

UC Merced

UC Merced Electronic Theses and Dissertations

Title

Sp[i]CALU: A dynamical systems model of the creative aspect of language use

Permalink

<https://escholarship.org/uc/item/9g7311xr>

Author

Lichtenstein, Patricia

Publication Date

2017

Peer reviewed|Thesis/dissertation

*Sp[i]CALU: A Dynamical Systems Model
of the Creative Aspect of Language Use*

A DISSERTATION PRESENTED
BY
PATRICIA LICHTENSTEIN
TO
THE DEPARTMENT OF COGNITIVE & INFORMATION SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
COGNITIVE & INFORMATION SCIENCES

UNIVERSITY OF CALIFORNIA, MERCED
MERCED, CALIFORNIA
AUGUST 2017

© 2017 - *PATRICIA LICHTENSTEIN*
ALL RIGHTS RESERVED.

Sp[i]CALU: A Dynamical Systems Model of the Creative Aspect of Language Use

ABSTRACT

In this dissertation I present research on the creative aspect of language use. I focus on Blind Variation and Selective Retention, a process whereby a creative system produces an outcome by first generating with no imposed constraints all variant outcomes and then retains from these variegated candidates the optimal outcome by gradually introducing selective constraints. I carry out a dynamical systems analysis of a model incorporating a maximum entropy construction based on pairwise correlations among interacting elements of the system and a Metropolis walk in energy space. I conduct computational and behavioral experiments to test the validity of the outcomes emerging from the model. The scientific motivation is to understand the processes underlying creative cognition, including how previously impossible outcomes can be discovered and produced by a creative system, and what factors contribute to the viability of creative outcomes.

First I study and analyze a Potts Hamiltonian model of Blind Variation and Selective Retention. When systems operate near a critical point, I show that the energy landscape described by the model can provide reasonable estimates of empirically observed data. I also show that the energy space decomposes into several clusters that promote discovery of viable unobserved states. I compare my results with other computational models and findings from human subject experiments, and show that my model consistently predicts outcomes that are novel, surprising,

valuable, and intelligible.

The model I develop demonstrates the idea that creative processes should be viewed as emergent, collective phenomena. This idea represents the crux of a long-standing debate in creative cognition. The challenge has been to move beyond concepts by developing robust mathematical models that can be probed and analyzed. The approach I propose provides a framework for meeting this challenge.

Contents

| | | |
|----------|--|-----------|
| 1 | SETTING THE SCENE | 1 |
| 1.1 | The contribution of the CALU to human activity | 2 |
| 1.2 | Scientific investigation of the CALU | 3 |
| 1.3 | Current contribution | 16 |
| 2 | CASTING THE PLAYERS | 19 |
| 2.1 | Distributional semantic models | 19 |
| 2.2 | Clustering by association | 21 |
| 2.3 | Clustering with affinity propagation | 24 |
| 2.4 | Summary | 27 |
| 3 | SPINNING THE SCRIPT | 33 |
| 3.1 | Networks of interacting semantic fields | 35 |
| 3.2 | Solving the model | 36 |
| 3.3 | Testing and investigating the model | 37 |
| 3.4 | Landscaping with basins of attraction | 44 |
| 3.5 | Discussion & summary | 47 |
| 4 | GAUGING THE PERFORMANCE | 49 |
| 4.1 | Sp[i]CALU <i>in silico</i> | 50 |
| 4.2 | Sp[i]CALU <i>in feras</i> | 51 |
| 4.3 | Cross-linguistic semantic dissonance | 53 |

| | | |
|-----|---|-----------|
| 4.4 | Intelligibility experiments | 59 |
| 4.5 | General discussion | 67 |
| 4.6 | Conclusion | 72 |
| 5 | CURTAIN CALLS AND COMING ATTRACTIONS | 73 |
| | APPENDIX | 83 |
| | REFERENCES | 94 |

