Usage-based constructionist approaches and Large Language Models

Adele E. Goldberg

Princeton University

Abstract

The constructionist framework is more relevant than ever, due to efforts by a broad range of researchers across the globe, a steady increase in the use of corpus and experimental methods among linguists, consistent findings from laboratory phonology and sociolinguistics, and striking advances in transformer-based large language models. These advances promise exciting developments and a great deal more clarity over the next decade. The constructionist approach rests on two interrelated but distinguishable tenets: a recognition that constructions pair form with function at varying levels of specificity and abstraction, and the recognition that our knowledge and use of language are dynamic and usage based.

Keywords: constructionist approach; ConstructionNet, usage based, large language models

1. Introduction

I use the term *constructionist approach* to emphasize two claims (Goldberg 2006).[1] First, language is comprised of a dynamic network of CONSTRUCTIONS, at varying levels of complexity and abstraction, which pair each form with a conventional range of functions. Equally important, languages are learned or CONSTRUCTED on the basis of the linguistic input witnessed, together with general cognitive, pragmatic, and processing factors. These and several other basic tenets of the constructionist approach are stated below:

(1) All levels of description are understood to involve form-function pairings, including filled and partially filled words (aka morphemes); filled and partially filled idioms, and partially lexically filled and fully abstract schematic grammatical patterns.

(2) Constructions are understood to be learned on the basis of the input and general cognitive mechanisms and are expected to vary cross linguistically.

(3) Cross-linguistic generalizations are explained by the functions of the constructions and general cognitive constraints.

---

[1] For thoughtful descriptions of the broad landscape of related frameworks please see Gonzálvez-García & Butler (2006); Hoffman & Trousdale (2013); Ungerer & Hartmann (2023).

(4) An emphasis is placed on subtle aspects of the way we construe events and states of affairs. The functions of constructions include any conventional aspect of meaning, information structure, rhetorical influence, register or speaker attitude.

(5) A "what you see is what you get" approach to syntactic form is adopted.

(6) Language-specific generalizations among constructions are captured via dynamic networks.

(7) The totality of our knowledge of language is captured by a dynamic network of constructions: a CONSTRUCTIONNET. [2]

1.1. Constructions all the way down

While some constructionists distinguish words from complex constructions, or fully specified collocations and idioms from partially schematic constructions, there is value to using a single term, *constructions* (or *signs*), because of the profound parallels between all types of linguistic units (see also Diessel et al. 2019; Fillmore et al. 1988; Goldberg 1995). Each can be lexically filled, partially filled, or fully abstract, and each can be compositional to variable degrees.

Each construction serves a function, or more typically, a range of related functions. That is, both words and grammatical constructions tend to convey a conventional range of polysemous and occasionally homonymous meanings (e.g., Lakoff 1987; Langacker 1987; Goldberg 1995). There should be little controversy about this point since there would be no reason to access and articulate a construction that had no impact on the comprehender. When researchers occasionally argue for a functionless construction, the evidence typically rests on the existence of a form that is associated with a range of arguably unrelated functions. For instance, Jackendoff (2002) suggests the English verb-particle construction is a syntactic pattern that serves no generalizable function. Yet he does not argue that any particular verb-particle combination is functionless. The observation is instead that forms tend to be reused for multiple purposes (Croft 2001: 132-34; Culicover & Jackendoff 2005: 42; MacDonald 2013). This observation holds of traditional lexical items as well as grammatical constructions, likely due to efficiency considerations (MacDonald 2013; Piantadosi et al. 2012). However, just as few would suggest that a word is meaningless just because it can be used to convey multiple functions, neither is there reason to posit meaningless grammatical patterns. In this way, the recognition of parallels between traditional lexical items and grammatical constructions (recall tenet [1]) reveals as a non-issue, the question as to whether constructions exist that have no function. Example constructions at varying levels of complexity and abstraction are offered in Table 1.

Table 1. English constructions at varying levels of complexity and abstraction. Lexically filled aspects are in italics

Constructions	English Examples (can be lexically filled, partially filled, fully abstract)

---

[2] I suggest the term *ConstructionNet* as a replacement for the term *constructicon since* the latter is prone to being auto corrected or misinterpreted.

Words	*predate, going, saunter, afraid, walk, walked*
Words with open slots (MORPHEMES)	*pre*-N, V-*ing*, *V-ed*
Constructions that convey states, events or relationships between multiple states or events (w/ more specific instances in italics)	Verb + clausal complement *believe* [clause]; *know* [clause] Double object construction *gimme* Obj$_2$; *tell* Obj$_{1rec}$ Obj$_{2;}$ Gossip construction *It's nice of you to be here; It was stupid of me.* *the* Xer, *the* Yer construction: *The bigger they come, the harder they fall*
Constructions with functions related to structuring discourse (w/ lexically specified instances in italics)	Information questions *What does that mean*? Polarity questions *Does it matter? Is that a thing?* Relative clauses *things you can do* Passives *Mistakes were made*
Idioms, collocations	*happily ever after* *raise the roof*
Idioms/collocations with open slots	*hazard* <a guess> *I hope this* < message> *finds you well*

1.2. Definition

My understanding of constructions has evolved as I've gained a better appreciation of human memory and learning. Rather than abstract constructions being reified entities that exist independently of their instantiations, the description in (8) is more accurate:

(8) A construction is an emergent cluster of lossy (imperfect) memory traces that align within our high-dimensional conceptual space on the basis of shared form, function, and contextual dimensions (Goldberg 2019:7).

The definition of construction in (8) is based on evidence of the usage-based nature of our knowledge of language, briefly reviewed in the following section.

2. The usage-based nature of language

Each language can be viewed as a complex solution the challenge of communicating about a combination of familiar and new ideas, in familiar and new contexts, with familiar and new individuals. For this reason, our knowledge of language is incredibly rich and complex (Beckner et al. 2009; Arnon & Christiansen 2017; Diessel et al. 2019; Herbst 2011; Hunston & Francis 2000; Huddleston & Pullum, 2002; Langacker 1988; Wray 2013). One of the aims of usage-based constructionist approaches is to understand how language users are able to learn, represent, combine and employ constructions appropriately. The usage-based approach demands we recognize our vast associative memory and the importance of social and communicative factors that give rise to the distributional patterns evident within each community of language users (see Croft this issue; van Trijp this issue).

Not every constructionist emphasizes the usage-based nature of language. Indeed, this aspect was only in my own peripheral vision early on (e.g., Goldberg 1995: 135). A far deeper appreciation of statistical information and discourse factors came into focus by the time I wrote *Constructions at Work* (Goldberg 2006), as I interacted with and read more work by colleagues such as Ron Langacker, Joan Bybee, Liz Bates, Wallace Chafe, Jeff Elman, Knud Lambrecht, Mike Tomasello, the other authors and editors of this issue, and others.

Yet the usage-based nature of language is tacitly endorsed by nearly every psychologist and machine learning expert, in addition to those of us who explicitly describe our perspective as usage-based (Abbot-Smith & Tomasello 2010; Ambridge & Lieven 2011; Arnon & Snider 2010; Boas 2008; Diessel & Hilpert 2016; Dunn 2019; Kidd et al. 2010; Hilpert 2015; Ibbotson 2022). The frequencies of constructions and the frequencies of their subparts simultaneously influence language processing and language change (Baayen & Prado Martin 2005; Bybee 2010; Gries 2010; Goldberg & Lee 2021; Gries & Hilpert 2010; Traugott & Trousdale 2013). Relationships among constructions and the forms of constructions are shaped by users' goals and conversational demands over diachronic time (e.g., Francis & Michaelis 2017; DuBois 2014; Givón 2014).

Since new information is related to old information, constructional generalizations emerge as clusters of related instances within the high-dimensional network embedded in each brain, with its nearly 100 billion neurons and roughly 60 trillion connections. As discussed at some length in Goldberg (2019), memory traces that cluster together to form constructions involve partially overlapping patterns of connections. Our brain's incredibly rich network is dynamic: each person's ConstructionNet is shaped by millions of exposures to language (Beckner et al. 2009; Bybee 2010; McClelland et al. 2010; Gries & Hilpert 2008; Perek 2015; Traugott 2014). ConstructionNets continue to change as speakers are exposed to new contexts, new semi-idiomatic expressions (*Ok, boomer*; *living one's best life; I did a thing)*, new individuals, different dialects, and/or new languages. Several foundational aspects of memory and learning are relevant to the dynamic nature of ConstructionNets including the following (Goldberg 2019: 6):

- Speakers balance the need to be expressive and efficient while conforming to the conventions of their speech communities.

- Our memory is vast but imperfect: memory traces are partially abstract ("lossy").[3]

- Lossy memories are aligned when they share relevant aspects of form and function, resulting in overlapping, emergent clusters of representations: constructions.

- New information is related to old information, resulting in a rich network of constructions.

- During production, multiple constructions are activated. If they cannot combine, they compete with one another to express our intended messages.

- During comprehension, mismatches between what is expected and what is witnessed fine-tune our network of learned constructions via error-driven learning.

---

[3] Representations are "lossy," a term from computer science, in the sense that they are not fully specified in all detail. Models are also lossy representations, a point we return to in section 4.1.

The usage-based perspective allows constructional knowledge to be both remarkably specific and flexible. A cluster of lossy overlapping memory traces that comprise a construction will be more specific, the narrower the range of contexts it is witnessed used in. That is, when witnessed utterances share similar contexts of use, the learned cluster will be correspondingly narrow or specific. When witnessed instances are more variable, the construction will be applied to new cases more broadly (Suttle & Goldberg 2011; Perek & Goldberg forthcoming). Yet even highly specific constructions are extended flexibly on occasion, because speakers need to use constructions in ever-changing contexts to convey an open-ended range of messages (Casasanto & Lupyan 2011; Christensen & Chater 2022; Christianson & Ferreira 2005; Christianson 2016; Christianson & Ferreira 2005; Cuneo et al. 2024; Ferreira et al. 2002; Goldberg & Ferreira 2022; Rambelli et al. 2022). The specificity and flexibility of language can most easily be illustrated by example, which brings us to the section 2.1.

