What Fixed Expressions Reveal About Grammaticization

Jaden Daniel González

Department of Linguistics, Yale University

April 2025

Advisor: Maria Mercedes Piñango

Acknowledgements

To my cohort of fellow Class of 2025 Linguistics seniors:

Thank you for your moral support and for sharing your experise, and congratulations! You are all incredible people and academics, and I feel so lucky to have worked so with and alongside each of you.

To the members of the Yale Language & Brain (LaB) Lab:

Thank you for welcoming me into this community of support. Your weekly encouragement and community of support.

To my second reader Tom McCoy:

Thank you for your feedback and words of encouragement, which always came at the right time.

To Raffaella Zanuttini and Jason Shaw:

Thank you for your advisory and unwavering patience throughout this process.

To my primary advisor, MMP:

Thank you for your continued dedication to my growth as a student and linguist. Your encouragement (and accountability!) led me to discover and center the scientist within my approach to this field. Your generosity and compassion will continue to have a lasting impact on me as a researcher and individual.

Abstract

Fixed expressions have received much attention in the literature due to their unique properties and behaviors, including (but not limited to) stereotyped form, conventionalized meaning, resistance to syntactic transformation. The juxtoposition between these properties and the aparent compositionality these expressions demonstrated by comprising multiple lexical items has given rise to a body of literature surrounding questions about how to design their underlying phrase structure. Motivated by the observation that speech formulas and other fixed expressions tend to be preserved in cases of aphasia with severe impairment generative, this paper seeks to reconcile the contrast between fixed expressions and novel speech within a generative framework. This proposal implements existing frameworks of lexical interface to argue for a bidirectional pathway from compositional to holistic representations in the Mental Lexicon. By this model, fixed expressions move fluidly along this continuum by a process called "grammaticization," which refers to the parsing of stored gestalts into operationalizable according to acquired grammatical conventions. To illustrate this mechanism, I consider two specific types of fixed expressions: formulaic speech and idioms. Crucially, my analysis of these phenomena considers syntactic derivation to be a process of construal that is conditioned by metalinguistic idiosyncracy.

Introduction

Generativity is the foundational theory of linguistic analysis. The generative framework is characterized by a categorical distinction between grammatical and lexical features of language. By this framework, phrase structure is generated by the Grammar and populated by discrete items stored in the Lexicon to be realized as language. Fixed expressions present a unique and noteworthy challenge this framework. A fixed expression can be defined as a communicative utterance comprised of multiple lexical items and whose form and meaning exhibit some degree of conventionalization. Idioms—such as [kick the bucket], [spill the beans], [cat's out of the bag]—are one example of this phenomenon. Other examples include speech formulas, also called formulaic speech. This refers to a range of discourse rituals such as greetings and farewells (*Hi, how are you? I'm fine, and you?*) as well as conversational interjections such as ([*Get out!*], [*No way!]* [*Oh, I see*]).

Many linguistic analyses parse between lexical items as a diagnostic for identifying underlying phrase structure. These approaches tend to treat prosodic words as atomic units, except for morphemes that can be attributed to a single terminal node (i.e. tense markers). Fixed expressions are generally parsable into multiple prosodic words which, by this diagnostic, is indicative of phrasal composition mechanisms that existing frameworks deem attribitable to the Grammar alone. Nevertheless, the conventionalization of their specific composition and meaning suggests a holistic representation, which are generally attributed to the Lexicon. This juxtoposition has been notoriously difficult for linguists to reconsile, as it requires direct immersion within the complex interface of these pillars of the language faculty: the Grammar and the Lexicon.

Research Question

This essay expands upon existing theoretical frameworks of the language faculty to propose a plausible, research-based explanation for the properties and characteristic behaviors of fixed expressions. The proposal takes the form of a narrative review, which is motivated by the following question: How can a generative framework of the language faculty account for fixed expressions?

