Stretched Tree Metric Grammars: A usage-based grammatical formalism that supports generation, parsing and morphological innovation

Whitney Tabor
University of Connecticut
whitney.tabor@uconn.edu

Hyosun Lee
University of Connecticut
hyosun.lee@uconn.edu

Abstract

Grammatical theories which specify grammars by means of symbolic well-formedness constraints (e.g., Context Free Grammars, HPSG, LFG, Minimalism, Dependency Grammars, etc.) are ill-suited to model the (semantically and statistically) gradual character of grammatical change as it manifests in successive historical corpora. Grammatical theories which claim that the language system is subject to change based on what speakers do in life (i.e., usage-based accounts) are better-suited to handle such phenomena. Nevertheless, current usage-based theories (e.g., Cognitive Grammar, Construction Grammar) lack a clearly formalized model that specifies how usage can affect the grammatical system. In this paper, we describe Stretched Tree Metric Grammars (STMGs), a new formal model of syntax and semantics that exhibits usage-based effects. We show that the model can generate and parse simple sentences. Then we show how it supports morphological innovation in appropriately limited circumstances. We conclude by noting that STMGs are closely related to Large Language Models (LLMs), but they have the benefit of being more analytically interpretable.

1 Introduction

Although the trajectories that languages follow as they evolve over historical time are difficult to predict in detail (Weinreich et al., 1968), much evidence indicates that grammar change is highly constrained by the nature of the language system at any point in time (Gottfurcht, 2008; Heine et al., 1991; Hopper, 1991; Hopper and Traugott, 2003; Lichtenberk, 1991; Roberts, 1993; Schmid, 2020). Saussure, concerned about the tendency that 19th century linguists exhibited to conflate historical and synchronic properties of language, argued for a focus on the synchronic (Saussure, 1916), and Chomsky, highlighting the importance of the language acquisition process accomplished by children early in life with no significant access to historical states of a language, argued for a focus on language as a property of individual brains, not on community trends (Chomsky, 1986). These emphases have been fruitful, but they do not help account for the historical patterning just mentioned.

By contrast, work in usage-based grammar (Beuls and Van Trijp, 2016; Bybee, 1985; Bybee et al., 1994; Langacker, 1987, 1988; Schmid, 2020; Steels, 2011) has long taken grammatical change phenomena as an important part of the data that a theory of language should capture. Working in the framework of Fluid Construction Grammar, Van Eecke and Beuls (2018) observe that languages appear to exhibit two kinds of "creative" behavior—the first type, which we label "generalization", refers to the use or comprehension, by a language user, of known constructions or rules in combinations which the user has not previously encountered. (unusual lexicalization fits this bill—e.g., "He loped down the hall chortling that he would revisit when the marmots faltered.") The second kind, which we label "innovation" refers to uses which go beyond the repertoire of existing types. (Morphological innovations like "unfriend" and grammaticalization developments like English "a lot" becoming a quantifier (Brems, 2003) fall under this heading.)

Focusing on innovation, one perspective proposes that language users have the capacity to diverge, in a single utterance, from the patterning exhibited by their current language system; on this view, such divergence occurs frequently in natural usage, but not all such divergences are embraced by the culture and developed into deep new structural properties of the grammatical system (only a few are). A related view holds that language users have the capacity to covertly re-analyze the structure of a grammatical utterance, thus establishing a different systematic interpretation of observed data from what previous generations assumed; only when

Proceedings of the 30th Conference on Computational Natural Language Learning, pages 760–770
July 3-4, 2026

that covert reanalysis is extended to new forms that could not be generated by the earlier grammatical system does the change become evident as evidence for learners (Langacker, 1977); the change will generally spread gradually from this point in time across the grammatical system (and to new users) because language users are, on the whole, cautious, or “sneaky" (De Smet, 2012) and thus eschew highly noticeable divergence at each moment.

We find many of these observations insightful and relevant, but we do not think they address a core question: How do the kinds of things that people do in their lives result in changes in a grammatical system? We call this the Life ↔ Grammar question.[1] To answer this question, we take a strong usage-based position: *every* instance of language use modifies grammar. Nevertheless, we suggest, in line with arguments of Petitot (2011), that the Life ↔ Grammar question has, thus far, not been very effectively answered because both formalist and usage-based linguists have been working with an inappropriate formal substrate.[2] Specifically, the rules of context-free grammars (Hopcroft and Ullman, 1979), the feature-structures of HPSG (Pollard and Sag, 1994), the graph structures of dependency grammars (**?**), the merge operation of Minimalism (Chomsky, 1995), the frames of Cognitive Grammar (Langacker, 1987), and the construction templates of Construction Grammar (Fillmore et al., 1988; Goldberg, 2006; Lakoff, 1977) are all discrete combinatory mechanisms. These devices are not irrelevant, but they inhabit the kind of space that mathematicians call a *discrete topology* where each structure is assigned a type, and the only possible distances between two structures are 0 (meaning, they have the same type) or 1 (meaning they are of different types). If we adopt, instead, a continuum topology with a real-valued metric (i.e. arbitrarily small distances between structures are possible), then we will be in a much better position to explain how grammatical systems and events of life can causally interact.

To argue for this position, we will show an example of a system of the sort we have in mind. Before doing this, we hasten to acknowledge that many researchers, including many of the researchers cited above, as well as many other cognitive scientists not explicitly concerned with historical language change, have argued that continuum structure plays a key role in addressing outstanding questions in the field (Rumelhart et al., 1986; Spivey, 2007). Indeed, some have shown how to build complex linguistic structures in continuum spaces (Cho et al., 2020; Coecke et al., 2010; Tabor, 2000). However, to our knowledge, none have addressed the Life ↔ Grammar question with these theories. It is also notable that LLMs, which seem to have broken through a barrier in making usable AI, are metric space language models (i.e., they belong to the class we are promoting here). But LLMs are analytically opaque. The particular type of model we focus on here, called a *Stretched Tree Metric Grammar* (STMG), has, like an LLM, a backpropagation-trained neural network at its core, but its internal structure is much simpler and is more transparently related to recursive symbol structure than the weights and hidden activation patterns of an LLM.