Listing of figures

| | | |
|-------|--|----|
| 2.3.1 | Cluster 11. Terms for things that flow and float are associated through verb semantic fields with terms used to discuss economics and terms used to discuss intelligence and mild emotions. | 28 |
| 2.3.2 | Cluster 16. Terms for spilling and seeping are associated through verb semantic fields with terms used to discuss illness and terms used to discuss emotions that deplete. | 29 |
| 2.3.3 | Cluster 29. Terms for cooking ingredients are associated through verb semantic fields with terms used to discuss math and terms used to discuss communicable emotions. | 30 |
| 2.3.4 | Cluster 38. Terms for valuable things and people who deal with valuable things are associated through verb semantic fields with terms used to discuss heredity and terms used to discuss eternal emotions. | 31 |
| 2.3.5 | Cluster 39. Terms for bodily functions and fluids are associated through verb semantic fields with terms used to discuss governance and terms used to discuss emotions of transfer. | 32 |

| | | |
|-------|---|----|
| 3.0.1 | Spin systems are complex systems that can be modeled as lattices where each node can be in one of any of the possible spin states of the given system. These can be states of magnetization, velocity, chemical composition, or electrical activation (for example). Global characteristics of the system can be modeled as emerging from local interactions of spin states across neighboring sites in equilibrium and non-equilibrium conditions. | 34 |
| 3.3.1 | Probabilities of site correlations of modeled training data plotted against probabilities of site correlations of observed testing data. Correlations are for the $k = 1$ through $k = 4$ nearest semantic fields. The probabilities have been sorted into bins from 0 to 1 where bin-width is 0.01. | 38 |
| 3.3.2 | Probabilities of site correlations of observed training data plotted against probabilities of site correlations of observed testing data. The scatter here is of a commensurate magnitude with that of the model as shown in Figure 3.3.1, which indicates that the model is nearly as accurate as the data allow. | 39 |
| 3.3.3 | Comparison of probabilities of site correlations between observed training data and modeled training data plotted against probabilities of site correlations of observed testing data. The comparison indicates that the magnitude of scatter of the modeled data is slightly greater than that of the observed data. | 40 |
| 3.3.4 | The observed data (the red line) and the model data (the green line) both follow a Zipfian distribution where FREQUENCY is plotted over RANK. Emergent global behavior where microcanonical entropy is inversely proportional to energy are in proximity to a critical point. | 42 |

| | | |
|-------|---|----|
| 3.3.5 | Mutual information across semantic field sites is relatively weak when calculated across metaphor ensembles. With regard to the CALU and creative cognitive processes, this has implications relevant to the theories that emphasize the importance of divergent thinking for creative ideation. If it could be shown that creative metaphors were more likely to be found by including more numerous and more diverse metaphor ensembles, such a finding would support those theories. | 43 |
| 3.3.6 | Within single metaphor ensembles, the mutual information between semantic fields strengthens. | 44 |
| 3.4.1 | The entropy of all the metaphor ensembles from frequency counting, from the independent model, and all K interacting semantic field models indicate that the local interactions are having a strong effect on the global features of the system. | 45 |
| 3.4.2 | Creative metaphors can be retrieved by percolating the system to recover the attested vectors in closest proximity in energy space to the unattested vectors. Discrepancies in the active semantic fields yield creative metaphors. | 47 |
| 4.4.1 | Instructions seen by participants in Experiment 1. | 61 |
| 4.4.2 | Instructions seen by participants in Experiment 2. | 64 |
| 4.4.3 | Instructions seen by participants in Experiment 3. | 65 |
| 4.5.1 | Agreement rates per item by condition. | 71 |

To H.

Acknowledgments

First, I would like to express my deepest gratitude to my advisor Teenie Matlock. I have benefited not only from her knowledge and great insight as a linguist and cognitive scientist, but also from her kindness and patience. The dynamic Cognitive and Information Sciences program at the University of California, Merced is a testimony to Teenie's dedication to the scientific community at large and her personal commitment to academia. I would also like to thank my advisor Stephanie Shih whose inspiration and energy helped me very much in pursuing my own research goals. My knowledge of linguistic phenomena was greatly enhanced through working with Stephanie. My dissertation committee members Michael Spivey and Anne Warlaumont brought valuable suggestions at every stage of my dissertation research. My deepest acknowledgement is to Henry for being my condition of possibility.

Everything should be at once surprising and inevitable.

Beethoven

A simile that has been absorbed into the forms of our language produces a false appearance which disquiets us. 'But this isn't how it is!' – we say. 'Yet this is how it has to be!'

Wittgenstein

1

Setting the Scene

THE ABILITY TO CREATE effective solutions to unexpected problems is essential for adaptive success during periods of environmental flux. *Homo sapiens* excel in this ability, referred to as creativity, to the extent that many scholars point to the Creative Explosion of the Upper Paleolithic as the turning point in the remarkable influence and longevity of the species [24, 41, 67, 91]. The study of creativity spans the fields of anthropology, philosophy, psychology, artificial intelligence and neuroscience. Considerable interest is currently directed toward understanding the cognitive and neural basis of creativity. In this dissertation, I primarily consider the Blind Variation and Selective Retention (BVSR) hypothesis of creative processing [21] and how it can be modeled in a manner compatible with neurobiology. The models and experiments I present here are built upon linguistic data, and it is my hope to extend my findings to encompass broader knowledge of the creative aspect

of language use and language systems dynamics.