<u>Roadmap</u>

I begin with a brief review of critical proposals concerning the architecture of langugage cognition. This is meant to situate the reader within the current understanding of language processing accoridng to existing literature. I then narrow the focus on fixed expressions by outlining the linguistic contrast between novel and formulaic speech, also consulting the relevant literature and identifying the area for futher contribution this essay seeks to address. I then introduce my proposal for an additional cognitive mechanism of natural language which I term *grammaticization*. I define this mechanism as a bidirectional pathway between holistic and compositional representation in the Mental Lexicon. In the following sections, I build my argument by implementing this expanded framework in my analasys of two linguistic phenemona: idioms and speech formulas. Finally, I conclude by discussing the implications of this proposal for approaches to linguistic analyses and potential areas for further research.

The Architecture of Generative Language Processing

Mainstream Generative Grammar (MGG) is motivated by a distinction between grammatical and lexical mechanisms of the language faculty. Linguists have significantly

expanded upon this framework to implement theretical models that account for natural langauge phenomenon that require more dynamic architectures of analyses, including fixed expressions.

<u>Parallel Architecture</u>

One such proposal is Jackendoff's (2007) proposal of a Parallel Architecture. This framework divides the language facuty into three separate but coordinated components: phonology, syntax, and semantics (see Figure 1). Jackendoff (2007) proposes each of these are independent computational systems which simultaneously compose their respective components of language. To illustrate, Jackendoff (2007) highlights the case of idioms—whose conceptual representation (meaning) is independent of the explicit lexical items which comprise the utterance (phonology) and whose structural composition (syntactic) is not responsible for its idiomatic interpretation (since the construction VP + DP-complement is not unique to idioms). The specific example he uses (*kick the bucket*) will also be considered in the present paper.
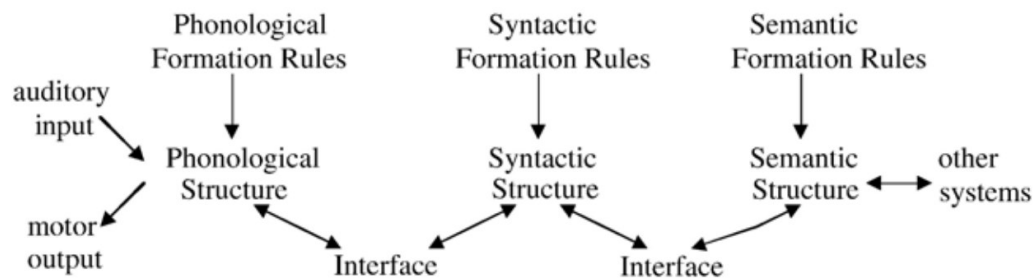

Figure 1. "The Parallel Architecture" (Jackendoff, 2007)

<u>Neuro-Cognitive Architecture of Language</u>

Piñango (2024) builds upon this framework by integrating neurological research concerning the role of long-term and working memory networks in language cognition to

propose a model of Real Time Memory-Based Meaning Construal (see Figure 2). Faithful to the Parallel Architecture, the model proposes that meaning is constructed dynamically as language is processed, relying heavily on memory retrieval mechanisms. Rather than deriving meaning from a fixed underlying grammatical construction, Piñango (2024) designs meaning as a process of construal achieved by accessing and integrating prior experiences—both articulatory and contextual—stored in memory to interpret incoming linguistic stimuli. Crucially, meaning is not discretely organized according to syntax categories; rather, it is construed in real time through an approximation of acquired associations between patterns of sounds or signs (phonological representations) and conceptual representations. These associations are inventoried in the Mental LexiconAmong the many virtues of this framework is that meaning need not be interpreted according to the apparent syntactic structure of the lexical items they comprise. Instead, an utterance may be treated as having flexible, context-dependent and experience-driven meanings independent of their syntactic and phonological topography.
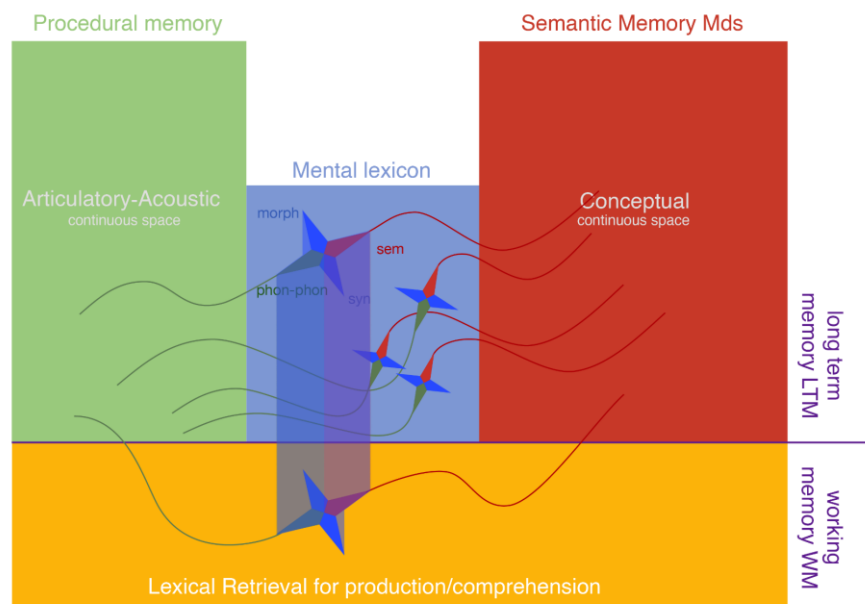