Foreshadowing our main conclusions, there are two distinctive properties of STMGs that support our approach to the Life ↔ Grammar question:

1. STMGs model grammaticality with a continuous valued quantity called “harmony", a close relative of the *harmony* of Harmony Theory (Smolensky and Legendre, 2006; Cho et al., 2020).

2. Following Elman (1991) and current LLM work, STMGs encode syntactic and semantic structure in one space, not as a relation between a syntactic space and a semantic one as is done, for example, in model-theoretic semantics (Montague, 1970)–see Tabor (2021).

Regarding (1), STMGs exhibit a hilly-landscape structure as shown in Figure 1. The x-axis corresponds to an encoding space—generally high dimensional and similar in its role to the space of a hidden layer in an LLM, but shown here as 1-dimensional for simplicity's sake. The y-axis indicates harmony. The green curve, which we refer to as the “grammar manifold", shows harmony as a function of encoding state. The encoding states correspond to word-forms in context, henceforth

[1] The Life ↔ Grammar question is closely related to the symbol grounding problem (Harnad, 1990) as well as the symbol ungrounding problem (Rączaszek-Leonardi and Deacon, 2018), but it puts the focus on (if and) how language systems and events in the world influence one another, independent of the question of symbols.

[2] Some linguists reject formalization altogether. We are open to this position provided that it is not imposed as a universal requirement on language research: we think there is value in particular kinds of informal thinking, and also in particular kinds of formal thinking.

referred to as “events". There is a threshold value, called the "observability threshold", shown by the horizontal blue line: events with harmony less than the threshold are unlikely to occur, while events with harmony above the threshold are likely to occur (i.e., they are found in corpora) and they are predicted to feel fairly grammatical to speakers.[3] Language innovation involves the occurrence of pragmatic circumstances that give items that are just below the observability threshold (shown in red in the figure) a boost that causes them to become observable (this happens because the act of usage slightly adjusts the shape of the grammar manifold, bringing something that was previously below the blue line, above it). For example, English does not have a verb “to marmalade" meaning to spread marmalade on something (cf. “to butter toast"). However, if someone were preparing breakfast in a cafeteria with a friend, and marmalade and butter were both on hand, they might launch into a question with, "Do you want to butter your toast or...", and then, for the sake of mild playfulness as well as syntactic parallelism, conclude with "marmalade it?" instead of the grammatically licensed "put marmalade on it?". In this fashion, something which had not been previously said, but was on the verge of being able to be said, has become said. The news in what we propose below is not the observation that such events happen—this is uncontroversial—nor is it in the claim that the change alters the grammatical system—usage-based approaches generally argue this. Rather it is in the specification of how the occurrence of such an event could modify a plausibly detailed formal grammar model. The theory makes testable predictions because the set of imminently innovatable expressions (red segments in Figure 1) occupies a small subset of the set of all unobserved expressions (events along those parts of the x-axis where the grammar manifold is below the blue line). In Section 3.2 below, we give examples predicted by our model of imminent and far-from-imminent potential innovations. These are predicted to have different felicity values, testable via judgment studies that support the detection of graded value judgments (Keller, 2000) and also with more elaborate psycholinguistic methods (Bornkessel-Schlesewsky et al., 2020).

As for the second central feature of STMGs, that

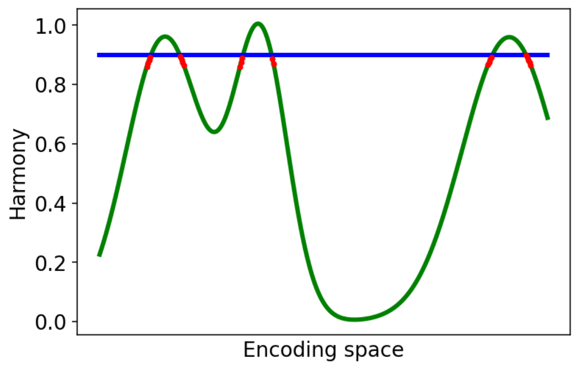

Figure 1: Schematic picture of the grammar manifold (green curve) as a function of position in encoding space (the encoding space is specified by the hidden layer of a neural network.)

syntax and semantic distinctions are encoded in the same space, a core part of this claim is that syntactic distinctions are encoded as big differences in location in the encoding continuum, and semantic distinctions are (generally) encoded as smaller differences within the regions of the big differences.[4] This makes it so that we only need to posit one kind of innovation, syntactico-semantic innovation. We claim that the only possible changes are small adjustments, predicting gradualness, but the accumulation of many small (semantic) changes can amount to a syntactic change.[5]

So far, we have identified key conceptual features of STMGs. We now turn to specifying concretely how they work.

2 Metric Grammars and Stretched Tree Metric Grammars

By “Metric Grammars", we mean dynamical systems on complete metric spaces that recognize/ generate sequences of symbols. This is a very broad mathematical class, capable of modeling formal languages of a wide variety of computational complexities in the traditional sense (including finite state languages, context free languages, and context sensitive languages among others—(Moore, 1998)). LLMs are in this class. Here, we focus on “Stretched-Tree Metric Grammars" (STMGs)

[3] By “word-form" we mean the physical word-form—phonetic or orthographic; in what follows we sometimes, for familiarty and brevity, refer to “words" but, by this, we mean “word-form events".