2.1. Hazard a guess

The phrase *hazard a guess* occurs only 187 times in the billion-word corpus of Contemporary American English, COCA (Davies 2008), but the transitional probability of *a guess* following *hazard* used as a verb is very high: P(*a guess | hazard*) = .49, in the same corpus. That is, roughly half of the time *hazard* is used as a verb, it appears in the phrase *to hazard a guess*. Speakers display an implicit awareness of this type of highly specific information. When 20 English-dominant adults were recruited on Prolific crowd-sourcing platform and asked to provide the next word in the following sentence, roughly half of them supplied the indefinite phrase, *a guess*:

(9) *I can't try to hazard _____*

At the same time, the familiar phrase *hazard a guess* is partially compositional, with the word *guess* being particularly transparent.[4] Because *guess* is meaningful, it may appear in the plural, with modification, and/or as the topic in a passive, as illustrated in the attested examples (10)-(13), respectively (also from COCA).

(10) *If I had to hazard guesses.*

(11) *[they] can hazard a pretty sophisticated guess.*

(12) *I am going to hazard a wild guess.*

(13) *one can only now hazard an educated guess.*

---

[4] The verb *to hazard* is semantically related to its more common use as a noun. When one hazards a guess, there is typically little evidence on which to base the guess, which makes the guessing somewhat fraught or hazardous. Occasionally, *to hazard* is used to imply some communicative signal other than a guess. In this case, it retains the implication that the action makes the speaker vulnerable as in the attested examples from COCA:
    a. *I hazarded that maybe it was glamorous living in exile with a tennis legend.*
    b. *She hazarded a backward glance.*

Notably, because the statistics are so skewed toward the particular word, *guess,* the meaning "guess" tends to be implied even when other terms are used, as in the attested example in (14), which makes the fact that the prediction is a guess explicit in the second clause:

(14) *I would hesitate to hazard a particular percentage, but I would guess that…*

This brings us to an implication of the fact that the input for most constructions tends to be skewed, as discussed in Section 2.1.

2.2.   Skewed input

Constructions, like many other distributions, are commonly skewed, often in what looks like a quasi-Zipfian way (Piantadosi 2014). When this occurs, the construction appears to become implicitly associated with the meaning of its most frequent instance (Goldberg 2006; Goldberg et al. 2004; but cf. Perek 2016 for a different case). For example, the verb *give* is the grammatical "head" of roughly half of all tokens of the English double object construction, and its meaning – transfer from one animate being to another – is associated with the construction even when other verbs are used in the construction. This offers an explanation for why *She baked him something* entails that she intended to give him whatever it was she baked.[5] Similarly, the verb *make* accounts for roughly 40% of the instances of the *way* construction (e.g., *She made her way into the room*), and the construction implies that a real or metaphorical path is created (i.e., made); the implication that a path is created is what imbues the construction with its interpretation of self-propelled motion despite difficulty or obstacles (Goldberg 1995).

3.   The usage-based nature of language is a challenge for symbolic formalisms

While all constructionists recognize Chuck Fillmore as an inspirational figure, my own thinking has been at least as influenced by George Lakoff, my PhD advisor at Berkeley. In the first course I took in the department, Lakoff shared chapters of his then-new book, *Women, Fire and Dangerous Things* (1987). The title was intended to trick readers into assuming that he was claiming that women, fire, and dangerous things all share something in common. The book dispels that misconception by explaining that linguistic elements in languages are rarely tamed by a list of features or by a definition. Instead, most grammatical elements, word roots, and constructions are polysemous in that they can convey a range of distinguishable *but related* interpretations. This provides another parallel between words and abstract grammatical constructions. I've found the observation that strict definitions regularly fail to account for familiar concepts to be one of the most profound I've encountered, and I do my best to keep it forefront in my mind. It is the root cause of my skepticism regarding symbolic formalisms.

---

[5] The paraphrase, *She baked something for him,* on the other hand can alternatively be interpreted to mean that she baked something on his behalf (to be given to someone else), or that she baked something intending to throw it at him.

3.1. Symbolic, feature-based formalisms

Each language provides an exquisitely finely tuned formalism for communication. No simplification of a natural language into interpretable symbols will ever equal natural languages themselves as a means of expressively and efficiently conveying an open-ended range of messages. What better way to express the rich meaning involved in words like *extradite, renege, fireworks* than with the lexical items themselves? An early valiant attempt to define "drink" by one eminent linguist led to the cockeyed "CAUSE LIQUID to MOVE into one's MOUTH." This representation incorrectly predicts that gargling is a type of drinking, and fails to distinguish sipping, gulping, and chugging. The meaning of LIQUID also begs for further scrutiny, as glass and sand are sometimes classified as liquids by physicists, yet swallowing a mouthful of sand or glass is not considered drinking. And if we dare venture beyond LIQUID, we face the need to distinguish GIN, VODKA, and ORANGE CRUSH in some way other than simply promoting English terms into capital letters. Interpretable features are especially impotent when used to represent lexical semantics (Fillmore 1975). It is futile to decompose the meaning of words such *Bible-belt, tailgate, gaslight, TikTok, contention, soccer* or idioms such as *Ok Boomer; I feel seen*; *cast the first stone.*

Grammatical constructions likewise offer effective and succinct means of conveying who did what to whom, as well as distinguishing questions from statements, commands from requests. English has a special "gossip construction" that evaluates how the actions of sentient beings reflect on them (e.g., *It's big of them to be there*, Goldberg & Herbst 2021). Grammatical constructions commonly constrain how parts of an utterance relate to other utterances that came before or will come afterwards, the certainty of the information being conveyed, or the relative status of speaker and hearer, to name just a few of the many functions assigned by constructions.

Although I sometimes use simple decompositional representations to capture the type of general and abstract meanings associated with common argument structure meanings in English (e.g., CAUSE-MOVE), I do this only in an effort to provide a representation that may be comprehended at a glance. This is possible for highly frequent constructions because they generalize across many, varied instances, so their meanings are necessarily very general. I also sometimes employ grammatical terms such as N(noun) or V(verb), but this again only intended to provide information via a shorthand that I assume is familiar to readers. Each time I employ a formal notation, I am unsatisfied and humbled. Syntactic terms such as *noun, subject, passive* do not refer to consistent categories across different languages (e.g., Croft 2001; Fried 1994; LaPolla 1993), nor even within a single language (e.g., Croft 2001; Culicover 1999; Goldberg 2006; Ross 1973). Efforts to collapse the complex and massive expressive power that each language boasts into a fixed set of features will fall short of capturing the communicative potential of natural languages, but doing so may be of practical value in certain circumstances.

Other contributors to this issue have developed formalisms that suit their intended goals. Proponents of Sign-Based Construction Grammar provide a unification-based symbolic formalism for the sake of explicitness and in order to offer a common descriptive language (Michaelis this issue; Trousdale & Michaelis 2010; Boas & Sag 2012; Bergen & Chang 2005). As Michaelis (this issue) describes it, the use of a "rigid" formalism can provide a "way of seeing." Likewise, Fluid Construction Grammar offers a fully implemented computational tool that can be used to test the compatibility of representations in a way that captures the interactive online nature of language processing as a means of communication (Steels & de Beule 2006; van Trijp 2014, this issue). In this way, Fluid Construction Grammar has made enormous efforts and taken great strides in grounding meaning in the goals of agentive actors in real and computational situations.

3.2.   Combining constructions: an example

The editors asked each of us to analyze the sentence in (15) in terms of a combination of constructions. I confess to not being confident that I understand what it is intended to mean, due to the use of what for me is a positive polarity item, *rather,* in the negative context (Israel 2001) and my lack of familiarity with golf. Therefore, below I analyze the sentence in (16) instead.

(15) *Wasn't it rather McIlroy who seemed never to be outdriven when playing in contention?*

(16) *Wasn't it actually Everett who consistently demonstrated remarkable linguistic skills, effortlessly speaking multiple languages?*

What is most interesting about the utterance in (16) (and presumably the one in [15]) is its complex interpretation. Due to the combination of constructions it employs (see i-viii below), (16) conveys a hedged assertion, namely that the speaker believes the polyglot at issue is Everett; it also presupposes that someone had incorrectly suggested (or thought) that a different person was a remarkable polyglot. Example (16) combines the following constructions:

(i)   A *yes/no* question construction (polar interrogative), which includes the SAI construction (a):

   a. Subj-aux inversion (SAI) construction (*wasn't it*), which includes:

      i.   A negative clitic (*n't*), which presupposes the relevance of the positive counterpart (e.g., Horn 1989; Lakoff 2014).

   Because polar interrogatives literally question the veracity of the associated assertion, they can be used to imply that the associated assertion is false. In (16) the associated assertion is negated, so that the question implies that Everett *was* the polyglot.

(ii)   The *it*-cleft construction (*It BE ___ <relative clause>)* includes b:

   b. a relative clause construction, in this case a subject-oriented non-restrictive relative clause (since *Everett* is interpreted the subject argument of the relative clause).

   The *it*-cleft presupposes the content of the relative clause and makes the head noun a contrastive focus. In (16), *Everett* is the contrasted focus, and it is presupposed that someone else was suggested to be the polyglot.

(iii)   The focus element (*actually*) emphasizes the contrastive interpretation of the *it*-cleft in (16). It treats *Everett* as the focus and implicitly corrects a mistaken belief (whether previously asserted or presupposed), in this case that someone other than Everett was the polyglot.

(iv)   Several instances of the noun phrase construction (*it, Everett, remarkable linguistic skills, multiple languages*).