Figure 2. "The Dynamics of Language within its Memory Infrasturcture" (Piñango 2024, p. 4)

<u>Gestalt Processing</u>

Another critical contribution to this framwork, particularly the role of working memory in lexical encoding and retrieval, is the notion of gestalt processing. Blanc's (2012, 2013) six stages of Natural Language Acquisition (see Figure 3) suggests that children interpret linguistic inputs holistically (rather than lexeme by lexeme[1]) and use pattern recognition to identify meaningful chunks called *gestalts*. These gestalts serve as units of meaning and may comprise many lexical items. As children acquire more exposure to these patterns in various phonological and pragmatic contexts, they experiment by mitigate these gestalts into smaller units, decomposing and reassembling identifiable patterns contained within. By this process, typically developing children eventually learn to parse linguistic analytically according to grammatical conventions they acquire through both implicitly (by exposure and immersion) and explicitly (schooling and correction). Nevertheless, this model centers perception and experience as the driving force of lexicalization, like Piñango (2024), but from the vantage of purely topographical observation rather than the neuroarchitecture of memory. In doing so, it foregrounds the inexrticable influence of both phonetic environment and pragmatic context in both meaning construal and linguistic structure. These mechanisms are crucially relevant to the presently proposed analysis of fixed expressions and their unique linguistic behavior.

---

[1] Here, lexeme refers to the any formal unit of linguistic meaning. This could be a morpheme or an entire prosodic word. Crucially, these are typically attributable to some degree of underlying syntactic structure, be it a terminal or entire X-phrase. A gestalt may correspond to lexeme boundaries, but such correspondence should be interpreted as purely coincidental for the purpose of this essay, especially at early stages of language development.

Stage	Examples
1. Communicative use of whole language gestalts	"Let's get out of here." "Want some more?"
2 - A. Mitigation into chunks 2 - B. Recombining these chunks	(1) "Let's get" + "out of here" (2) "Want" + "some more?" (1) "Let's get" + "some more?" (2) "Want" + "out of here"
3. Further mitigation: isolation of single words, recombination of words, and generation of original two-word phrases	"Get – more." "Want – out?"
4. Generation of first sentences	"I got more."
5. Generation of more complex sentences	"I wanna go out."
6. Generation of the most complex sentences	"I don't want any more, but you can have mine." "How long do you wanna play outside?" "Do I really have to go out to play today?" "How 'bout if you go out and play instead?"

Note. Adapted with permission of Marge Blanc, copyright owner. Originally published in *Natural Language Acquisition on the Autism Spectrum: The Journey from Echolalia to Self-Generated Language*, Madison, WI: Communication Development Center, p. 23.