[4] This approach aligns with findings that LLMs generally learn syntactic distinctions early and semantic/pragmatic distinctions late (Bunzeck and Zarrieß, 2024).

[5] Such accumulating cases fall under the heading of *grammaticalization* (Hopper and Traugott, 2003), also sometimes called *constructionalization* (Traugott and Trousdale, 2016). We will not treat such cumulative changes in this paper, but the model is designed to predict them in addition to morphological innovations, which happen more simply and quickly.

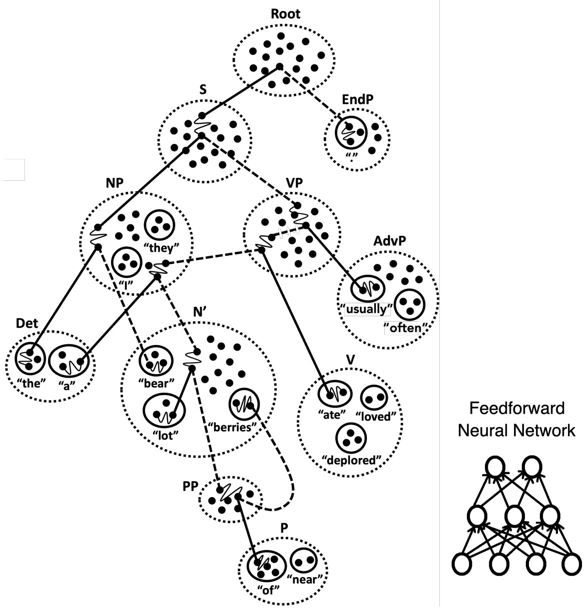

Figure 2: Schematic diagram of a Metric Grammar. The dots are vector space encodings (in R^2, the plane of the page) of grammatical nodes. All points are linked, either to sequenced pairs of other points, or to word points. The neural network in the lower right maps from pairs of points in the metric space to points in the metric space and also from one-hot encodings of words (not shown) to points in the metric space—in doing so, it implements the daughter(s) $\leftrightarrow$ mother links specified by constructions or context-free grammar rules. The linking of daughter sequences, or words, to mothers, is present for all points shown, but for ease of viewing, the diagram only illustrates the links involved in a parse of the sentence, "The bear usually ate a lot of berries".

(Tabor and Lee, 2024; Villata and Tabor, 2022), a different proper subset of Metric Grammars. These generate structures similar, though not identical, to the phrase structure trees generated by Context Free Grammars (CFGs) (Hopcroft and Ullman, 1979). An STMG can be parametrized on the basis of a corpus of CFG-parsed sentences. In the present work, we will generate such a corpus with a probabilistic CFG (or "PCFG").

Figure 2 shows a schematic illustration of an STMG, with a parse of one particular sentence highlighted. The metric space is R^2 (distance metric = cosine distance, all points in the first quadrant). The core STMG is specified by a set of triplets and lone-points, along with a neural network. Each triplet consists of a point called its "Mother", and two additional points called "Left Daughter" and "Right Daughter". The Left Daughter is joined to the Mother by a solid line, and the Right Daughter to the same Mother by a dashed line. The Left Daughter comes before the Right Daughter in the speech stream. The lone points are associated with lexical forms.

2.1 Generation

A metric grammar can function as a probabilistic generator of sentences, similar to a PCFG. The process is initiated at a point of type Root (circled at the top of Figure 2). We call this initially-chosen point a "Daughter Above". Then, another point is selected at random from all points in the space, with a weighting that favors points close to the Daughter Above (See (1): $\mathbf{x}$ is the location of the Daughter Above; $\alpha = 4$ is a free parameter,) . It is likely that another Root point will be chosen. This new point is called a "Mother Below". If the new point is a lexical point, then its word is generated and the process terminates (resulting in a one-word sentence). If the new point is a branching node, then the process is repeated, treating the Left Daughter and the Right Daughter of the Mother Below, respectively, as Daughters Above. When no more branching nodes remain to be explored, the process terminates. The result is a "Metric Grammar Tree" (MGTree) with terminal nodes that specify a sequence of words.

$$p_i(\mathbf{x}) = \frac{d(\mathbf{x}, \mathbf{x}_i)^{-\alpha}}{\sum_j d(\mathbf{x}, \mathbf{x}_j)^{-\alpha}} \qquad d(\cdot, \cdot) = \textit{cosine dist.} \tag{1}$$

MGTrees are similar to, but not the same as, parses generated by traditional grammars. Like traditional parses, MGTrees have a hierarchical structure reflecting constituency relationships. Unlike traditional tree nodes, the nodes of the MGTree lie in a metric space, so distances between them can be calculated. Moreover, the Daughters Above and associated Mothers Below are often separated from one another by a positive distance (highlighted by the squiggly lines in Figure 2). It is the presence of these positive separation-distances between Daughters Above and Mothers Below that motivates the modifier "Stretched Tree" in the name "Stretched Tree Metric Grammar" (STMG). Each such distance is divided by the diameter of the entire point set and subtracted from 1, producing a value in [0, 1]. This number is called the *harmony* of the Daughter-Mother bond. The product of these bond harmonies yields a *sentence harmony* (also in [0, 1]) and this number is taken as a model of the sentence's (predicted-to-be-gradient) grammaticality.