(v)   A verb phrase adjunct which is discontinuous from what it modifies (*effortlessly speaking multiple languages* modifies *Everett*, not *skills*) [aka a "dangling participle"].

(vi)   Two lexical adverbs modifying different verb phrases (*consistently, effortlessly*).

(vii) Morphological inflection constructions (*V-ing, N-s, V-past*).

(viii) Lexical items, each associated with related words as well as their own range of functions: *was, not, it, actually, who, consistently, demonstrated, remarkable, linguistic, skills, effortlessly, speaking, multiple, languages*

The list in (i)-(viii) clarifies that the utterance in (16) is a combination of many constructions, although the descriptions in (i)-(viii) fail to do justice to any of them, as dozens of papers have been written on each. Moreover, listing constructions obscures the fact that each exists as part of a network of related items in the ConstructionNet: the polar interrogative construction in (i) is related to the tag question construction (e.g., *wasn't it?*) and to information *(wh-)* questions. The subject-auxiliary construction is in reality a family of constructions in English (e.g., Goldberg 2006). The *it*-cleft construction is related to the presentational relative clause construction (e.g., *there was a guy who*) and to *wh*-clefts (Kim & Michaelis 2020); it can also be used as a scene-setting device, with information structure quite distinct from that described in (ii): e.g., *It was 1967 when young people from around the world were drawn to San Francisco by the promise of peace, love, and understanding* [COCA, 1997, SPOK]. Subject-oriented relative clauses are related to non-subject oriented relative clauses. And obviously, the lexical item *was* is related to *were, be*, and *is;* the adverb *effortlessly* is related to the lexemes *effortful, effort,* and to the morphological constructions *N-less, Adj-ly*. In fact, each "construction" is a family of constructional interpretations (Barðdal et al. 2011; Desagulier 2016; Goldberg & van der Auwera 2012; Goldberg & Herbst 2021; Goldberg & Jackendoff 2004; Goldberg & Michaelis 2017; Gonzálvez-García 2009; Kapatsinski & Vakareliyska 2013; Jong- Kim and Sells 2013; Lakoff 1987; Ungerer 2022).[6]

Moreover, none of the labels or features used in (i)-(viii) captures their usage-based nature. There are no uniform tests that hold of all words we generally call adjectives in English (Goldberg 2019), let alone any tests that apply to all adjectives in all languages (Croft 2001). While there are constructions with comparable functions across languages (Croft 2022) and while each construction is *motivated* and not random, the specifics of each construction are not strictly predictable (e.g., Lambrecht 1994). The need to explain the usage-based motivation and complexity of every feature and construction leaves me wary of symbolic formalisms.

4. A Game Changer: Large Language Models

---

6 This is not to say that all of the content we infer from an utterance is specified in any particular construction or by their combination directly. As Remi van Trijp (this issue) emphasizes, each utterance provides only cues that enable people to create an enriched situation model (or mental model, see Johnson-Laird 1983; Christiansen & Chater 2022).

Since the functions of constructions are essential for communication, I had been privately skeptical that Large Language Models (LLM), which are trained only on text, would ever produce or comprehend language in a way that might be mistaken for a human being (the famous Turing Test; French 2000). But then a new generation of models burst onto the scene at the end of November in 2022, beginning with ChatGPT, and this has led me to reverse my perspective. I am far from alone: the prominent New York Times newspaper swiftly published half a dozen essays including "This changes everything" (Klein 2023), and "Our New Promethean Moment" (Freidman 2023). As is shown in Figure 1, during the four months following the release of ChatGPT and then GPT4, which followed quickly on its heels, searches for *GPT* in Google quickly shot up to meet the average daily searches for *Excel,* software used daily all over the world, by millions (Figure 1).
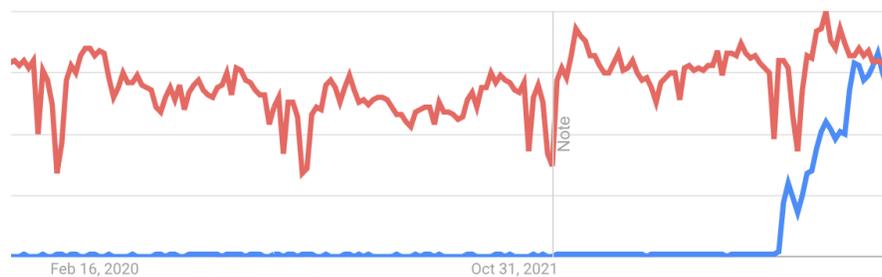
Figure 1. Google Trends data from February 2020 to May 2023

ChatGPT and GPT4 are generative pre-trained transformer models, a type of Large Language Models (LLMs). *Generative* here simply means that the models generate novel outputs; it is unrelated to generative linguistics, and in fact, generative linguists are generally skeptical, arguably naively so (e.g., Chomsky et al. 2023). In fact, as described below, LLMs share far more with the usage-based constructionist approaches to language than traditional generative approaches (Weissweiler et al. 2023). Table 2 provides six striking parallels, with the final one, new to ChatGPT and GPT4, potentially being quite profound: the new models are specifically trained to be helpful to human users (Section 4.6). There are to be sure, important differences in how the parallels arise. Each is discussed briefly in turn below.

Table 2. Parallels between the usage-based constructionist perspective and LLMs

PARALLELS	Usage-based constructionist approach to language	GPTChat, GPT4 & similar recent LLMs
LOSSY COMPRESSION and INTERPOLATION	Human brains represent the world imperfectly (lossy), with limited resources (compressed); we generalize from familiar to related cases (via interpolation/ coverage/ induction)	Every model involves lossy compression and interpolation; all neural net models interpolate to generalize within the range of the training data.
CONFORM TO CONVENTIONS	Humans display an inclination to conform to the conventions of their communities.	Pre-training to predict the next word in texts requires that outputs conform to conventions in the input.
COMPLEX DYNAMIC NETWORK	Structured distributed representations at varying levels of complexity and abstract are learned from input and an understanding of others' intentions and real-world grounding; they can be flexibly extended.	Structured distributed representations at varying levels of complexity and abstraction are learned from massive amounts of input text; they can be flexibly extended.
CONTEXT-DEPENDENT INTERPRETATIONS	Humans use linguistic *and non-linguistic* context for interpretation.	Only linguistic context is available, via a thousand words of preceding text.
RELATIONSHIPS AMONG DISCONTINUOUS ELEMENTS	Made possible via working memory and attention.	Made possible by attention heads (Transformer models)
GOAL is to BE HELPFUL	Humans display natural tendency to be helpful to others in their communities.	Special training provided by InstructGPT taught GPTs to provide responses humans find helpful.

4.1. Lossy compression and interpolation

In an effort to demystify the then-new ChatGPT model, an essay in the *New Yorker* magazine (Chiang 2023) argued that it was (simply) analogous to a blurry image of the web, in that it involved the standard mechanisms of lossy compression and interpolation. Yet *every* model involves lossy compression, since no model is a veridical replication of reality, since models don't have infinite recourses. And interpolation is what allows models to "fill in" missing information by averaging neighboring vectors when missing information falls within the training space.

The usage-based constructionist approach likewise recognizes that the human brain involves lossy compression and interpolation insofar as memories are imperfect (lossy), and brain resources are vast but finite at any given time (requiring compression). Humans also readily generalize from familiar cases to similar new cases (interpolation or *coverage*, Suttle & Goldberg 2011; Goldberg 2019).

But lossy compression and interpolation cannot on their own explain the sudden and impressive leap the new generation of models have taken, since the past five decades of neural net (connectionist) models have employed lossy compression and interpolation (e.g., McClelland & Rumelhart 1986a, 1986b). While the earlier models enjoyed successes within limited domains, they never approached the stunning performance of current GPT models.

4.2. Conform to conventions

The "pre-training" involved in current LLM models (i.e., the P in GPT) has been used in some form or other for decades: models are trained to predict the next word in large amounts of coherent texts of natural language. Initial predictions are random, but the models learn to iteratively self-correct based on receiving the next word that actually appears in the text. As in certain earlier neural network models, each word is divided into substrings, which allows the models to learn morphology from regularities in written text. Importantly, this pre-training regime forces the models to conform to human input, which ensures that they produce the conventions reflected in that input as best they can. We have also known at least since Elman (1990) that models trained in this way learn hierarchical structure, which essentially groups strings of text into coherent units insofar as single words can substitute for the string in other texts (Langacker 1997). Finally, the next-word prediction task allows for the emergence of constructions with open but constrained slots: an increase in entropy in which word will appear next indicates an open slot, and the distribution of potential next-words serves to constrain the type of filler that may appear in that slot (Dunn 2022).

The same key attributes are recognized to be critical in the usage-based constructionist approach: the inclination to conform to others is required for a community to converge on a shared system; the tendency to try to construe meaningful units from continuous pieces, and of course the possibility for constructions to include open but constrained slots. Humans spontaneously display an inclination to conform to other members of their community in comparison to other apes. For instance, both preschool-aged children and chimps are able to learn to perform multi-step processes in order to receive a reward, but only children persist in conforming to the same process when an easier solution is evident. In these contexts, children recognize that there is a conventional or "correct" way to perform the activity and they conform their behavior accordingly (Gergely, Bekkering, & Király 2002; Horner & Whiten, 2005). Humans also naturally segment the natural world into meaningful units in vision, memory, and language (Chater 2018). We construe parts of scene that move together as parts of the same entity (Ostrovsky 2009), we come to recognize parts with relatively high transitional probabilities as units (Saffran et al. 1996), and we understand contiguous words that combine to form a coherent unit to be a semantic constituent.