Figure 3. The Six Stages of Natural Language Acquisition (Stiegler, 2015; adapted from Blanc 2012, 2013)

<u>Dual Process Model</u>

Continuing in the vein of developmental communicative sciences, the dissociation between "novel" and "formulaic speech" arose from studies of preserved speech in severe aphasia. Van Lanker Sidtis (2012) writes, "…following damage to the language areas of the brain, while newly generated speech is impaired (Code, 2005), in many cases, a great variety of overlearned expressions (different ones for different persons) are retained with normal-sounding competence" (p. 64). Here, "newly generated" or "novel" speech refers to utterances that are generated by the speaker in real time. Novel speech is originated by the speaker and is generally used for narrative function—something like "I went to the store today and saw a bird outside…

"Overlearned expressions" refers to phrases that have become conventionalized in form and usage due to high frequency; in other words, fixed expressions. These phrases tend to serve metalinguistic communicative functions but can still be used in a narrative context: "Yesterday,

Ester totally [spilled the beans]!". Idioms and a wide range of other fixed expressions (see Figure

4) have been classified in the literature as "formulaic speech" (Sidtis 2012; Torrignton Eaton &

Burrowes 2022).

Table 1. List of familiar language categories in order of Sidtis's (2004) continuum from most reflexive to most conventionalised, operational definitions, and exemplars.

Category	Operational definition	Examples
1 Pause fillers (Fuller, 2000; Clark & Fox Tree, 2002)	Five recognized, grammatically optional word and non-word items generally used during word retrieval or reformulation; annotated by "&" in CLAN	*um, uh, like, I_mean, you_know*
2 Interjectional phrases (Ameka, 1992; Goffman, 1978; Wilkinson & Kitzinger, 2006)	Response cries with emotional valence that reflect state of mind such as surprise, disgust or sympathy including expletives, and not directed to a listener (i.e., not in 2nd person)	*For Christ sakes! Oh man! Thank goodness!*
3 Vocatives (MacWhinney, 2000)	Utterance-initial discourse markers annotated in CLAN by "‡" that are not phatic interjections or parts of interjectional phrases	*well, so, okay, alright, anyway, yeah*
4 Phatic interjections Stivers, 2019; Ward, 2004)	Responses to direct or implied polar questions and statements requiring a variant of yes/no (including dialogue within story-telling), but excluding non-lexical sounds	*Nope. I'd say. Yes ma'am.*
5 Greetings and farewells (Ameka, 1992)	Conversational routines used specifically in the beginning or end of an interaction	*How've you been? Happy to see you. Gotta go.*
6 Speech formulas (Van Lancker Sidtis & Rallon, 2004)	Context-bound, able to stand alone, purposeful (i.e., realizing a pragmatic function)	*I have no idea. Just go with it. Thanks so much!*
7 Idioms and proverbs (Van Lancker Sidtis & Rallon, 2004)	Idioms: largely non-transparent, contextually-free expressions. Proverbs: expressions intended to make a point and may be non-transparent or literal.	*Pours out her heart. He cleans up well. The time has come and gone. We're gonna whip this thing.*
*8 Formulaic sequences (Jeong & Jiang, 2019; Nekrasova, 2009)	Structurally complete multi-word expressions (grammatically optional) with a unitary but neutral semantic representation and are relatively fixed	*that kind of thing, as a matter of fact, at that time, first of all, or whatever*
*9 Lexical bundle (Conrad & Biber, 2005; Jeong & Jiang, 2019)	Multi-word utterances with relatively neutral meanings that are structurally incomplete (spanning phrasal boundaries) and may perform a bridging function; includes sentence stems	*to the point where, that depends on, kind of like a, in the process of, I'm not sure if*

* indicates lexical bundle category in Sidtis's (2018) three-category model, whereas all other items are categorized as formulaic expressions.

Figure 4. "List of familiar language categories"[2] (Torrignton Eaton & Burrowes 2022, p.

1403)

---

[2] There appears to be some degree of variability regarding how these phenomena are termed in the literature. In my understanding, and for the purpose of this essay, "familiar language" is synonymous with "formulaic speech"—the former centers the method by which language becomes formulaic: overfamiliarization; wheras the latter centers the underlying cognitive processing that differentiates from "novel". Fixed expressions fall within this category of "familiar language" and are thus analyzed a type of formulaic speech,

Crucially, the term "formulaic" refers to how the speech is processed. According to Sidtis (2012), "Early hints that formulaic expressions are stored and produced differently in the brain arose from observations of adults with language disturbance, usually those with nonfluent aphasia due to a left hemisphere stroke, who preserve some kinds of speech while sustaining serious loss to generative language competence (Code, 2005)" (p. 68). This invaluable observation made especially salient the existence of a linguistic and cognitive contrast between novel and formulaic speech. This is contrast is formalized by the Dual Process Model. There is some strong evidence for localized differences in brain activity across these types of speech (Sidtis 2012). In this essay, I implement the Dual Process Model in conversation with the above frameworks to propose a grammaticization mechanism within the language faculty, which I support with analysis of the topographical (observable) properties and characteristic behaviors of fixed expressions.