2.2 Encodings

The points shown in Figure 2 are located in a way that makes the figure roughly resemble a context free parse tree in the standard format used by linguists (root node at the top, lexical elements at the bottom). The figure is a caricature (drawn in this way for clarity of illustration): the actual points, generally lie in a space of dimension higher than 2, and encode the distributional context of each word and phrase in the corpus, similarly to the way the encodings of the words (and, we suspect, the phrases) processed by an LLM reflect their contextual occurrence tendencies. To initialize an STMG, however, we do not train it on a raw corpus of words, the way LLMs are trained. Instead, we start with a corpus of tree structures generated by a PCFG. The PCFG approximates aspects of the distributional behavior of a natural or artificial language that interests us. We define a 1-1 map from the cardinality K vocabulary of the language to the natural numbers, $\{1, \ldots, K\}$—this is the *vocabulary index* of each word. Consider a word w_0, that occurs in the stream of the corpus surrounded by preceding and following words: $\ldots w_{-3}w_{-2}w_{-1}w_0w_1w_2w_3\ldots$. The local encoding, C_0, of w_0 is given by

$$\begin{aligned} C_0 &= \tfrac{1}{2}(Ibit(w_{-1}) + \tfrac{1}{2}(Ibit(w_{-2}) + \ldots)) \\ &\qquad\qquad \oplus \\ &\quad \tfrac{1}{2}(Ibit(w_1) + \tfrac{1}{2}(Ibit(w_2) + \ldots)) \end{aligned} \tag{2}$$

where $Ibit(w_i)$ is the K-space one-hot vector with a 1 at the vocabulary index of word w_i and 0s in every other position, and $\oplus$ denotes vector concatenation. The vector C_0 is a form of fractal encoding (Barnsley, 1988; Tabor, 2000): it is a kind of data compression which assigns each bi-infinite context to a unique point in $2K$-space, emphasizing the information provided by more proximal words over that provided by more distant words. An exactly parallel encoding scheme is employed for each phrase in the corpus, using the one-hot encodings of the words preceding and following the phrase. The STMG coding of the Mother of a given word or phrase in the corpus is then formed by averaging the local encodings of a random sample of words/phrases that share the same mother node in the PCFG parses of all sentences in the corpus. For branching treelets, the Daughter points are initialized to the locations of their attached Mothers Below. This gives rise to the clusters corresponding to lexical and syntactic types shown schematically in Figure 2.

2.3 Neural Network

Let there be M mother nodes. Each corresponds to a point in $2K$-space. We form the cardinality M^2 set of vectors in $4K$-space corresponding to all possible ordered pairs of mother nodes. We then train, via backpropagation, an auto-encoder neural network with one hidden layer to map each pair of mother nodes to itself (the number of hidden units needs to be large enough to allow this mapping to be learned—for the very simple language we discuss below, we used 10 hidden units). We then remove the $4K$ dimensional output layer and the hidden-to-output weights of this trained network, and replace it with an untrained map from the hidden layer to $2K$-dimensional space. We train this grafted network on the function defined by the treelets of the STMG. That is, for the second network training, the inputs are grammatical sequences of daughter nodes (fractally encoded), and the outputs are the mother nodes that generate these daughter sequences (also fractally encoded). Let S stand for the encoding space of the STMG. The second trained network now implements a continuous map from $S \times S$ to S. We then augment the input units of the network to include one-hot codes for lexical items and perform a third network training to map, via the hidden layer to the fractal encodings of their lexical classes as specified by the PCFG. This positions the model to function as a parser of arbitrary sequences of words drawn from the vocabulary.

2.4 Parsing

Parsing is accomplished by self-organized sentence processing (Kempen and Vosse, 1989; Smith and Tabor, 2018; Tabor et al., 2004; Tabor and Hutchins, 2004; Velde and Kamps, 2006; Vosse and Kempen, 2000). This is a kind of continuous, bottom-up parse formation where nearby words and phrases bond together more strongly if they form a felicitous pairing. If it is possible, the system will generally stabilize in a highly grammatical configuration. This is a form of dynamical optimization. Since our focus here is not on online processing, we have not implemented full, self-organized parsing, but, instead, have implemented a (simpler) stochastic search for an optimal binary tree structure over the input words. The words are encoded as the fractal codes of their contexts (Verheyen et al., 2025) and then parsed as the nearest mother point in the metric space (cosine distance).

3 Grammatical Innovation

To model innovation, the STMG system employs repeated generation and parsing in an alternating fashion, with the key additional assumption that the model inhabits a world that can bias its production choices. This feature corresponds to the standard assumption that, at most moments of speaking, a speaker chooses what they will say according to their expressive needs and what their grammatical system supports. The work of much formalist linguistics suggests that, at any point in historical time, formal rules can do a reasonable of specifying all and only the utterances deemed grammatical by a particular speaker, and to a great extent by all the speakers who belong to a particular speech community. Something similar is true of the STMGs we have studied here: at a given point in historical time, the STMG's inventory of output structures can be fairly well characterized by a system of symbolic grammatical rules. However, unlike the symbolic rule system, the STMG is capable, in such a state, of generating utterances that are not within the output of its current symbolic description. It needs to be pushed a little bit in order to do this. The pushing is fairly gentle—of the sort portrayed above in the scenario involving butter and marmalade—but even one small push results in slight modification of the STMG's grammatical system, and many successive pushes in a consistent direction over long period of time can produce radical revision of the grammar.

At each step of the repeated generation and parsing mentioned above, the STMG replaces treelets of the generated parse in its metric space with treelets of a newly crafted parse of the same sentence. Under these assumptions, the STMG functions as a feedback dynamical system. Dynamical systems theory makes a distinction between autonomous dynamics and driven dynamics. Autonomous dynamics refers to the changes that a system undergoes as a function of its internal state. Driven dynamics refers to the changes a system undergoes due to forces on it that come from its environment. In the present case, the repeated cycle of generation and parsing implements autonomous dynamics of the STMG. The choice to use particular expressions, possibly mildly novel ones, that stem from speaker's expressive impulses and needs, amounts to an external driver of the system. The autonomous dynamics models generalization to previously unobserved items within the repertoire of a synchronic symbolic description of the STMG's generative character. The combined autonomous dynamics and external driver model innovation, including morphological innovation and grammaticalization.

