of knowledge are in play (Kecskés, 2000).

Lexical bundles can more easily find their linguistic mates, but many of these, too, have conventionalized meaning, function, and form. An expression in the target language must be found that does the same work in discourse. What about collocations? It's been said that all languages have certain ways of saying things. Although collocations are more likely to embody standard grammatical structures and convey usual lexical meanings than the other two classes, it remains a fact that collocations are made up of certain words combined in certain ways (Gries & Wulff, 2005). Successful rendition of collocations will require deft of hand and an artistic eye and ear in the

[9] It is worthy of contemplation that producing familiar phrases in creatively distorted ways is often very funny.

translation process. If, as hinted by Professor Fillmore, the entire texture of human language is shot through by formulaicity, then translators are flying by the seat of their pants and they deserve our deep admiration. Perhaps these textured differences account for Steiner's (1975) view that each language has its own intrinsic, essentially untranslatable essence.

Incidence of Familiar Language Exemplars

It's a jungle out there

With the goal of characterizing human language and developing models of language competence, linguists have traditionally engaged in the practice of generating samples of language out of whole cloth (as it were), producing words, sentences, idioms, and other expressions, which are submitted to the linguists themselves and/or to experimental subjects for endorsement (as grammatical or natural), ratings, judgements, and other responses. In a more naturalistic approach, corpus studies utilize computer sweeps to uncover actual usage in various types of discourse, focusing on a selected linguistic question. Still another source of information about the nature of familiar language and its storage and processing by users lies in the media. Examples from newspaper articles, ads, and cartoons reveal that speakers know fixed expressions and, further, these exemplars from the media offer blatant proof about what is known: the nuances, appropriate social settings, their form and meaning, their possible provenance, and the legal limits of distorting the expression for altered communication or humor. In these pages, copious examples from the media are included to expose knowledge by the language community of the large repertory of familiar phrases and their complex constellations of features.

This is the era of corpora. Transcriptions of spoken discourse and written texts from many disciplines have been made available for linguistic analysis (Altenberg, 1991). Going beyond earlier frequency word counts, corpora, with the help of methodologies for their use (Burnage & Dunlop, 1992a, 1992b), offer actual use of language in many contexts. This provides a great wealth for the study of familiar phrases.

Everyday Discourse

Like water off a duck's back

Quantification in language is challenging work, qualified by age of the users, definitions of linguistic objects to be counted, and the sources of the language information. Situation parameters influencing the type and amount of usage include the participants and how they are related, their attitudes toward each other and the talk, whether the situation is formal or informal, the topic or theme, and the type of discourse (conversation, lecture, or pronouncement) (Biber, 2001). Approaches to quantification of word and phrase knowledge have led to estimates (for an adult at age 20) of 42,000 lemmas (word types) and 4200 nontransparent multiword expressions (verb plus particle and idioms), with acquisition in increasing years of 6000 new lemmas. When inflected and combined forms are considered, word repertory is very large (Brysbaert, Stevens, Mandera, & Keuleers, 2016).

The number of fixed expressions known to members of a language community, when all classes and subsets of known expressions are considered, may be very large, approaching tens of thousands (some say more) and they manifest a dizzying variety of types and shapes. Estimates range from one third to one half of language product of the language user (Schmitt, Sonbul, Vilkaité-Lozdiené, & Macis, 2019). Many incidence studies have combined various types of familiar expressions, while others have made separate counts of subsets, such as idioms, proverbs, or collocations. Various estimates emerge from counts, depending on many factors: criteria for identification, targeted expressions, size and type of corpus, and the selected algorithms used on the samples. Cowie (1992) performed a study on verb–noun collocations in newspaper language, concluding (p. 1):

> …the high incidence of such familiar expressions in news coverage suggests that the professional skill of reporters owes less to verbal inventiveness than to the memorisation and re-use of existing locutions.

Biber, Johansson, Leech, Conrad, and Finegan (1999) estimated that lexical bundles occur between 10 and 40 times per million words. Using computer-search criteria, Altenberg (1998), estimated that London-Lund Corpus (Greenbaum & Svartvic, 1990) contained 80% recurrent word-combinations. Soskin and John (1963) found that 75% of expressions were other than information statements; this classification system was based on meaning and intention of the speaker.[10]

Moon (1998c) performed a descriptive study of formulaic expressions and idioms in an 18 million-word corpus of contemporary English, the Oxford Hector Pilot Corpus (Glassman et al., 1992), and other texts. Many known expressions had 0 occurrence, and overall, idioms occurred only rarely. Moon (1998b) reports that "simple formulae" are the most frequent overall. Strässler (1982) counted about one idiom in every four and one half minutes of text. Norrick (1985) reports only one complete proverb, plus a few proverbial allusions, in the 43,165 line corpus transcribed conversation published by Svartvik and Quirk (1980).[11] A comparative frequency count of proverbs in French and English conversational corpora was once described by Arnaud and Moon (1993).[12] Using spontaneous speech collected in Canadian corpora, Sorhus (1977) found that approximately 20% of the material consisted of fixed expressions.

Fixed, familiar expressions are frequently repeated in conversations, a strategy that further enhances affiliative sentiment. In telephone conversations (Kingsbury et al. 1997), 75% of all verbal repetitions were performed on familiar phrases (Van Lancker

[10] C. Fillmore, who was among the first to publish papers bearing the banner for familiar expressions, many years later stated that he believed nearly all of language to be formulaic. Instead, he mentioned, it is the novel expression that requires defense (personal communication).