Finally, there is evidence that humans spontaneously predict the next word that will be uttered while comprehending language (Kutas & Federmeier 2011). For instance, the N400 ERP component, detectable from EEG recordings on the scalp, while people listen to text, correlates quite well with how predictable each word is in context (e.g., Nieuwland & Van Berkum 2006). Less predictable words result in a higher amplitude N400 and highly predictable words results in a negligible N400. Yet, despite the efficacy of predict-the-next-word training, it is unlikely to be sufficient to explain the dramatic improvement in recent LLM models, because it has been used for decades.

4.3. Complex dynamic network of constructions at varying levels of abstraction and complexity

Today's LLMs include far more layers, with exponentially more nodes and connections than earlier connectionist models. This accounts for their characterization as "deep learning" models (Graves et al 2013). And ChatGPT was trained on 300 billion words of text scraped from the internet in all languages found online. The massive amount of input allowed it to learn the thousands of collocations, idioms, and semi-idiosyncratic constructions within the vast training data, a key hallmark of usage-based constructionist approaches. The compression involved requires a rich network of conventional constructions to partially share representational structure with related constructions, in a spirit similar to the clustering described in section 2.

To be clear, ChatGPT and GPT4 receive orders of magnitude more input than any human could absorb in a hundred lifetimes. And no human can learn any new language only by scanning text or by listening to the radio, even for a billion years. We would be unable to glean any meaning whatsoever. Humans, however, have access to real or imagined grounding in various contexts, and importantly, humans are adept at inferring the intentions of others (Tomasello 2003, 2010). I had been skeptical that models trained only on text could converse coherently with humans, but ChatGPT has proven my intuition wrong.

4.4. Context-dependent interpretation

In order for natural language to be an efficient and flexible form of communication, it provides only cues to the intended message, rather than specifying the message in its entirety (Christiansen & Chater 2022). For instance, we are far more likely to say *Hey* in greeting than *I, the speaker appearing before you, hereby acknowledge and greet you informally with this utterance that I expect you can hear and interpret as intended*. Context and background knowledge supplement the cues that utterances provide to allow us to understand others' intended messages (Goldberg 2015). If a young child asks for a drink, we understand they are likely to want water, juice, or milk, while if an adult at a bar ask for a drink, it is more likely to be a Manhattan or a Mojito. Today's GPT models have no access to non-linguistic contexts, but they make use of the large amount of linguistic context: each token of input includes thousands of words of the preceding text.

4.5. Semantic relationships among discontinuous elements

The *T* in *GPT* stands for *transformer*. The essential innovation in transformer models is that they include "attention heads", which allow each part of the input to be weighted differently in different layers (Vig 2019). Attention heads enable discontinuous relationships of all kinds to be captured, therefore serving a role similar to working memory and attention in humans, albeit a far more powerful one. The attention heads allow GPT models to produce and respond coherently to all sorts of long-distance dependencies including those within sentences (*if/then, wh*-questions, etc.), and those that hold across sentences, paragraphs and pages of text.

4.6. NEW: Goal is to be helpful

While early generative pre-trained transformer models (GPT-1, GPT-2) were impressive in many ways (e.g., Hawkins et al. 2020; Dasgupta et al. 2022; Grand et al. 2022; Mahowald 2023), they were prone to grammatical errors and regularly produced totally incoherent sequences of text. This was perhaps not surprising, since humans do not generate language with the goal of producing the most likely next word, and as noted, human learners are able to glean others' intentions in context (e.g., Tomasello, 2003). Our goal in producing and comprehending language is to convey messages and respond appropriately to others.

I find it remarkable that what seems to have enabled the new generation of GPT models to engage in human-like conversations was additional training that taught them to align with human conversational goals. That is, ChatGPT and GPT4 were trained to be helpful, or more specifically: helpful, true, non-toxic, respectful, and humble about their own certainty. This special training came from a separate "InstructGPT" model,

created on the basis of human rankings of sets of machine-generated responses according to how "helpful, honest and harmless" the responses were (Ouyang et al. 2022). The InstructGPT model then used reinforcement learning based on the human feedback (RLHF) (Christiano et al. 2017). This helpfulness-training resulted in ChatGPT performing better, with a mere 1.3 billion parameters, than the same model that had *175* billion parameters but no helpfulness training (Ouyang et al. 2022).

Grice's (1975) essential point was that human language users are cooperative: we are, and assume others are: relevant, truthful, brief, and mannerful. The goal of being helpful is mighty close to being cooperative: one would seem to need to be relevant, generally honest, allow for turn-taking (one way to interpret the idea of being "brief"), and be appropriate in context (or "mannerful"). In fact, the assumption that others are being helpful or cooperative is a prerequisite for natural language, present in young children (Tomasello 2010). On the other hand, a great deal of experimental and observational data has found that non-human primates rarely if ever interpret communicative signals as intended to be helpful (Tomasello 2009, 2016). Chimps fail to understand an experimenter's pointing gesture to where food is hidden, even while they understand grabbing gestures, in which an experimenter reaches for food as though competing for the resource (Call et al. 2000; Herrmann & Tomasello 2006).

What I find most intriguing about the latest LLMs is the way they succeed as well as they do. The assumption that others intend to communicate cooperatively and the inclination to conform to relatively arbitrary social norms are prerequisites for languages (and complex cultures) to emerge in humans. This combination of prerequisites explains why none of our primate cousins are able to learn a language anywhere near as complex as the natural language of humans (Tomasello 2010, 2019).

4.7.   With great power comes great responsibility

Many have rightly observed that there is a dark side to successful language models that ought not be ignored (e.g., Bender et al. 2021; Klein 2003; van Dis et al. 2023). Their statistical nature makes them prone to exaggerating biases in their input and incorporating falsehoods in their responses (recall the lossy compression). The first of the new generation of models, ChatGPT, was accurate perhaps 80% of the time, potentially enough to lull users into accepting its "botsplanations" at face value. The models can easily be used to generate propaganda and disinformation. There will surely be anticipated and unanticipated consequences of the new technology that individuals and societies need to think through carefully. Yet legitimate concerns about current and future societal impacts do not imply we should bury our heads in the sand and pretend that the latest models do not produce and comprehend languages that they have had sufficient exposure to. They evidently do.

The speed of innovation itself is remarkable. I am keenly aware anything I write will likely be outdated before this paper is published. I encourage any readers who have not yet had a chance to play with the latest models to find the time to do so.

5. GPTs at work

In what follows, I include a series of representative responses to prompts I provided to GPT4 in March of 2023. Like humans, GPT models are not deterministic. What follows includes responses to the first (and

only) time I provided a prompt. Your results will vary. While it is remarkably impressive overall, an illustrative instance in which it fails in an illuminating way is included as well (Figure 9).

5.1.   Intention-reading and social inferences

Inspired the idea that pointing is unique to humans among all apes (Tomasello 2010), I prompted GPT4 with the following: "If I'm walking with a friend and I point to a bicycle parked by a house and wink, what do you think I might mean?". Its response was coherent and quite human-like (Figure 2).

When you point to a bicycle parked by a house and wink at your friend, it could mean a variety of things depending on the context and your relationship with your friend. Some possible interpretations might include:

1. You're suggesting the bicycle would be a fun or convenient mode of transportation.
2. The bicycle might be a type or brand that you or your friend like or have talked about before.
3. The bicycle may remind you of a shared experience, joke, or memory related to cycling or a specific event.
4. You could be playfully implying something mischievous or joking about taking the bike.

It's essential to consider the context and your shared experiences with your friend to understand the intended meaning fully.

Figure 2. GPT4 displaying an ability to interpret the description of a pointing gesture

In another test of GPT4's ability to make appropriate social inferences, I asked: "When Tim's husband said he was at the gym all morning, Tim turned red. What might have happened?". GPT4's response was again remarkably appropriate (Figure 3).

When Tim's husband said he was at the gym all morning, and Tim turned red, it could indicate a variety of emotions or reactions. Some possibilities include:

1. Surprise: Tim might be surprised to hear that his husband spent the whole morning at the gym, perhaps because it's unusual behavior or because he had other expectations for their plans that day.
2. Embarrassment: Tim could be embarrassed for some reason, such as if he had planned a surprise or if he thought his husband was somewhere else and had told others about it.
3. Anger or suspicion: Tim might be upset or suspicious about his husband's statement, perhaps suspecting that he was not actually at the gym or that he is hiding something.
4. Physical reaction: Tim's turning red could also be unrelated to his husband's statement and instead be due to a physical reaction, such as blushing, overheating, or experiencing an allergic reaction.

The exact reason for Tim's reaction depends on the specific context and their relationship dynamics.

Figure 3. A simple probe that elicited a range of plausible inferences

I tested the model on whether it could supply reasonable and distinct inferences when given single-word utterances, *fire!* vs. *coffee!*. Chat-4's unedited and appropriate responses are provided in Figure 4 (graciously overlooking a typo):

Figure 4. GPT4's markedly distinct and highly appropriate interpretations of the single word utterances, *fire!* and *coffee!*

5.2. GPT correctly interprets unusual examples

Curious how GPT4 would respond to instances in which constructional meaning should coerce the interpretation of the utterance, I asked it: "what does 'she sneezed the foam off the cappuccino' mean?" and then: "what does '3 computers ago' mean?". While these "novel" inputs may be contained in GPT4's vast training data, it is doubtful that the intended interpretation of either was detailed, because human readers do not require the meanings to be spelled out. Remarkably, GPT4 responded with appropriate and detailed descriptions of both interpretations (Figure 5):

The phrase "She sneezed the foam off the cappuccino" is a figurative and humorous way of saying that someone sneezed with such force or intensity that it seemed like it could blow the foam off a cappuccino. It is not meant to be taken literally but instead paints a vivid picture in the listener's mind of the strength or suddenness of the sneeze. This expression can be used to add humor or color to a description of an event or situation involving a sneeze.