Table 1: A PCFG with an ambiguous word, “dog", that can be either a noun or a verb.

1.00	Root	$\rightarrow$	S EndP
1.00	S	$\rightarrow$	NPsubj VPtense
1.00	VPtense	$\rightarrow$	Modal VPinf
0.67	VPinf	$\rightarrow$	Vtrans NPobj
1.00	NPsubj	$\rightarrow$	Detsubj Nsubj
1.00	NPobj	$\rightarrow$	Detobj Nobj
	VPinf	$\rightarrow$	sleep (0.18), gambol (0.09), howl (0.06)
	Modal	$\rightarrow$	may (0.55), could (0.27), should (0.18)
	Vtrans	$\rightarrow$	dog (0.48), like (0.24), abhor (0.16), resemble (0.12)
	Det_{subj}	$\rightarrow$	the (0.55), a (0.27), some (0.18)
	Det_{obj}	$\rightarrow$	the (0.55), a (0.27), some (0.18)
	N_{subj}	$\rightarrow$	dog (0.43), cat (0.22), vole (0.15) ferret (0.11), lemur (0.09)
	N_{obj}	$\rightarrow$	dog (0.43), cat (0.22), vole (0.15) ferret (0.11), lemur (0.09)
1.00	EndP	$\rightarrow$	.

3.1 Morphological Innovation: Denominal Verbs

Now, we turn to a simple case study, in which the model predicts the innovation of new denominal verbs (analogous to, for example, *to marmalade the toast*) but not new de-determiner verbs (e.g., *The caterer wanted to many the tables.*), or new denominal modals (e.g., *Josie edge like jalapeños*).

To train the model, we used a corpus of 400 sentences generated by the PCFG shown in Table 1. Note that in this grammar, one word, “dog", is ambiguous. It can serve as either a noun or a verb. To keep the pattern relatively simple, we used modal verbs in every sentence (so the main verb appeared in infinitive form) and we used singular nouns preceded by a determiner (1).

(1) a. A lemur may like a dog.
b. The cat might resemble a ferret.

The noun-verb ambiguity in this language is relatively easy for the STMG to handle because there are reliable cues in the surrounding words to the class membership—every verb is immediately preceded by a modal and every noun is immediately preceded by a determiner. Also making the learning easier, the grammatical function distinction, Subject vs. Object, was marked on the mother of each constituent of each noun phrase in the PCFG.

3.1.1 Generation

Once the model was trained, we used it as a generator to generate new parses in the fashion described in Section 2.1 above. We generated 200 novel sentences in this fashion. Every one of them corresponded to a grammatical parse under the source PCFG, indicating that the model had emulated the structure of the PCFG very well.

3.1.2 Parsing

The model showed essentially perfect parsing of trained sentences (Average Harmony = 1.00). We also tested the model on several kinds of non-trained sentences (2).[6]

(2)
a. The lemur may resemble the vole. (Grammatical, Trained: Unambiguous)
b. The cat should abhor the ferret. (Grammatical, Generalization: Unambiguous)
c. The ferret could dog the lemur. (Grammatical, Generalization: Denom. Verb)
d. The ferret may cat the lemur. (Precedented Innovation: Denom. Verb)
e. The ferret may resemble lemur vole. (Unprecedented Innov.: Denominal Determiner)
f. The ferret lemur follow the vole. (Unprecedented Innov.: Denominal Modal)

3.1.3 Simulation

To test the significance of our main prediction that there should be a 3-way distinction in harmony values with Generalization > Precedented Innovation > Unprecedented Innovation, we trained 30 metric grammars on 40 different samples from the PCFG of Figure 1. Each trained model was tested on types b (Generalization, N = 20), c (Generalization, N=20), d (Precedented Innovation, N = 40), e (Unprecedented Innovation, N = 20), and f (Unprecedented Innovation, N = 20) from (2).

Table 2: Descriptive statistics of sentence harmony (H) by condition.

Condition	*M*	*SD*	Min	Max
Generalization	0.994	0.005	0.953	1.000
Precedented Innovation	0.969	0.024	0.754	0.995
Unprecedented Innovation	0.861	0.066	0.583	0.985

3.1.4 Analyses and Results

Table 2 shows descriptive statistics by condition. A linear mixed-effects (LME) model predicting sentence harmony (H) was fitted with Condition as a fixed effect and Participant as a random intercept, using the `lme4` package (version 1.1.37; Bates et al. 2015). p-values were estimated via the `lmerTest` package (version 3.1.3; Kuznetsova et al. 2017), and pairwise comparisons were then conducted with the `emmeans` package (version 1.11.1; Lenth 2025). Results are presented in Table 3 and Table 4.

The results provide evidence that the STMG assigns very high harmony to generalizations, minimally degraded harmony to precedented innovations, and lower harmony to unprecedented innovations. In the dynamical framework we have outlined, this result amounts to a prediction that precedented morphological innovations (e.g., denominal verbs) are ripe for innovation, while unprecedented morphological innovations are not (Figure 1).

3.2 Evolution with a driver

We implemented a simple version of the evolution cycle described above. It uses a batch-processing approach as follows:

1. Use a PCFG approximating the language under study to generate a corpus of CFG tree structures.

2. Use the corpus to parameterize a metric grammar (use fractal encoding of each word and phrase in sequence to determine the metric grammar points; train a neural network to map sequences of daughter points to mother points).