Proposal

These theoretical models have contributed significantly to further articulating the complex mechanisms of the langauge faculty within the framework of generativity. In this essay, I attempt to implement these frameworks in conversation with each other to propose my own contribution to the discussion of language cognition. First, I posit that gestalt processing is a perseverating mechanism of the language faculty, which is supported by language acquisition research across developmental profiles (Stiegler 2015, Blanc et al. 2023). I argue this acquisition mechanism allows fixed expressions to become relexicalized independently of their constituent lexical items and develop holistic representations in the Mental Lexicon. I argue this accounts for

the retention of fixed expressions and other formulaic speech in cases of impaired generativity. Moreover, I argue that the intact language system can parse these holistic representations according to grammatical conventions in real-time during lexical retrieval. I propose this process of grammaticization restores the compositional represention of a given fixed expression, allowing processor to seemlessly integrate the expression into larger generative structures. As outlined in the following sections, idioms and other fixed expressions demonstrate this phenomenon by their ability to undergo syntactic transformations, such as V°-to-T° movement (He $_{TP}$[kick$_i$-ed $_{VP}$[$\_i$ $_{NP}$[the bucket]]].) or passivisation (even when unsupported by the semantics: #[The bucket was kicked by John]). Taken together, grammaticization can be conceptualized as a bidirectional pathway between compositional and holistic lexical represenations.

Formulemes

Sidtis (2012) outlines four characteristic properties of formulaic speech: conventionalized meaning, stereotyped form, cannonicity, and pragmatic hypersensitivity. Crucially, she employs the term "formuleme" to refer to the unit of language that fits this profile. In this section, I elaborate on these properties with examples and consider the unique linguistic bevaiors that arise as a result of these properties. I interpret these behaviors to reveal an underlying process of re-lexicalization motivated by gestalt processing. I illustrate how framework accounts for how compositional phrases may develop holistic representations—formulemes—stored in the Mental Lexicon.

<u>Fragmentation</u>

Speech formulas are what might be colloquially referred to as "non-literal". In other words, their meaning is not necessarily derived as "a sum of its lexical content" (Sidtis 2012, p. --). For example, the meaning of [*what's up!*]—something akin to "hey!" or "hi!"—is not a direct composition of the individual meanings of [*what*], [*is*], and [*up*]. Like any lexial item, the meaning and form of a formuleme may be considered independently despite being inextricably linked by association. Due to this conventionalized meaning, a speech formula may exhibit variation in form (including intonation and emphasis) across speakers or across instances (as seen below). Speakers might omit ([*what up*] or even [*sup*]), insert ([*what the hell is up!*]), or replace ([*what's going on?*]) items within the phrase. Nevertheless, a maximally representative base form is easily identifyable by speakers' intuition and is presumably reflective of the most frequent form. In this case, [*what's up*] is the stereotyped form of this paticular speech formula. These can be represented as alloforms: [what's up] ~ [what up] ~ [sup]; [it's a pleasure to meet you] ~ [pleasure to meet you] ~ [pleasure].

The alloforms of each formulaic greeting are interesting to consider from a syntactic perspective. In novel speech, the omission of words from these phrases might be analyzed as a case of ellipsis. By this account, the expression would be assigned a syntactic phrase structure representation, and the surface form is illustrated as a realization of elliptical substitution:
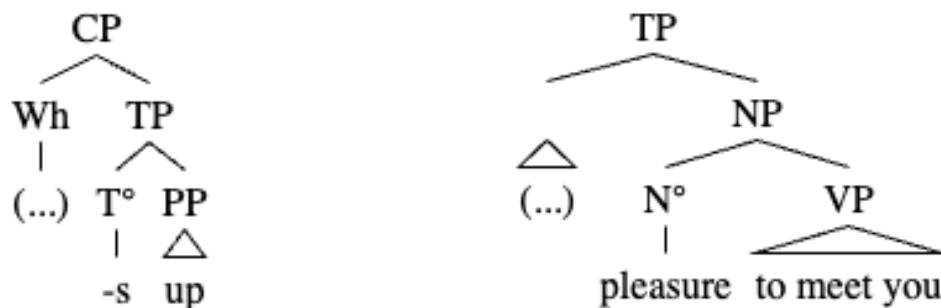
Figure 5.   **Figure 6.**