"3 computers ago" is an informal way of referring to a point in time when someone had three computers prior to their current one. It is used to give a rough idea of how much time has passed or how much technology has advanced since then. For instance, if someone says, "I remember using Windows XP 3 computers ago," they are likely indicating that they have had three newer computers since the one running Windows XP, suggesting that it was a relatively long time ago.

Figure 5. GPT4's interpretation of *she sneezed the foam off the cappuccino* and *three computers ago* which require constructional meaning

Another example of appropriately interpreting novel input comes from GPT4's interpretation of a novel Phrase-as-Lemma (PAL) construction (Shirtz & Goldberg forthcoming), as illustrated in examples (17)-(20), from the COCA corpus (Davies 2008):

(17) *A don't-mess-with-me driver*

(18) *It's not a "call Ronan Farrow" scenario*

(19) *We're at the people-are-moving-to-Jersey stage of nationwide collapse*

(20) *This is my "can you believe this bull***t?" face.*

Using corpus analysis, survey data, and cross-linguistic comparison, we provide motivation for the form and function of phrases that are treated syntactically as if they were words. In particular, we argue that novel uses of the PAL construction are ideally suited for conveying what comedians call "observational humor." The reasoning for this is sketched in Table 3.

Table 3. Why the Phrase-As-Lemma construction has the interpretation it has

Claim	Basis for the claim
(a) The construction treats a phrase formally as if it were a word root.	(e.g., Trips and Kornfilt 2017)
(b) The concept associated with a lexeme is what psycholinguists refer to as a *lemma*.	Definition
(c) Lemmas evoke familiar, recurrent semantic frames.	(Fillmore 1985; Geeraerts 2006)
(d) PALs are therefore understood to convey a type of event or situation that the speaker expects the listener to find familiar.	(a)-(c)
(e) Observation humor involves talking about familiar events that are not usually talked about.	Definition

(f) Novel PALs express events, presumed to be familiar, that are not often talked about, convey observational humor.	(d)-(e)

We confirmed the hypothesized function of the PAL construction with survey data that asked participants to compare pairs of sentences that either included a PAL or a non-PAL paraphrase (Shirtz & Goldberg forthcoming). Results showed that the sentences with PALs implied more shared background between speaker and listener and were judged to be more witty and more sarcastic than non-PALs. With this as background, I asked GPT4 what the following means: "I'm officially 'slows down at all of the yellow traffic lights' years old". Remarkably, GPT4 recognized the "humorous" flourish of the PAL construction (Figure 6) and interpreted the phrase accurately:
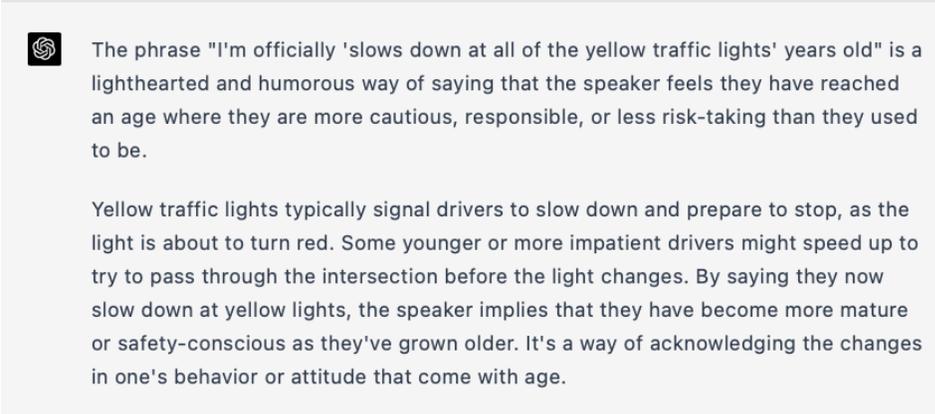
The phrase "I'm officially 'slows down at all of the yellow traffic lights' years old" is a lighthearted and humorous way of saying that the speaker feels they have reached an age where they are more cautious, responsible, or less risk-taking than they used to be.

Yellow traffic lights typically signal drivers to slow down and prepare to stop, as the light is about to turn red. Some younger or more impatient drivers might speed up to try to pass through the intersection before the light changes. By saying they now slow down at yellow lights, the speaker implies that they have become more mature or safety-conscious as they've grown older. It's a way of acknowledging the changes in one's behavior or attitude that come with age.

Figure 6. Probe of the PAL construction (see Shirtz & Goldberg forthcoming)

5.3. GPT4 on a simple math problem

GPT4's ability to solve at least simple arithmetic word problems can be impressive, as illustrated in Figure 7. The response combines arithmetic and world knowledge when asked: "If 40 people can fit on a bus, how many buses are needed to drive 84 people 6 blocks?" (I included *6 blocks* in the prompt in an unsuccessful attempt to misdirect the model).
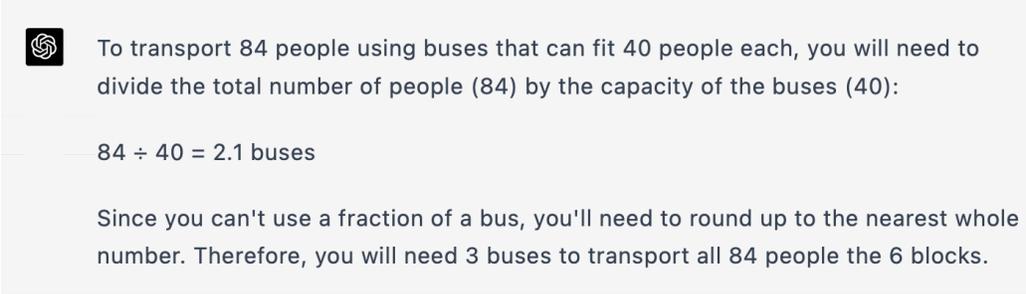
To transport 84 people using buses that can fit 40 people each, you will need to divide the total number of people (84) by the capacity of the buses (40):

84 ÷ 40 = 2.1 buses

Since you can't use a fraction of a bus, you'll need to round up to the nearest whole number. Therefore, you will need 3 buses to transport all 84 people the 6 blocks.

Figure 7. GPT4 response to a mathematical word problem

5.4. GPT4 appropriately characterizes conceptual metaphors

Can GPT4 make sense of conceptual metaphors? An initial prompt resulted in the response in the left panel of Figure 8. When then asked to "tell me like I'm in first grade", the simplified description at the top right of Figure 8 was given; notice it includes a novel yet sensible metaphorical extension: *going up, up, up into the sky.* The final response comes from a prompt to generate a novel metaphor and once again, GPT4's response is competent (Figure 8).
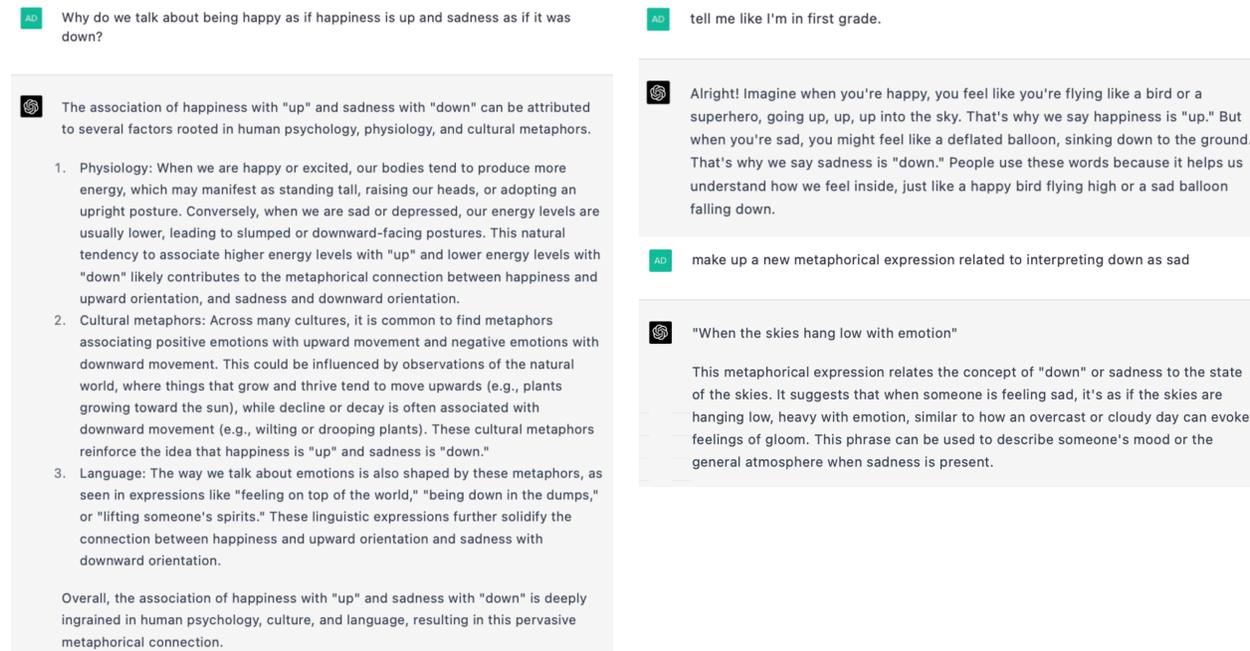

Figure 8. GPT4 on conceptual metaphors

5.5. Over-reliance on associations can lead GPT models (and humans) astray

Insight into how ChatGPT worked comes from a series of examples posted on Twitter by @PaulMainwood (February 22, 2023).[7] In one, Mainwood cleverly provided ChatGPT with a twist on a well-known riddle, written by undergraduate students at Boston University in 2008, and intended to highlight implicit sexism. A version of the familiar riddle follows in (21):

(21) Familiar riddle (included in training data): "A boy was rushed to the hospital after a terrible car crash in which his father was killed. The surgeon looks at the boy and says 'I can't operate: he's my son'. How is this possible?"