1.1 THE CONTRIBUTION OF THE CALU TO HUMAN ACTIVITY

The creative aspect of language use (CALU) is integral to human activity. Conversational creativity enters into our daily lives forging interpersonal relationships and group identity [68], [128]. It is an ordinary everyday occurrence, which makes it an aspect of ordinary everyday cognition, a cognitive poetics of everyday language ([47], [25], [93]).

Linguistic creativity sparks innovation in activities vital to human progress, with many well-known episodes coming from the annals of scientific discovery and artistic achievement ([31], [88]). Werner Heisenberg (whose father was a philologist and professor of philology at the University of Munich) reflected repeatedly on the relationship between language and scientific progress, writing the following in the manuscript *Reality and its Order* [59].

There is another way of representing reality that can be set against the “static” one; we may call it “dynamic” representation. It is made possible by those infinitely complex associations among words. An idea expressed in this way is not meant to be as faithful a representation of reality as possible but to be the seed for further series of ideas. The issue is not the accuracy but the fruitfulness of concepts. As a result of the complex associations among words, new ideas attach themselves to one idea, further new ideas emerge from these, until finally in hindsight, a faithful depiction of an area of reality under consideration takes shape from the abundance of substance in the space the ideas have traversed and measured. This sort of representation is based in the vitality of the word itself. Here, a sentence can, generally speaking, not be “right” or “wrong”. But one may call a sentence “true” when it fruitfully leads to an abundance of other ideas. The opposite of a “right” sentence is a “false” one. But the opposite of a “true” sentence will often be another “true” one.

Here, Heisenberg is suggesting that scientific creativity can be driven through the play of linguistic forms and functions, by traversing the matrix of words and transforming it to identify new concepts and transmute old ones. From this vantage point, the CALU serves as a core tool in our cognitive toolbox for expanding and deepening our concepts about the world and the phenomena in it.

1.2 SCIENTIFIC INVESTIGATION OF THE CALU

1.2.1 INFINITUDE AND THE CALU

Given the centrality of the creative aspect of language use to progress in science, the phenomenon itself has yet to be ensconced as one worthy of, or even amenable to, scientific study. Early generative theory identified the scientific explanation of the creative aspect of language use as the “central fact to which any significant linguistic theory must address itself” and thus “a theory of language that neglects this creative aspect is of only marginal interest” [?]. Generative grammar were proposed as the mechanism behind the creative aspect of language use. The mechanism consists of the recursive application of a finite set of rules. The underlying assumption of the generative model is that the CALU is inherently and inextricably a matter of mapping a finite set of inputs to an infinite variety of outputs. In generativist theory, the discrete infinitude of possible output strings is considered a universal property of human languages, a notion rooted in Port-Royal grammar and logic [3] and von Humboldt’s philosophy of language [60].

Yoking the creative aspect of language to infinite cardinality, however, impedes the scientific study of it in several respects. A finite-set-to-infinite-set mapping mechanism such as a generative grammar is insufficient, for instance, as a scientific model of the CALU. Consider phrase structure rule (1).

1. Adj \rightarrow very Adj

Over the six-word vocabulary music, lifts, his, very, good, mood, an infinite set of sentences music lifts his very good mood, music lifts his very very good mood,

... , music lifts his very very very very very good mood can be generated, but is this illustrative of the CALU? Consider phrase structure rule (2).

2. Adj \rightarrow not Adj

Over the six-word vocabulary music, lifts, his, not, good, mood, again an infinite set of sentences can be generated music lifts his not good mood, music lifts his not not good mood, ... , music lifts his not not not not not good mood, and again we can question whether this demonstrates the CALU in action. According to [27], where the CALU is described as a process involving “unboundedness, novelty, freedom from stimulus control, coherence and appropriateness to situations”, it clearly does not. Instead, [27] claims that generative grammar provides the necessary but not sufficient mechanism for enabling the CALU and concedes that the fundamental organization and behavior of the CALU remains a “mystery”. This latter concession damns the CALU to scientific purgatory in a manner that is paralleled in some of the broader literature on creative processes and creative cognition. The reasoning goes that creativity by its very nature is unpredictable and has no associated causal structure, making it in a sense “supernatural” and inaccessible to investigation via the scientific method [66].