Elliptical representation of [sup]   Elliptical representation of [Pleasure to meet you]

The elliptical framework is attractive, especially considering that formulaic fragmentation seems be systematic and perhaps sensitive to underlying features: [what's up] ~ *[what's]; [it's a pleasure to meet you] ~ *[meet you]. However, it is not supported by empirical research on processing formulaic speech. As mentioned, formulaic speech is consistently observed to be preserved in individuals with "significant loss to generative language competence" (Sidtis 2012, p. 68). Ellipsis is a syntactic mechanism that requires generative procedures such as substitution and coreference. Therefore, while this representation may provide a potential derivation of these forms prior to acquisition by the speaker, it is not viably account for how speech formulas are acquired and produced.

Formulaic fragmentation does, however, resemble the second of Blanc's six stages of natural language acquisition: mitigation. In the echolalic literature, mitigation refers to the manipulation of lexicalized chunks using decomposition and recombination (Blanc et al. 2023; Stiegler 2015). To expand on this point, I reference Sidtis's (2012) characterization formulaic speech as "overlearned expressions" (p. 64), which highlights the high frequency of circulation these expressions have in the speech community. The speech formula [*what's up*], for example, is frequently encoutered by many speakers of English as a consistent gestalt, always appearing in a specific pattern and specific pragmatic context. It follows, then, that the processor lexicalizes this gestalt as a formuleme, a symbolic unit represented holistically and independently of the lexical items comprised within. The phonological form gives the illusion of an underlying syntactic structure but is a singular association between a sound pattern and its corresponding meaning.

Following this framework, fragmented alloforms of speech formulas may be acquired as such due to frequency rather than derived in real time using syntactic operations.

<u>Borrowing</u>

Another core feature of speech formulas is that they are recognizable to members of the speech community as features of the language. In the UK, for example, a common formulaic greeting is "Are you okay?" or "Are you alright?" As a speaker of United States English, I did not retrieve the conventionalized meaning "hello" when first encountering this form and instead interpreted the question computationally to arrive at a quite literal interpretation which might have prompted an unexpectedly in-depth response. I have, however, acquired speech formulas from speech communities I do have membership in, specifically Puerto Rican Spanish. Before learning Spanish in middle school, my frequent exposure to the language in my environment facilitated my acquisition of specific rituals —[*Bendición*] '(Please give me your) Blessing'; [*Dios te bendiga*] 'God bless you'— and expletives —[*Ten cuida'o!*] 'Be careful!'. Crucially, I understood the meaning of these items and could produce them appropriately and with native-like accuracy (because they had been acquired as such) without any formal understancing of Spanish syntax. I include this annecdotal evidence (which I invite future research to formalize with empirical study) to highlight that formulaic speech is not necessarily generated by syntax during production; rather, it is retrieved holistically from the Mental Lexicon. The role of syntax within this framework is more explicitly outlined in the following section.

<u>Summary</u>

In summary, formulaic speech is a natural language phenomenon that evidentiates the perseverating mechanism of gestalt processing that fuels lexicalization. Linguistic meaning is

constantly being renegotiated according to the input a speaker's recieves as well as the output they produce. Speech formulas exemplify this phenomenon by undergoing a process of re-lexicalization that overrides their compositional structure. Speakers with unimpaired generative competence may still mitigate and assign underlying structure to holistic representations that resemble compositional phrases (as will be explored in the following section). Speakers whose generative competence is compromised may still produce the fossilized representations they acquire before incuring the unfortunate trauma.