People who hear the riddle for the first time are sometimes flummoxed until it is revealed that the surgeon is the boy's mother. Mainwood explains a different situation to ChatGPT: it is not a riddle at all but is strongly but vaguely reminiscent of the original riddle. He stated that the man at the wheel was the child's biological father and that the surgeon is the child's adoptive father. Strikingly undeterred, ChatGPT blindly

---

forged ahead and provided the standard answer to the standard riddle, incongruously responding, "The surgeon is the boy's mother". I tried the same prompt on GPT4, and it performed similarly, confidently but incorrectly stating that the surgeon was the boy's "adopted mother" (Figure 9).

Figure 9. GPT4's response to a prompt inspired by @PaulMainwood, illustrating its over-reliance on context

GPT4's response indicates that it associates the prompt with a specific context – the standard riddle – which was undoubtably encountered in its training. The type of error in Figure 9 is potentially interesting. We humans are also prone to context-based errors, which have been described as a result of "good-enough" processing (Christianson 2016; Ferreira et al. 2002; Goldberg & Ferreira 2022). For example, when asked: "How many pairs of animals did Moses take on the ark?", people commonly fail to notice that the query asked about Moses, rather than Noah.[8] Similarly, students are often misled by math and physics word problems that vary from the specific types of content they had been previously exposed to (Bassok 1990). In fact, when I shared Figure 9 with two quite brilliant colleagues, each made the same error that ChatGPT and GPT4 did, by failing to notice that the prompt was *not* the standard riddle.

In the narrow domain of human natural language production and comprehension, GPT models make every phrase structure grammar and every syntactic parser that came before look like line drawings next to Gaudí's *Sagrada Familia*. These models are lacking in spatial reasoning, and in complex math and logic problems; and of course they lack knowledge of world events outside their exposure. Like *la Sagrada Familia*, the models are works in progress. Advancements will continue. It is up to humans to put the models to work in ways that benefit humanity. And it will be left to cognitive scientists and linguists to explore how they work.

---

[8] Unsurprisingly, since the example is a class example of good-enough processing, GPT4 was *not* fooled by this particular question, responding "Moses did not bring animals onto an ark; it was Noah who brought animals onto the ark according to the biblical story found in the Book of Genesis..."

6. Looking ahead

I fully agree with Martin Hilpert (this issue) that the future of construction grammar is in excellent hands. Researchers ought to follow their own curiosity wherever it takes them. But before closing, I offer what I personally take to be the most promising directions for new work over the coming decade.

- GPT models put on full display the power of usage-based constructionist models. Systematic investigation of such models will likely help us better understand parallels and differences with human language and cognition (Hawkins et al. 2020; Mahowald forthcoming; McCoy et al. 2021; Piantadosi 2023). At the very least, since we know that context always matters, we need to move away from static representations of the ConstructionNet and embrace dynamic models to the extent possible (see also Barak & Goldberg 2017; Dasgupta et al. 2022; van Trijp 2014, 2015; Steels & de Beule 2006).

- Fieldwork will always be highly valuable and large-scale cross-linguistic comparisons will help us better understand shared aspects of our semantic and pragmatic construal of the world, as well as the processing pressures that result in languages patterning as they do (Bohnemeyer et al. 2007; Croft 2001, 2022; Haspelmath 2010; Kemmerer 2011; Majid et al. 2004).

- A fuller, deeper appreciation of information structure and lexical semantics can unlock puzzles that have long been assumed to require syntactic stipulations, including island constraints, scope, anaphora, and binding (Ackerman & Nikolaeva 2014; Cole et al. 2014; Culicover & Jackendoff 2005; Cuneo & Goldberg 2023; Francis & Michaelis 2017; Goldberg & Michaelis 2017; Israel 2001; Namboodiripad et al. 2022).

- Laboratory phonology and sociolinguistics are thriving subfields of linguistics. Each field has long provided compelling evidence for the usage-based approach to language. Researchers equipped to hypothesize and test potential parallels between phonological and grammatical phenomena will be in a position to offer coherent and insightful perspective across subareas (e.g., Bybee 2010; Docherty & Foulkes 2014; Harmon & Kapatsinski 2017).

- We ought not feel constrained to focus only on traditional questions. For instance, emotion drives most everything we do, so it is worthwhile to better understand its role in communication and language (e.g., Citron & Goldberg 2014; Foolen 2012). We also need to incorporate implications of communicative gestures (Congdon et al. 2018; Khasbage et al. 2022; Steen & Turner 2013; Willems and Hagoort 2007) and conversational dynamics (Du Bois et al. 2003; Hopper & Thompson 1980; Stephens et al. 2010) to more fully understand natural languages.

- Applications of the constructionist approach to education, atypical language development (e.g., Goldberg & Abbot-Smith 2021), and language documentation (e.g., Bast et al. 2023) are among the most exciting new directions for constructionists to develop.

7. Conclusion

Let's allow ChatGPT to have the final word, offered in the style of Ovid (Left, Figure 10) and Dr. Seuss (Right, Figure 10).
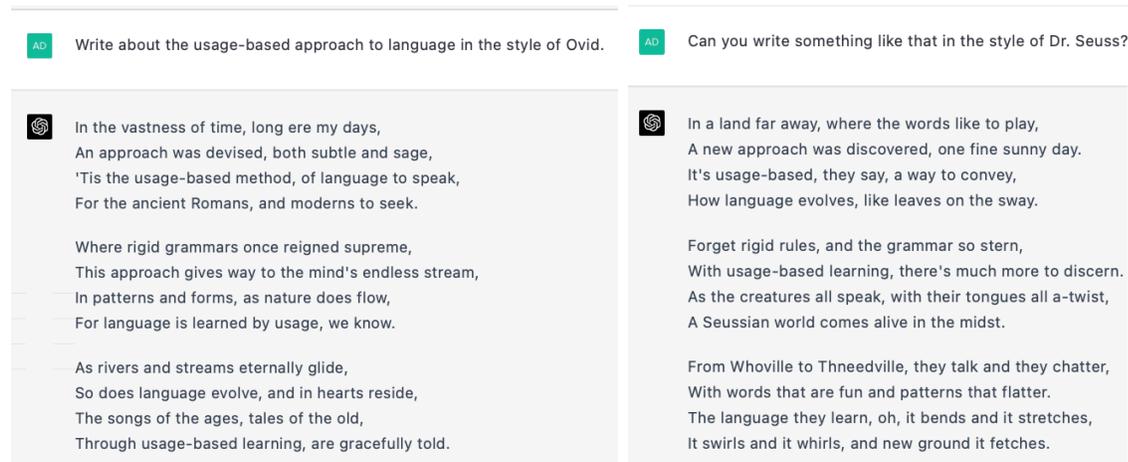

Figure 9. Final remarks, generated by GPT4 (March 2023)

References

Abbot-Smith, K., & Tomasello, M. (2010). The influence of frequency and semantic similarity on how children learn grammar. *First Language*, *30*(1), 79–101. https://doi.org/10.1177/0142723709350525.

Ackerman, F., & Nikolaeva, I. (2014). *Descriptive typology and linguistic theory: A study in the morphosyntax of relative clauses*. CSLI Publication.

Ambridge, B., & Lieven, E. V. M. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1 L2 differences. *Topics in Cognitive Science*, *9*(3), 621–636.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Baayen, R. H., & del Prado Martin, F. M. (2005). Semantic density and past-tense formation in three Germanic languages. *Language*, *81*(3), 666–98. https://doi.org/10.1353/lan.2005.0112.

Barak, L., & Goldberg, A. E. (2017). Modeling the partial productivity of constructions. In *The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding*, [Technical report SS-17-02] (pp. 131–138). AAAI Press.

Barðdal, J., Kristoffersen, K. E., & Sveen, A. (2011). West Scandinavian ditransitives as a family of constructions: With a special attention to the Norwegian 'V-REFL-NP'construction. *Linguistics, 49*(1), 53–104. https://doi.org/10.1515/ling.2011.002.

Bassok, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 522–533. https://doi.org/10.1037/0278-7393.16.3.522.

Beckner, C., Ellis, N. C., Blythe, R., Holland, J., Bybee, J., Christiansen, M. H., Larsen-Freeman, D., Croft, W., & Schoenemann, T. (2009). Language is a complex adaptive system. *Language Learning*, *59*(1), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🦜. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Bergen, B., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitivegrounding and theoretical dimensions* (pp 147 – 190). John Benjamins. https://doi.org/10.1075/cal.3.08ber

Boas, H. C. (2008). Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics*, *6*, 113–144. https://doi.org/10.1075/arcl.6.06boa

Boas, H., & Sag, I. (Eds.). (2012). *Sign-Based Construction Grammar*. CSLI Publications.

Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretke, I., Kita, S., Lupke, F., & Ameka, F. (2007). Principles of event segmentation in language. *Language, 83*(3), 495–532.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.

Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review, 22*(2–4), 381–410.

Call, J., Agnetta, B., & Tomasello, M. (2000). Cues that chimpanzees do and do not use to find hidden objects. *Animal Cognition*, *3*(1), 23–34

Casasanto, D., & Lupyan, G. (2011). Ad hoc cognition. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society (*p. 826). Cognitive Science Society.

Chater, N. (2018). *Mind is flat: The remarkable shallowness of the improvising brain*. Yale University Press.

Chiang, T. (2023, February 9). ChatGPT Is a Blurry JPEG of the Web. *The New Yorker*.

Chomsky, N., Roberts, I., & Watumull, J (2023, March 8). Noam Chomsky: The False Promise of ChatGPT. *The New York Times*.

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Curran Associates Inc.