But does the necessary mechanism for the CALU necessarily involve infinitude or unboundedness? Pullum & Scholz [97] argue that it is not, offering as an illustration the Japanese haiku verse form. A haiku is composed of 17 phonological elements called morae, of which there exist around 100 in Japanese. So, while the space of possible mora is strictly finite and the space of possible combinations is not unbounded (somewhat less than 10^{17} given phonotactic, semantic, pragmatic, and aesthetic considerations), the combinatorial potential is nevertheless vast enough to afford the generation of unique, coherent, and compelling verse into the foreseeable future.

The generative framework thusly failed to establish a scientific approach to the CALU due to its focus on unboundedness and consequent perseverance in designing a finite-set-to-infinite-set mechanism that could serve as the foundation of the

CALU.¹ In the process, “freedom from stimulus control, coherence and appropriateness to situations” were relegated to the mystery bin of scientifically unapproachable phenomena. All are features of context: coherence is a feature of linguistic context and freedom from stimulus control in this case is related to variation as conditioned by the linguistic environment; appropriateness is a feature of pragmatic context, with freedom from stimulus control here related to variation as conditioned by the situational environment (socio-cultural, interpersonal, current moment, and so forth). Variation in context then may be more fertile ground for planting the seeds of a scientific approach to the CALU.

1.2.2 THE CALU AND LINGUISTIC CONTEXT

Variation as conditioned by linguistic context plays a dominant role in the poetic function, one of six major functions operable during any given instance of verbal communication according to Jakobson.²

Within this framework, the productivity of language arises from the interaction of two mutually integrated operations, substitution and contexture. Substitution involves the selection of a linguistic unit (henceforth referred to as a sign) from a repertoire of prefabricated signs. The sign repertoire can be thought of as a lattice where the nodes are signs and the undirected edges are similarity relations between the signs (relations such as similarity, dissimilarity, synonymy, antonymy, contrast, opposition, metonymy, and metaphor). Substitution also involves the selection of a sign template. A sign template is equivalent to a construction in contemporary cognitive grammar (for example: subj + verb, subj + verb + obj, obj + verb + prep

¹Michael Spivey has suggested (personal communication) that an exploration of the CALU in terms of unboundedness and infinite cardinality may still be informative when modeled as a Cantor’s dust fractal. In such a model, the semantic space will be initialized as a uniform matrix of semantic fields (as is the case in the spin models presented here). An energy landscape corresponding to the observed data can then be simulated using the method of successive removals. Random fluctuations in or percolation of the system will lead to potentially creative novel expressions.

²The other five functions are emotive (expressive), conative (appellative), metalingual (“glossing”), referential (denotative), and phatic. Most acts of verbal communication incorporate some combination of the six functions.] The poetic function “projects the principle of equivalence from the axis of selection into the axis of combination” [130].

+ subj, subj + verb + prep + obj). Contexture combines the selected signs and sign templates according to relations of contiguity. Contexture can be thought of as arising from a lattice where the nodes are the selected signs and the directed edges are the contiguity relations specified by the selected template.

To make the above more concrete, consider the process of conveying a message about music acting positively on someone's mood. The relevant sign repertoires would consist of substitutable signs along the lines of music, song, tune, melody, rhythms, noise, cacophony, clamour, ..., improve, benefit, enhance, lift, elevate, brighten, worsen, deplete, dampen, ..., mood, spirit, feelings, state of mind, well-being, emotions, happiness, sadness, The relevant templates would be subj + verb + obj, obj + verb + prep + subj, subj + verb + prep + obj. If the selected signs are "music", "elevate", "spirit", and the selected template is obj + verb + prep + subj, then contiguity relations specify the contexture "spirit is elevated by music".

In the sentence "Our spirits are elevated by music" the constituents are not necessarily related by equivalencies (e.g., "music" as a sign is not characterized by its linguistic similarity to the sign "elevate") but by contiguities (e.g., contiguities in time and space, agency, and causation). The poetic function, in superinducing substitution upon the process of combining, might specify a contexture such as "Melodies elevate, spirits soar", where the constituents are related by phonetic, inflectional, syntactic, and semantic similarities and repetitions.

In addition to accounting for linguistic context, this approach to describing the processes underpinning the CALU produces language that overtly draws our attention to the "complex association of words" that Heisenberg assigns importance to in creative ideation processes in general. And it produces language that, I suspect, appeals more forcefully to intuitive notions of how the CALU functions in poetic discourse as well as spontaneous creativity in language use [121]. It falls short, however, in that it cannot account for pragmatic context and its contribution to the CALU.