3. Generate a new corpus of metric grammar trees with the metric grammar.

[6] In these test examples, "Grammatical, Trained" means the form is generated by the PCFG and occurred in the training corpus, "Grammatical Generalization", means it can be generated by the PCFG but did not occur in the training corpus, "Precedented Innovation" means the novel usage has a parallel in a trained PCFG form (denominal verb case), "Unprecedented Innovation" means the novel usage has no PCFG precedent.

Table 3: Linear mixed-effects model results.

Predictor	***b***	**SE**	**95% CI**	***t***	***p***
(Intercept)	0.9943	0.0022	[0.9900, 0.9986]	453.24	$<.001^{***}$
Generalization vs Precedented Innovation	-0.0254	0.0016	$[-0.0286, -0.0222]$	-15.78	$<.001^{***}$
Generalization vs Unprecedented Innovation	-0.1332	0.0016	$[-0.1364, -0.1300]$	-82.69	$<.001^{***}$

Model: $H \sim$ Condition + (1|Participant). Since test items were newly generated for each model, they were not included as random effects. CI = confidence interval. $^{***}p < .001$.

Table 4: Pairwise comparisons results.

Contrast	***b***	**SE**	**95% CI**	***t***	***p***
Generalization − Precedented Innovation	0.0254	0.0016	[0.0216, 0.0293]	15.78	$<.001^{***}$
Generalization − Unprecedented Innovation	0.1332	0.0016	[0.1294, 0.1371]	82.69	$<.001^{***}$
Precedented Innovation − Unprecedented Innovation	0.1078	0.0016	[0.1040, 0.1117]	66.90	$<.001^{***}$

p-values were Bonferroni-corrected for 3 comparisons. CI = confidence interval. $^{***}p < .001$.

4. To model driven evolution, replace a small number of sentences in the corpus with borderlined-observable examples (e.g., “marmalade the toast", “cat an X" in the simulation, red events in Figure 1.)

5. Reparse the new corpus with the metric grammar.

6. Return to Step (2).

When Step 4 is skipped, these steps implement autonomous evolution. Applying the procedure starting from the PCFG of Table 1, repeating over multiple generations, the model stably produced all and only sentences recognized by the initial PCFG.

To implement driven evolution, we ran the procedure with Step 4 included. Specifically, at Step 4, we randomly deleted 4 trees (2%) from the 200-tree corpus formed at step (3) and replaced these with 4 trees employing “cat" as a verb (note that “cat" is never used as a verb in the original corpus.) Figure 3 shows the harmony of the "cat"-as-verb sentences as we increased the number of novel "cat"-verb sentences in the training corpus. The average harmony of the novel denominal verb form goes from slightly below the average harmony of attested grammatical forms, to slightly above them as the percentage of novel forms in the training corpus increases from 0% to 8%. By contrast, the rate of use of denominal determiner and denominal modals remains below the level of the forms that are likely to be attested. Thus, the model demonstrates a process by which usage can change the grammatical status of a word from being borderlined acceptable to being clearly acceptable. Figure 3 suggests that there is some tendency for the training of the novel verb to destablize the behavior of

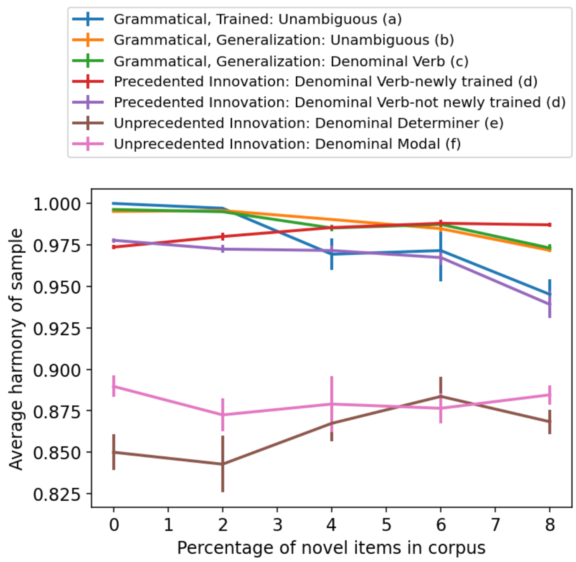

Figure 3: Change in the average harmony of test sentences with progressively increasing frequency of a novel denominal verb in the training corpus (Error bars are standard errors of the means across 5 runs, each with one STMG tested on 10 items per condition).

the existing grammatical constructions—we comment on this observation below.

4 Conclusions

We started by noting that existing treatments of linguistic structure in a range of theories are challenged by the Life $\leftrightarrow$ Grammar question, which asks how events that occur in the world can give rise to changes in the form of a grammar. We argued that a viable approach to answering this question can be achieved by constructing grammars in a continuum topology rather than a discrete topology. We presented Stretched Tree Metric Grammars (STMGs) as a concrete example of how to do this, and studied a very simple model of English denominal verb innovation with an STMG.

4.1 Strengths

The model exhibited a form of usage-grammar interaction that typifies naturally occurring historical change episodes and, to our knowledge, has not previously been demonstrated in a model. The model also works similarly to an LLM in that it computes in a metric space and small, progressive changes in its training set can lead to changes in the grammatical properties that the model exhibits. Since the model is more easily interpretable than an LLM, understanding why it shows this property may shed light how LLMs compute and on how they differ from classical symbolic language models, which do not show this mutability property.