Christiansen, M. H., & Chater, N. (2022). *The language game: How improvisation created language and changed the world*. Hachette UK.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good enough, underspecified, or shallow language processing. *Quarterly Journal of Experimental Psychology, 69*(5), 817–828. https://doi.org/10.1080/17470218.2015.1134603.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition 98*(2), 105–135. https://doi.org/10.1016/j.cognition.2004.10.006.

Citron, F. M. M., & Goldberg, A. E. (2014). Metaphorical sentences are more emotionally engaging than their literal counterparts. *Journal of Cognitive Neuroscience, 26*(11), 2585–2595. https://doi.org/10.1162/jocn_a_00654.

Cole, P., Hermon, G., & Yanti (2014). The grammar of binding in the languages of the world: Innate or learned? *Cognition, 141*, 138–60. https://doi.org/10.1016/j.cognition.2015.04.005.

Congdon, E. L., Novack, M. A., Brooks, N., Hemani-Lopez, N., O'Keefe, L., & Goldin Meadow, S. (2018). Better together: Simultaneous presentation of speech and gesture in math instruction supports generalization and retention. *Learning and Instruction, 50*, 65–74. https://doi.org/10.1016/j.learninstruc.2017.03.005.

Croft, W. (2001). *Radical Construction Grammar*. Oxford University Press.

Croft, W. (2022). *Morphosyntax: Constructions of the world's languages*. Cambridge University Press.

Croft, W. (this issue). Philosophical reflections on the future of construction grammar (or, Confessions of a radical construction grammarian). *Constructions and Frames*, *16*(2), XX-YYY.

Culicover, P. W. (1999). Syntactic nuts: Hard cases, syntactic theory and language acquisition. *Cognitive Linguistics*, *10*(3), 251–261.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press.

Cuneo, N., & Goldberg, A. E. (2023). The discourse functions of grammatical constructions explain an enduring syntactic puzzle. *Cognition*, *240*, 105563. doi.org/10.1016/j.cognition.2023.105563

Cuneo, N., Floyd, S., & Goldberg, A. E. (2024). Word meaning is complex: Language-related generalization differences in autistic adults. *Cognition*, *244*, 105691. https://doi.org/10.1016/j.cognition.2023.105691

Dasgupta, I., Lampinen, A. K., Chan, S. C. Y., Creswell, A., Kumaran, D., McClelland, J. L., & Hill, F. (2022). Language models show human-like content effects on reasoning. arXiv. http://arxiv.org/abs/2207.07051.

Desagulier, G. (2016). A lesson from associative learning: Asymmetry and productivity in multiple-slot constructions. *Corpus linguistics and linguistic theory*, *12*(2), 173-219.

Davies, Mark (2008). The Corpus of Contemporary American English (COCA): One Billion Words, 1990-2019.

Diessel, H., Dabrowska, E, & Divjak, D. (2019). Usage-based construction grammar. *Cognitive Linguistics*, *2*, 50–80.

Diessel, H., & Hilpert, M. (2016). Frequency effects in grammar. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.120.

Docherty, G. J., & Foulkes, P. (2014). An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua, 142*, 42–56. https://doi.org/10.1016/j.lingua.2013.01.011.

Domanchin, M., & Guo, Y. (2017). New frontiers in interactive multimodal communication. In A. Georgakopoulou & T. Spilioti (Eds.), *The Routledge handbook of language and digital communication* (pp. 377–380). Routledge.

Du Bois, J. W. (2014). Towards a dialogic syntax. *Cognitive linguistics*, *25*(3), 359–410.

Du Bois, J. W., Kumpf, L. E., & Ashby, W. J. (2003). *Preferred argument structure: Grammar as architecture for function.* John Benjamins.

Dunn, J. (2019). Frequency vs. association for constraint selection in usage-based Construction Grammar. In E. Chersoni, C. Jacobs, A. Lenci, T. Linzen, L. Prévot & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics* (pp. 117–128). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-2913.

Dunn, J. (2022). *Natural language processing for corpus linguistics*. Cambridge University Press.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158.

Fillmore, C. J. (1975). An alternative to checklist theories of meaning. *The Annual Meeting of the Berkeley Linguistics Society*, *1*, 123–131.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni Di Semantica*, *6*(2), 222–253.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language, 64*, 501–538.

Foolen, A. (2012). The relevance of emotion for language and linguistics. In A. Foolen, U. M. Lüdtke, T. P. Racine & J. Zlatev (Eds.), *Moving ourselves, moving others: Motion and emotion in intersubjectivity, consciousness and language* (pp. 349–369). John Benjamins.

Francis, E., & Michaelis, L. (2017). When relative clause extraposition is the right choice, it's easier. *Language and Cognition, 9*, 332–70. https://doi.org/10.1017/langcog.2016.21.

French, R. M. (2000). The Turing test: The first 50 years. *Trends in Cognitive Sciences, 4*(3), 115–122. https://doi.org/10.1016/S1364-6613(00)01453-4.

Fried, M. (1994). Grammatical functions in case languages: Subjecthood in Czech. *The Annual Meeting of the Berkeley Linguistics Society*, *20*(1), 184–193.

Geeraerts, D. C. N. (2006). *Words and other wonders: Papers on lexical and semantic topics*. Mouton de Gruyter.

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755-755.

Givón, T. (1979). From discourse to syntax: Grammar as a processing strategy. In T. Givón (Ed.), *Discourse and syntax* (pp. 81–112). Brill.

Goldberg, A (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199268511.001.0001.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. The Chicago University Press.

Goldberg, A. E. (2015). Compositionality. In N. Riemer (Ed.), *The Routledge handbook of Semantics* (pp. 419–433). Routledge.

Goldberg, A. E. (2016). Subtle implicit language facts emerge from the functions of constructions. *Frontiers in Psychology, 6*, 1–11. https://doi.org/10.3389/fpsyg.2015.02019.

Goldberg, A. E. (2019). Explain me this: Creativity, competition, and the partial productivity of *constructions*. Princeton University Press.

Goldberg, A. E., & Abbot-Smith, K. (2021). The constructionist approach offers a useful lens on language learning in autistic individuals: Response to Kissine. *Language, 97*(3), e169–183. https://doi.org/10.1353/lan.2021.0035.

Goldberg, A. & van der Auwera, J. (2012). This is to count as a construction. *Folia Linguistica, 46*(1), 109–132. https://doi.org/10.1515/flin.2012.4

Goldberg, A. E., Casenhiser, D. M., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 14*(3), 289–316.

Goldberg, A. E., & Ferreira, F. (2022). Good-enough language production. *Trends in Cognitive Sciences, 26*(4), 300–311. https://doi.org/10.1016/j.tics.2022.01.005.

Goldberg, A. E., & Herbst, T. (2021). The nice-of-you construction and its fragments. *Linguistics, 59*(1), 285–318. https://doi.org/10.1515/ling-2020-0274.

Goldberg, A. E. & Jackendoff, R. (2004). The English resultative as a family of constructions. *Language, 80*(3), 532-568.

Goldberg, A. E., & Lee, C. (2021). Accessibility and historical change: An emergent cluster led uncles and aunts to become aunts and uncles. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.662884

Goldberg, A. E., & Michaelis, L. A. (2017). One among many: Anaphoric one and its relationship with numeral one. *Cognitive Science, 41*, 233–258. https://doi.org/10.1111/cogs.12339.

Gonzálvez-García, F. (2009). The family of object-related depictives in English and Spanish: Towards a usage-based constructionist analysis. *Language Sciences, 31*(5), 663–723. https://doi.org/10.1016/j.langsci.2008.01.003.

Gonzálvez-García, F., & Butler, C. S. (2006). Mapping functional-cognitive space. *Annual Review of Cognitive Linguistics, 4*, 39–96. https://doi.org/10.1075/arcl.4.04gon.

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, *6*(7), 975–987. https://doi.org/10.1038/s41562-022-01316-8.

Graves, A., Mohamed, A., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *arXiv*. http://arxiv.org/abs/1303.5778.

Gries, S. T. (2011). Phonological similarity in multi-word units. *Cognitive Linguistics*, *22*, 491–510.

Gries, S., & Hilpert, M. (2008). The identification of stages in diachronic data: Variability-based neighbour clustering. *Corpora, 3*, 59–81. https://doi.org/10.3366/E1749503208000075.

Gries, S. T., & Hilpert, M. (2010). Modeling diachronic change in the third person singular: A multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics, 14*(3), 293–320. https://doi.org/10.1017/S1360674310000092.

Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, *98*, 22-44.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language, 86*(3), 663–687. https://doi.org/10.1353/lan.2010.0021.

Hawkins, R. D., Yamakoshi, T., Griffiths, T. L., & Goldberg, A. E. (2020). Investigating representations of verb bias in neural language models. *arXiv.* http://arxiv.org/abs/2010.02375.

Herbst, T. (2011). The status of generalizations: Valency and argument structure constructions. *ZAA, 4*(4), 347–368.

Herrmann, E., & Tomasello, M. (2006). Apes' and children's understanding of cooperative and competitive motives in a communicative situation. *Developmental Science*, *9*(5), 518–529.

Hilpert, M. (2015). From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics, 26*(1), 113–147. https://doi.org/10.1515/cog-2014-0001.

Hilpert, M. (this issue). The road ahead for Construction Grammar. *Constructions and Frames*, *16*(2), XX-YYY.

Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language, 56*(2), 251–299.

Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal cognition*, *8*, 164–181.

Huddleston, R., & Pullum, G. (2005). The Cambridge grammar of the English language. *Zeitschrift für Anglistik und Amerikanistik*, *53*(2), 193–194.

Hunston, S., & Francis, G. (2000). *Pattern grammar. A corpus-driven approach to the lexical grammar of English*. John Benjamins.

Ibbotson, P. (2022). *Language acquisition: The basics*. Routledge.

Israel, M. (2001). Minimizers, maximizers and the rhetoric of scalar reasoning. *Journal of Semantics, 18*(4), 297–331. https://doi.org/10.1093/jos/18.4.297.