1.2.3 THE CALU AND PRAGMATIC CONTEXT

Any act of verbal communication is a joint activity during which the communicative partners jointly coordinate their verbal and nonverbal behavior based on event-specific or socially-established common ground [28]. While this aspect of verbal communication is clearly present in spontaneous face-to-face communicative acts, it is also active during communicative acts in which interlocutors are separated by space or time. Processes involved in joint coordination include construal, perspectivation, and framing.

Creative acts of re-construal, re-perspectivation, and re-framing rely on many of the same devices identified as intrinsic to the poetic function of language. More recently, these types of parallelisms have been modeled as resonances in a dialogic syntax, which serve to coordinate intersubjective meaning construction [34]. Several recent studies have demonstrated how such resonances contribute to the CALU. Veale et al., [124] discuss an example of adversarial humor, from an exchange attributed to George Bernard Shaw and Winston Churchill, in terms of resonant parallelism.

Shaw: Am enclosing two tickets to the first night of my new play;
bring a friend ... if you have one.

Churchill: Cannot possibly attend first night; will attend second ... if
there is one.

The resonance created through the reuse of the *conditional-if* construction acts to draw attention to the message itself, rather than the content, by projecting paradigmatic equivalencies onto the syntagmatic process. The communication stream becomes less linear and the communicative process less automatic [49]. Clark [28] proposes that this type of layering is a part of all types of non-referential language use including humor, irony, sarcasm, play, and pretense. The process of layering involves the joint creation of a virtual world where communicative partners can simulate counterfactual events rather than report them. In the exchange between Shaw and Churchill, Shaw initiates a virtual world in which a virtual Churchill is

comprehensively unpopular, and Churchill subsequently populates it with a virtual Shaw who is comprehensively untalented.

Metaphoric language can also be viewed as a device of the CALU for creating resonance through the projection of paradigmatic selection onto syntagmatic combination. Feyaerts [39] cites the following comment from an exit poll interview with a voter during the Irish general election of 2011.

Voter: Well, finally Ireland will get a new captain, but unfortunately it will be to steer the Titanic ...

With the word “captain”, the voter invokes the conventional conceptual metaphor NATION IS A SHIP, which catalyzes the projection of semantic information associated with large oceangoing vessels onto the current topic of the state of the Irish nation in light of the election. Through a process of layering, or conceptual integration, the speaker blends the Irish ship of state to create further resonances. Conceptual integration, or blending, is a theoretical model of how our minds create meaning [37]. In the proposed model, meaning arises through the composition, completion, and elaboration of a series of mental spaces (e.g., virtual representations of the current communicative event including interlocutors and common ground, the current topic, the context framing the current topic, and the emerging construal with respect to the current topic).

Metaphoric language primarily serves to shape our understanding of abstractions [69]. Recent experimental findings underscore the influence of the framing effects of metaphoric language on how people reason about and evaluate abstract socio-political events. Landau et al., [70], for example, found that the framing effects of metaphoric language influence the stance that people adopt toward immigrants and immigration. Thibodeau & Boroditsky [123] found that framing effects influence how people reason about crime. Lichtenstein & Shutova [74] found cross-linguistic framing effects on evaluative judgement of economic change. Thibodeau & Flusberg [122] found framing effects on voting intentions. Matlock [84] provides a review of research on framing effects of metaphor in the political realm. Metaphoric language is integral to comprehending and communicat-

ing knowledge about complex phenomena and, consequently, must be an integral component of knowledge expansion and transmutation through the CALU.

The poetic function, dialogic syntax, conceptual metaphor, and conceptual integration together offer a firm foundation for a theoretical model of the CALU. To develop a computational model of the CALU consistent with these approaches calls for a turn toward research on creative cognition in general.

1.2.4 STAGES IN THE CREATIVE PROCESS

Much of the current work on creative cognition is predicated on Wallas' [129] four-stage model of the creative process, which itself was derived from Poincaré's formulation of creativity. A recent close reading of Wallas' original proposal has resulted in the following five-stage model [98].