4.2 Weaknesses

The simulation is very "toy". It will be helpful to see if it scales up. In this regard, one encouraging feature of the STMG framework is that it does not have to be initially parameterized with a PCFG. The parameterization procedure can also be applied to any collection of labeled, context-free parse structures. Thus, it is suitable for use with parsed corpora which have been curated by linguists. These are arguably more natural corpora than those generated by PCFGs.

We noted in Section 3.2 above that the introduction of the novel denominal verb seems to lower the harmony of existing constructions. We do not know if this is a robust tendency of the model, but we think it likely that it could be because the model is very sensitive—every event of usage has an effect on the entire encoding system. Although this property may be suitable for capturing morphosyntactic chain shift phenomena, it does not seem plausible for minor morphological innovations of the sort we are considering here. Future work on this model will need to thoroughly investigate its robustness under parametric perturbation.

Additionally, although we claim the model is more interpretable than an LLM, we have not actually provided an analysis of why innovation of denominal verbs in the model is more viable than other cross-category innovations. We believe the difference stems from the fact that the training grammar already includes a denominal verb, but it does not include any other cross-category items; the existing denominal verb, "dog" provides the basis for an analogy supporting innovation of other denominal verb forms (Misra and Mahowald, 2024). However, this interpretation is currently a speculation. Further analysis is needed to explore the question. Several methods may be useful here: (i) we can use training set ablation to discover what ingredients make the observed innovation viable; (ii) we can directly analyze the structure of the daughters-to-mother map learned by the neural network to see if the parses of the innovative denominal verb are benefiting from proximity to the existing denominal verb examples.

4.3 Future Work

Additional avenues seem interesting to explore. Do LLMs (or, more realistically Small Language Models or "SLMs") trained on synthetic data show similar innovation tendencies? Do the predictions of the STMG about what can and cannot be immediately innovated in English correspond with human judgments? One may check, for example, zero-derivation of de-adjectival verbs, de-adverbial verbs, de-modal verbs, etc. as well as many other morphological combination possibilities and non-possibilities—e.g., ?farness vs. *marinity (cf. obscenity). Also, we have touted STMGs as a general theory of language innovation, but we have only tested them here on a relatively short-term kind of change, morphological innovation. It is desirable to see how well the method models grammaticalization, a much more drawn out language change process, which can support more radical category shift (Hopper and Traugott, 2003).

Concluding, we suggest that the Metric Grammar framework may open new avenues for a conversation among a variety of researchers: usage-based theorists (the theory offers an equational characterization of innovation and entrenchment), classical formal theorists (it may help integrate formal semantics with distributional models—cf. Asher et al. (2017)), psycholinguists (it may bridge between the toy examples considered here and the complexities of natural stimuli), and LLM researchers (Exploring language stability via iterative autonomous dynamics may be a useful way to study LLMs as well as STMGs).

Acknowledgments

We thank Yan Kolupaev for helpful assistance during the early stages of STMG programming, and the members of Solab for their continual support and valuable feedback.

References

Nicolas Asher, Tim Van de Cruys, Antoine Bride, and and Márta Abrusán. 2017. Integrating Type Theory and Distributional Semantics: A Case Study on Adjective–Noun Compositions. *Computational Linguistics*, 42(4):703–725.

Michael Barnsley. 1988. *Fractals Everywhere*. Academic Press, Boston.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Katrien Beuls and Remi Van Trijp. 2016. Computational construction grammar and constructional change. *Belgian Journal of Linguistics*, 30:1–13.

Ina Bornkessel-Schlesewsky, Dietmar Roehm, Robert Mailhammer, and Matthias Schlesewsky. 2020. Language Processing as a Precursor to Language Change: Evidence From Icelandic. *Frontiers in Psychology*, 10.

Lieselotte Brems. 2003. Measure Noun constructions: An instance of semantically-driven grammaticalization. *International Journal of Corpus Linguistics*, 8(2):283–312.

Bastian Bunzeck and Sina Zarrieß. 2024. Fifty shapes of BLiMP: syntactic learning curves in language models are not uniform, but sometimes unruly. In *Proceedings of the 2024 CLASP Conference on Multimodality and Interaction in Language Learning*, pages 39–55. Association for Computational Linguistics.

Joan Bybee, Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World*. University of Chicago Press, Chicago, IL.

Joan L. Bybee. 1985. *Morphology: A Study of the Relation between Meaning and From*. John Benjamins, Amsterdam.

Pyeong Whan Cho, Matthew Goldrick, and Paul Smolensky. 2020. Parallel parsing in a gradient symbolic computation parser.

Noam Chomsky. 1986. *Knowledge of language : its nature, origin, and use*. New York : Praeger.

Noam Chomsky. 1995. *The minimalist program*. MIT Press, Cambridge, MA.

B. Coecke, M. Sadrzadeh, and S. Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *Linguistic Analysis*, 36:345–384.

Hendrik De Smet. 2012. The Course of Actualization. *Language*, 88(3):601–633. ERIC Number: EJ986712.

Jeffrey L. Elman. 1991. Distributed Representations, Simple Recurrent Networks, and Grammatical Structure. *Machine Learning*, 7:195–225.

Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language*, 64(3):501–538.

Adele E. Goldberg. 2006. *Constructions at Work*. Oxford University Press, Oxford.

Carolyn Gottfurcht. 2008. *Denominal verb formation in English*. PhD, Northwestern University.

Stevan Harnad. 1990. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1):335–346.

Bernd Heine, Ulrike Claudi, and Friederike Hunnemeyer. 1991. *Grammaticalization: A Conceptual Framework*. University of Chicago Press, Chicago.

John E. Hopcroft and Jeffrey D. Ullman. 1979. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Menlo Park, California.

Paul J. Hopper. 1991. On some principles of grammaticization. In Elizabeth Closs Traugott and Bernd Heine, editors, *Approaches to Grammaticalization, v. 1*, pages 17–36. John Benjamins.