Jackendoff, R. (2002). English particle constructions, the lexicon, and the autonomy of syntax. In N. Dehé, R. Jackendoff, A. McIntyre & S. Urban (Eds.), *Verb-Particle Explorations* (pp. 67–94). Mouton de Gruyter.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Kapatsinski, V., & Vakareliyska, C. (2013). [N[N]] compounds in Russian: A growing family of constructions. *Constructions and Frames, 5*(1), 69–87. https://doi.org/10.1075/cf.5.1.03kap.

Kemmerer, D. (2011). The cross-linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's area. *Language and Linguistic Compass*, *6*(1), 1–17.

Khasbage, Y., Carrión, D. A., Hinnell, J., Robertson, F., Singla, K., Uhrig, P., & Turner, M. (2022). The red hen anonymizer and the red hen protocol for de-identifying audiovisual recordings. *Linguistics Vanguard*. https://doi.org/10.1515/lingvan-2022-0017.

Kidd, E., Lieven, E. V. M., & Tomasello, M. (2010). Lexical frequency and exemplar-based learning effects in language acquisition: Evidence from sentential complements. *Language Sciences, 32*(1). https://doi.org/10.1016/j.langsci.2009.05.002.

Kim, J-B., & Michaelis, L. A. (2020). *Syntactic constructions in English*. Cambridge University Press.

Kim, J., & Sells, P. (2013). The Korean sluicing: A family of constructions. *Studies in Generative Grammar*, *23*(1), 103–130.

Klein, E. (2023, March 12). This changes everything. *The New York Times*.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621–647.

Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. The Chicago University Press.

Lakoff, G. (2014). *The all new don't think of an elephant!: Know your values and frame the debate*. Chelsea Green Publishing.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge University Press.

Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics*. John Benjamins.

Langacker, R. W. (1987). Foundations of Cognitive Grammar: Theoretical prerequisites (Vol. 1). Stanford University Press.

Langacker, R. W. (1997). Constituency, Dependency, and Conceptual Grouping. *Cognitive Linguistics*, *8*, 1–32.

LaPolla, R. J. (1993). Arguments against 'subject' and 'direct object' as viable concepts in Chinese. *Bulletin of the Institute of History and Philology*, *63*, 759–813.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 1–16. https://doi.org/10.3389/fpsyg.2013.00226.

Mahowald, K. (2023). A discerning several thousand judgments: GPT-3 rates the article + adjective + numeral + noun construction. *arXiv preprint arXiv:2301.12564.*

Majid, A., Evans, N., Gaby, A., & Levinson, S. C. (2011). The semantics of reciprocal constructions across languages. In N. Evans, A. Gaby, S. C. Levinson & A. Majid (Eds.), *Reciprocals and semantic typology* (pp. 29–60). John Benjamins.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends in Cognitive Sciences, 14*(8), 348–356. https://doi.org/10.1016/j.tics.2010.06.002.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group (1986). *Parallel distributed processing* (Vol. 2). MIT Press.

McCoy, R. T., Smolensky, P., Linzen, T., Gao, J., & Celikyilmaz, A. (2021). How much do language models copy from their training data? Evaluating linguistic novelty in text generation using RAVEN. *arXiv.* http://arxiv.org/abs/2111.09509.

Michaelis, L. A. (2001). Exclamative constructions. In M. Haspelmath, E. König, W. Österreicher & W. Raible (Eds.), *Language universals and language typology: An international handbook* (pp. 1038–1050). Mouton de Gruyter.

Michaelis, L. A. (2010). Sign-based Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of Construction Grammar* (pp. 133–152). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195396683.013.0008.

Michaelis, L., & Francis, H. (2007). Lexical subjects and the conflation strategy. In N. Hedberg & R. Zacharski (Eds.), *The grammar pragmatics interface: Essays in honor of Jeanette K. Gundel* (pp. 19–48). John Benjamins. http://doi.org/10.1075/pbns.155.04mic.

Michaelis, L. (this issue). Staying terminologically rigid, conceptually open and socially cohesive: how to make room for the next generation of Construction Grammarians. *Constructions and Frames*, *16*(2), XX-YYYY.

Namboodiripad, S., Cuneo, N., Kramer, M. A., Sedarous, Y., Sugimoto, Y., Bisnath, F., &Goldberg, A. E. (2022). Backgroundedness predicts island status of non-finite adjuncts in English. *Proceedings of the Annual Meeting of the Cognitive Science Society, 28*, 347–355.

Nieuwland, M. S., & Van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Ostrovsky, Y., Meyers, E., Ganesh, S., Mathur, U., & Sinha, P. (2009). Visual parsing after recovery from blindness. *Psychological Science*, *20*(12), 1484–1491.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *arXiv*. http://arxiv.org/abs/2203.02155.

Perek, F. (2016). Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics, 54*(1), 149–188. https://doi.org/10.1515/ling-2015-0043

Perek, F., & Goldberg A. E. (Manuscript in preparation). Choosing the best available option: Productivity is context dependent. University of Birmingham.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, *21*, 1112–1130.

Piantadosi, S. T. (2023). Modern language models refute Chomsky's approach to language. *Lingbuzz*, *7180*.

Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*(3), 280–291. https://doi.org/10.1016/j.cognition.2011.10.004

Rambelli, G., Chersoni, E., Blache, P., & Lenci, A. (2022). Compositionality as an analogical process: Introducing ANNE. In M. Zock, E. Chersoni, Y. Hsu & E. Santus (Eds.), *Proceedings of the Workshop on Cognitive Aspects of the Lexicon* (pp. 78–96). Association for Computational Linguistics.

Roose, K. (2023, February 17). A conversation with Bing's chatbot left me deeply unsettled. *The New York Times*.

Ross, J. R. (1973). A fake NP squish. In C-J. N. Bailey & R. W. Shuy (Eds.), *New ways of analyzing variation in English* (pp. 96–140). Georgetown University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Schrimpf, M., Blank, I., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J., Fedorenko, E. (2021). Artificial neural networks accurately predict language processing in the brain. *Proceedings of the National Academy of Sciences, 118*(45), e2105646118. https://doi.org/10.1073/pnas.2105646118

Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424. doi:10.1017/S0140525X00005756

Shirtz, S., & Goldberg, A. E. (2024). The English Phrase-As-Lemma Construction **The English Phrase-as-Lemma Construction: When a phrase masquerades as a word, people play along**. Manuscript submitted for publication.

Silbert, L. J. (2013). *The underlying neural correlates of real-world communication* [Doctoral dissertation]. Princeton University. http://arks.princeton.edu/ark:/88435/dsp01jw827b806

Steels, L., & de Beule, J. (2006). A (very) brief introduction to fluid construction grammar. In J. Allen, J. Alexandersson, J. Feldman & R. Porzel (Eds.), *Proceedings of the Third Workshop on Scalable Natural Language Understanding* (pp. 73–80). Association for Computational Linguistics.

Steen, F. F., & Turner, M. (2013). Multimodal Construction Grammar. In M. Borkent, B. Dancygier & J. Hinnell (Eds.), *Language and the creative mind* (pp. 255–274). CSLI. https://doi.org/10.2139/ssrn.2168035.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences, 107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107.

Suttle, L., & Goldberg, A. (2011). The partial productivity of constructions as induction. *Linguistics*, *49*(6), 1237–1269.

Tomasello, M. (2010). *Origins of human communication*. MIT Press.

Traugott, E. C. (2014). Toward a constructional framework for research on language change. *Cognitive Linguistic Studies, 1*(1), 3–21. https://doi.org/10.1075/cogls.1.1.01tra.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.

Traugott, E. C., & Trousdale, G. (2013). *Constructionalization and constructional changes* (Vol. 6). Oxford University Press.

Trips, C., & Kornfilt, J. (Eds.) (2017). *Further investigations into the nature of phrasal compounding*. Language Science Press. https://doi.org/10.5281/ZENODO.885113.

Ungerer, T. (2022). Extending structural priming to test constructional relations: Some comments and suggestions. *Yearbook of the German Cognitive Linguistics Association*, *10*(1), 159–182.

Ungerer, T., & Hartmann, S. (2023). Constructionist approaches: Past, present, future. *PsyArXiv.* https://doi.org/10.31234/osf.io/83dvj.

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature, 614*(7947), 224–226. https://doi.org/10.1038/d41586-023-00288-7

van Trijp, R. (2014). Long-distance dependencies without filler−gaps: A cognitive-functional alternative in fluid construction grammar. *Language and Cognition, 6*(2), 1–29. https://doi.org/10.1017/langcog.2014.8

van Trijp, R. (2015). Towards bidirectional processing models of sign language: A constructional approach in fluid construction grammar. In G. Airenti, B. G. Bara & G. Sandini (Eds.), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science* (pp. 668–673). CEUR Workshop Proceedings.

van Trijp, R. (this issue). Nostalgia for the future of Construction Grammar. *Constructions and Frames*, *16*(1), XX-YYY.

Vig, J. (2019). A multiscale visualization of attention in the transformer model. *arXiv.* http://arxiv.org/abs/1906.05714.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, *11*(3), 271–294.

Weissweiler, L., He, T., Otani, N., Mortensen, D. R., Levin, L., & Schütze, H. (2023). Construction Grammar provides unique insight into neural language models. *arXiv. arXiv:2302.02178*

Willems, R. M., & Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. *Brain and Language, 101*(3), 278–289. https://doi.org/10.1016/j.bandl.2007.03.004.

Wray, A. (2013). Formulaic language. *Language Teaching, 46*(3), 316–334. https://doi.org/10.1017/S0261444813000013.

Address for correspondence

Adele Goldberg

Psychology Department

Perestman-Scully Hall, Princeton University

Princeton, NJ 08544

adele@princeton.edu