Idioms

Idioms are another subcategory of fixed expressions, and thus they may be classified as a type of formulaic speech (Sidtis 2012; Eaton & Burrows 2022). Idioms are the poster child of the Parallel Architecture as they demonstrate the asymmetrical relationship between phonological, syntactic, and semantic representation (see Figure 7).
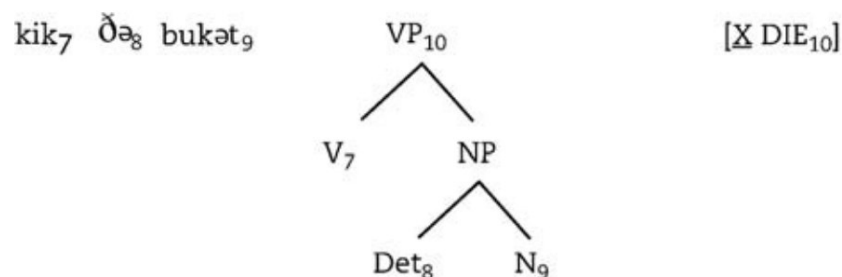

Figure 7. [kick the bucket] in the Parallel Architecture (Jackendoff 2007, p. 11)

Jackendoff (1997) hightlights that many idioms are resistent to passive structure, and he includes the example "#The bucket was kicked by John" in a list of examples that are

"impossible in an idiomatic reading (signaled by #)" (p. 166). I want to emphasize the use of # to mark semantic infelicity instead of syntactic ungrammaticalilty. According to Jackendoff's analysis, the infelicity arises from the fact that the semantic representation is not compositional; that is, there is no semantic representation that co-indexes with Det_8 or N_9 or NP, so those constituents cannot be separated from the phrase. Impossible, however, is perhaps hyperbolic. Many speakers intentionally misappropriate conventionalized phrases using syntactic transformations like passivisation for the purpose of humor or other discourse styles. Consider an instance where speaker A says to speaker B, "I thought John was still alive." To which speaker B responds, "Are you kidding? NP[That bucket]$_i$ was VP[kicked _$_i$] a long time ago!" In this case, the semantics flexibility is perhaps more salient. Perhaps the perceived inconsiderate prosody of this particular construction given its morbid implication has something to do with its infrequency. Nevertheless, we can conceptualize of NP[the bucket] as corresponding to a conceptual representation of John's life, and the V∘[kick] as corresponding to a separate conceptual representation akin to [X LOST].

Regardless of whether the passivized construction of this particular is semantically felicitous or maintains an idiomatic reading, I interpret the ability to manipulate the syntax without ungrammaticality to be indicative of the mechanism I call "grammaticization". This term is adopted from Nunberg et al.'s (1994) assertion that attempts to explain idiomaticity "have tended to overgrammaticize the phenomena—to ask the syntactic or semantic apparatus of the grammar to explain regularities that are in fact the consequences of independent rhetorical and discursive functions of the expressions" (p. 5). While the specific argument Nunberg makes is outside the scope of this essay, the notion of "overgrammaticize"-ation is not trivial and exceedingly relevant to the issue of holistic versus compositinoal representations of fixed

expressions. As these authors highlight, metalinguistic awareness plays a central role in linguists'

—and speakers'— conceptualization of language and its underlying structure. Crucially,

metalinguistic awareness of grammatical conventions and their applications is inherently

idiosyncratic. Zhao (2016) writes "grammar is usage-based and language knowledge comes from

language use, therefore, grammatical rules are not innate" (p. 375). Yet this is generally

unaccounted for by existing theoretical frameworks.

The proposal that accompanies my use of the term "grammaticization" is a formalism of

Jackendoff's (2007) theory of "syntactic integration: the building of a unified syntactic structure

from the fragments now present in working memory" (p. 14). In an earlier work detailing

theories about the lexical interface. Among other points, Jackendoff argues that some lexical

items have no syntactic structure at all: "*tra-la-la, e-i-e-i-o,* and *ink-a-dink-a-doo*" (Jackendoff

1994, p. 94). I take this point a step further to argue that syntactic structure is not innate to any

lexical item. When a phrasal utterance is encoded as a lexical gestalt, the compositional structure

is inherent in its phonological form. Consider, for instance, the occasion that [kick the bucket]

enters working memory through retrieval for production. The lexicalized gestalt is analyzed by

the processor during planning, and pattern recognition allows it to be parsed into three lexical

items (presumably matching items previously acquired): [kick] [the] [bucket]. The processor

then employs knowledge of grammatical conventions to assign a syntactic representation:

$_{VP}$[kick $_{NP}$[the bucket]].