Paul J. Hopper and Elizabeth Closs Traugott. 2003. *Grammaticalization, 2nd Edition*. Cambridge University Press, Cambridge, England.

Frank Keller. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality.

Gerard Kempen and T. Vosse. 1989. Incremental syntactic tree formation in human sentence processing: a cognitive architecture based on activation decay and simulated annealing. *Connection Science*, 1:273–290.

Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26.

G Lakoff. 1977. Linguistic Gestalts. *Linguistic Gestalts*, 13:236–287.

Ronal W. Langacker. 1988. A Usage-Based Model. In Brygida Rudzka-Ostyn, editor, *Topics in Cognitive Linguistics*, pages 127–164. John Benjamins.

Ronald Langacker. 1977. Syntactic Reanalysis. In Charles N. Li, editor, *Mechanisms of Syntactic Change*, pages 57–139. University of Texas Press.

Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar, v. 1*. Stanford University Press, Stanford, California.

Russell V. Lenth. 2025. *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.11.1.

Frantisek Lichtenberk. 1991. On the gradualness of grammaticalization. In Elizabeth Closs Traugott and Bernd Heine, editors, *Approaches to Grammaticalization, v. 1*, pages 37–80. John Benjamins.

Kanishka Misra and Kyle Mahowald. 2024. Language Models Learn Rare Phenomena from Less Rare Phenomena: The Case of the Missing AANNs. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 913–929, Miami, Florida, USA. Association for Computational Linguistics.

Richard Montague. 1970. English as a formal language. In B. Visentini et al, editor, *Linguaggi nella e nella Tecnica*, pages 189–224. Edizioni di Comunità.

Cris Moore. 1998. Dynamical Recognizers: Real-time Language Recognition by Analog Computers. *Theoretical Computer Science*, 201:99–136.

Jean Petitot. 2011. *Cognitive Morphodynamics: Dynamical Morphological Models of Constituency in Perception and Syntax*. Peter Lang AG, Internationaler Verlag der Wissenschaften, Bern, Suisse.

Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. The University of Chicago Press, Chicago.

Ian Roberts. 1993. A Formal Account of Grammaticalisation in the History of Romance Futures. *Folia Linguistica Historica*, XIII/1–2:219–58.

David E. Rumelhart, James L. McClelland, and the PDP Research Group. 1986. *Parallel Distriubted Processing. Explorations in the Microstructure of Cognition. Volume 1*. MIT Press, Cambridge, Massachusetts.

Joanna Rączaszek-Leonardi and Terrence W. Deacon. 2018. Ungrounding symbols in language development: implications for modeling emergent symbolic communication in artificial systems. In *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pages 232–237. ISSN: 2161-9484.

Ferdinand de Saussure. 1916. *Course in General Linguistics*. Columbia University Press. Pages: 336 Pages.

Hans-Jörg Schmid. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford University Press, Oxford, New York.

Garrett Smith and Whitney Tabor. 2018. Toward a Theory of Timing Effects in Self Organized Sentence Processing. In *Proceedings of the 16th international conference on cognitive modeling*, pages 138–143.

Paul Smolensky and Geraldine Legendre. 2006. *The Harmonic Mind: from neural computation to optimality-theoretic grammar*. MIT Press, Cambridge, MA.

Michael Spivey. 2007. *The Continuity of Mind*. Oxford University Press, New York.

Luke Steels, editor. 2011. *Design Patterns in Fluid Construction Grammar*. Constructional Approaches to Language. John Benjamins Publishing Company.

Whitney Tabor. 2000. Fractal encoding of context-free grammars in connectionist networks. *Expert Systems: The International Journal of Knowledge Engineering and Neural Networks*, 17(1):41–56.

Whitney Tabor. 2021. On the Relationship between Syntactic and Semantic Encoding in Metric Space Language Models. *Journal of Cognitive Science*, 22(2):35–67.

Whitney Tabor, Bruno Galantucci, and Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50:355–370.

Whitney Tabor and Sean Hutchins. 2004. Evidence for Self-Organized Sentence Processing: Digging In Effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2):431–450.

Whitney Tabor and Hyosun Lee. 2024. Metric Grammars. Poster presented at the 2024 Conference of the Cognitive Science Society.

Elizabeth Closs Traugott and Graeme Trousdale. 2016. *Constructionalization and Constructional Changes*. Oxford Studies in Diachronic and Historical Linguistics. Oxford University Press, Oxford, New York.

Paul Van Eecke and Katrien Beuls. 2018. Exploring the Creative Potential of Computational Construction Grammar. *Zeitschrift für Anglistik und Amerikanistik*, 66(3):341–355.

Frank van der Velde and Marc de Kamps. 2006. Neural Blackboard Architectures of Combinatorial Structures in Cognition. *Behavioral and Brain Sciences*, 29:37–108.

Lara Verheyen, Jonas Doumen, Paul Van Eecke, and Katrien Beuls. 2025. You Shall Know a Construction by the Company it Keeps: Computational Construction Grammar with Embeddings. In *Proceedings of the Second International Workshop on Construction Grammars and NLP*, pages 75–83.

Sandra Villata and Whitney Tabor. 2022. A self-organized sentence processing theory of gradience: the case of islands. *Cognition*, page 104943.

Theo Vosse and Gerard Kempen. 2000. Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75:105–143.

Uriel Weinreich, William Labov, and Marvin Herzog. 1968. Empirical Foundations for a Theory of Language Change. In W. P. Lehmann and Yakov Malkiel, editors, *Directions for Historical Linguistics*, pages 95–188. University of Texas Press.