[11] A comparative frequency count of proverbs in French and English conversational corpora was once described by Arnaud and Moon (1993).

[12] This book is no longer available.

Sidtis & Wolf, 2015). A transcript from a reality show revealed that 64% of repetitions were familiar language.

174.81 Speaker A: The week nights is pretty crazy.
257.00 Speaker B: um and like like fruit and oh my god it's just crazy
260.90 Speaker B: It's crazy how much he eats.
277.55 Speaker A: That's pretty crazy.
395.93 Speaker B: Now wait. What, got so much furniture in this tiny apartment. It's crazy

Overall, studies generally indicate that familiar language, known and stored prefabs, however named, classified and quantified, constitute a significant proportion of discourse, although a consensus number cannot realistically be established.

Literary Incidence

It has been brought to my attention

The use of fixed expressions as poetical device has been described by literary scholars. The efficiency of these expressions to establish nuanced meanings is exemplified in Poe's poem, *The Raven*. The word "nevermore" occurs 11 times, repeatedly accumulating connotations of loss and dread. This word has taken its cultural place in English speaking countries, as can be seen in creative humor. George Price in a New Yorker cartoon depicts a home, where two men are seated around a table, women are in the kitchen. On another side of room, a much older woman with kerchief and scrawny hands is sitting in a rocking chair looking at a television placed close to her on a stand. One man says to the other, looking at the older lady:

Gwen's aunt. She came upon a midnight dreary.

The expression in the cartoon alludes to the first line of Poe's poem *Once upon a midnight dreary*, evoking the dread and negativity of the setting. This example is one of thousands exhibiting the sheer fun of alluding to a familiar expression.

The extensive presence of familiar language in poetry and other literature, as well as song, and its effects on content and expression, and its impact on the reader can easily be shown. In some literary texts, incidence of familiar language is greater than in spontaneous speech (Kiparsky, 1976; Kuiper, 2000), presumably for artistic effect. Tilley (1950) identified numerous proverbs in the plays of Shakespeare, some of which, ironically, can claim their origin in those pages. In his studies of German novelist Günter Grass, Schweizer (Schweizer, 1978; Calhoun & Schweitzer, 2012) listed 194 idioms in 2876 pages of six novels, about 0.067 idioms per page. This corresponds to quantities reported in corpus studies listed above. The plays of Ionesco (e.g., *Les Chaises*, 1959) utilize an abundance of speech formulas to artistic effect (Klaver, 1989). An O. Henry story, *Whirligigs* (1920), is based on clichés.

Scholars of ancient poetry believed that the frequent occurrence of fixed expressions in the Iliad and the Odyssey reveal an essentially oral creation, fashioned to aid memory (Stolz & Shannon, 1976). Lord (1960) and Parry (1971) analyzed the presence and role of such language in the Odyssey and the Iliad. Hainsworth (1968)

elaborated on the Homeric style by showing the actual flexibility of the structures. Page (1959) estimated that about one-fifth of the Iliad is "*composed of lines wholly repeated from one place to another*" (p. 223), and that within the Iliad's 28,000 lines, there are approximately 25,000 repeated phrases. The Old English epic Beowulf contains less verbatim repetition of formulas than Homer; relying more on alliteration and rhyme to aid the memory of the bard (Foley, 1978). Performers probably memorize outline of the epic stories, some of the transitions, and then produce the formulaic and alliterative expressions that are provided. Similar devices have been described for African poetry (Gintsburg, 2019). Poetic devices, including assonance and rhyme, contribute to the retention of ballads across generations (Rubin, 1995).

A study of the incidence of familiar phrases in the screenplay *Some Like It Hot* (Wilder & Diamond, 1959) was conducted. Idioms, proverbs, and speech formulas were found to make up 24% of the total number of phrases in the text (Van Lancker Sidtis & Rallon, 2004) (Appendix IX), mostly conversational speech formulas, with some idioms and a few proverbs. Classification of these expressions in the screenplay confronts the usual challenges: examples are *All right, we're all set, When is the kick off; I better blow now; to err is human*. Of these, the majority (76%) are single occurrences of the expressions, manifesting the abundance of available familiar expressions. However, familiar expressions were also copiously repeated. In a later study of utterance repetition in *Some Like It Hot*, it was seen that 48% of all repeated material occurred on familiar phrases (Van Lancker Sidtis & Wolf, 2015).

The expressions in the *Some Like It Hot* screenplay were identified by the experimenters using native speaker intuition as well as formal and functional criteria (Figure 1.2). As a check on these selections, other speakers' knowledge of the expressions was also probed. A study designed to challenge the identification processes of the authors was conducted using a fill-in-the-blank (cloze) and a rating format. In the cloze procedure, 75 selected fixed-familiar and 25 novel expressions from the screenplay were randomized, each with a word missing. Participants were requested to fill in the word most appropriate to each item. For the recognition portion, the novel and fixed utterances were again presented in full and the task was to circle F (formulaic) or N (novel). (See Appendix XV for the protocol's randomized list and sample responses.) An example of each is given below.

Recall task:
That's a quarter of a _____. (novel expression)

Recognition task:
I ought to have my head examined! **F N** (fixed expression)

The results in Figure 1.2 support the intuitive identification of familiar versus novel utterances by the experimenters, showing significantly greater conformity in words entered for the formulaic than the novel expressions in the recall task, and high recognition of mode, formulaic or literal, of all the items.

Reflecting their role in transmitting wisdom, proverbs made an appearance only toward the end of the play, in the 15th, 18th, 20th, 21st line sets, where resolutions were reached, as can be seen in Figure 1.3.