Crucially, this construal is motivated by the need to integrate the idiom with other

generative structures, like TP, in real time. A speaker who generates the above structure may be

able to perform V°-to-T° movement to adapt the phrase to the intended tense: He $_{TP}$[kick$_i$-ed $_{VP}$[

$_i$ $_{NP}$[the bucket]]]. Another speaker may have lexicalized the preterite form [kicked the bucket]

and may employ a faithfullness constraint that disallows the omission of the preterite tense marker /-d/, making future tenses idiomatically infelicitous. Yet another speaker may grammaticize [kick-the-bucket] as a compound verb with a complex v°, allowing the /-d/ morpheme to undergo infixation. Lastly, this framework may also apply to formelemes like [*what's up*], which may be parsed as [*What is up!*] to allow for prosodic varation for stylistic purposes.

Summary

In summary, idioms demonstrate properties that suggest storage as holistic items in the Mental Lexicon. Crucially, however, the rigidity of syntactic form may be mitigated with the support of comositional semantics, as articulated above. This strongly suggests that the syntactic structure of the fixed expression may still be generated during retrieval without real-time composition. Grammticization is, thus, the act of performing syntactic integration on lexical gestalts. This framework implements language processing models to account for idiosyncracy in the application of grammatical conventions, specifically to fixed expressions. Most importantly, it maintains the position that idioms and other fixed expressions are stored holistically in the Mental Lexicon without flattening their syntactic structure during retrieval.

Conclusion

In conclusion, fixed expressions reveal the dynamic interface of compositionality and holistic representation that drives the language faculty. The theory of grammaticization which I propose and outline here accounts for the observed linguistic and cognitive contrast between novel and formulaic speech. Fixed expressions—which belong to the latter classification—move

bidirectionally along this continuum: towards holitic representation in the Mental Lexicon during acquisition, which may be gradual or relatively immediate; and towards compositional representation in working memory during retrieval.

This framework has important implications for linguists' understanding of natural language and its development. Future research should further investigate the role of grammaticization in communicative disorders, particularly ones that involve difficulty with composition at the syntactic or morphological level. Linguists should also continue to investigate formulaic speech and its relationship to social processing and other "non-linguistic" mechanisms indirectly related to language production. Finally, future scholarship should further investigate the role of memory in the acquisition, storage, and application of grammatical conventions and other metalinguistic features (as opposed to strictly lexicalization of meaning).

References

Blanc, M., Blackwell, A., & Elias, P. (2023). Using the Natural Language Acquisition Protocol to Support Gestalt Language Development. *Perspectives of the ASHA Special Interest Groups*, *8*(6), 1279–1286. https://doi.org/10.1044/2023_PERSP-23-00098

Jackendoff, R. (1997). *The Architecture of the Language Faculty*.

Jackendoff, R. (2007). A Parallel Architecture perspective on language processing. *Brain Research*, *1146*, 2–22. https://doi.org/10.1016/j.brainres.2006.08.111

Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. *Language*, *70*(3), 491–538. https://doi.org/10.1353/lan.1994.0007

Piñango, M. M. (2024). *What Experimentation Reveals About Linguistic Meaning and its Cognitive Structure*.

Stiegler, L. N. (2015). Examining the Echolalia Literature: Where Do Speech-Language Pathologists Stand? *American Journal of Speech-Language Pathology*, *24*(4), 750–762. https://doi.org/10.1044/2015_AJSLP-14-0166

Torrington Eaton, C., & Burrowes, L. (2022). Comparing patterns of familiar language use across spontaneous speech contexts in individuals with nonfluent aphasia and healthy controls. *Aphasiology*, *36*(12), 1399–1418. https://doi.org/10.1080/02687038.2021.1966375

Van Lancker Sidtis, D. (2012). Formulaic Language and Language Disorders. *Annual Review of Applied Linguistics*, *32*, 62–80. https://doi.org/10.1017/S0267190512000104

Zhao, G. (2016). A Cognitive Approach to Ellipsis. *Theory and Practice in Language Studies*, *6*(2), 372. https://doi.org/10.17507/tpls.0602.20