1. During the preparation stage, the intellectual resources and technical skills necessary for addressing a particular problem are accumulated. Data is collected and information is accessed. Deliberation, planning, and cognitive control govern successful execution of this stage of creative thought.
2. During the incubation stage, new streams of association are generated through the indiscriminate segmentation and recombination of existing ideas and images. Conscious and deliberate efforts toward solving the problem are avoided. Attention must be diverted to an unrelated activity, or allowed to wander. At this stage of creative thought, unconscious processes govern further progress in the search for a solution.
3. During the intimation stage, percolating streams of association at the fringe of consciousness can become fleetingly amenable to conscious awareness and deliberate influence. Important aspects of this stage are allowing the percolation process to continue naturally without external interference whilst remaining primed for fleeting moments of insight.
4. During the illumination stage, one of the percolating streams of association becomes fully potentiated as the focal point of conscious attention.

5. During the verification stage, the novel solution generated as a result of the prior four stages serves as the point of departure for a further deliberate, conscious, and controlled process of elaboration and verification using established disciplinary methods and techniques.

While the above again represents a theoretical model of creative cognition, a contemporary adaptation of Campbell's blind variation and selective retention model of creative processes represents further progress toward a viable computational model of creativity.

1.2.5 BLIND VARIATION AND SELECTIVE RETENTION

In general, blind variation and selective retention (BVSR) models the achievement of fit and order in interactive adaptive systems. The primary components of BVSR are variation via heterogeneous alterations on an existing form, systematic selection and elimination of the resulting variants, and retention, preservation, and in some cases multiplicative duplication of the selected variants. Absent any one of these components and neither the fit nor the order amongst interacting systems can increase.

The most problematic component of BVSR has been the "blind" variation on an existing form ([30], [32], [43], [108]). Though the term is meant to convey that BVSR is a model of discovery beyond what is known or what can be formulaically produced, it has been misinterpreted as amounting to random variation [23]. There are aspects of the modern statistical formulation of randomness, though, that are nonessential for adaptive BVSR. These include equiprobability of selection amongst variants and unrestrained variability. Statistical independence amongst successive variants can be beneficial in some applications but is also nonessential for adaptive BVSR. The aspect of randomness that is essential for BVSR is independence from the environmental condition for which a variant may prove to be adaptive. In other words, the likelihood of a variant occurring is independent of the environmental condition determining its chances of retention.

BVSR was proposed as a model for creative thought and scientific discovery in

the 1960s, though precursors can be found in ([5], [112], [79], [96], and [56]). Simonton [109] formalized the model based on the specification of novelty, value, and surprise as the three dimensions governing creativity. In Simonton's formalization, the creativity of a variant x_i is defined as $c_i = (1/p_i)u_i(1/v_i)$, where p_i is the variant's initial probability, u_i is the variant's actual utility, and v_i is the degree of prior knowledge of the utility u_i . A variant's sightedness is given by $s_i = p_i u_i v_i$. Blindness as the opposite of sightedness becomes $b_i = 1/s_i$. Highly sighted variants that are highly useful will have high initial probabilities because the high prior knowledge of their high usefulness is already known. Highly blind variants, in contrast, have initial probabilities that are less affected by their actual utility since prior knowledge of it is low. The distribution of Monte Carlo generated variants according to this formalization shows that blindness is positively correlated with creativity.

Evidence consistent with BVSR from the literature on the neuroscience of creativity indicate that creative processes are associated with recurrent activity across two large scale brain networks as mediated by a third. A considerable number of neurophysiological and neuroimaging studies have reported increased coupling between brain regions that are part of the default mode network and those that are a part of the cognitive control network [11]. The key nodes of the control network include the dorsolateral prefrontal cortex and the posterior parietal cortex. These exhibit strong intrinsic functional coupling during tasks that demand tight cognitive control over exogenous information, such as information maintenance and manipulation during tests of working memory, judgment, and decision making [86]. This functionality can be contrasted with that of the default mode network, which becomes engaged during processes not immediately related to or consequent upon external factors, such as mind-wandering [110], autobiographical planning [114], and mental simulation [4]. In concert with this contrasting functionality, the control and default mode networks appear to be intrinsically competitive networks in that activation in one suppresses activity in the other.³

³This latter point coheres well with anecdotal accounts of creative achievement in which a novel approach to a problem arrives unexpectedly during the performance of some unrelated and pos-

The salience network, whose core components are the anterior insula and the anterior cingulate cortex, has been shown to mediate dynamic interactions between other large-scale brain networks [86]. In an analysis of the time course of functional connectivity during a creativity-related task, Beaty et al., [10] found early coupling between hubs of the salience and default mode networks, intermediate coupling across hubs of all three networks, and late coupling between hubs of the salience and control networks. Beaty et al., [12] report similar brain network dynamics underpin the production of novel metaphoric expressions.

Several of the groups investigating the neuroscience of creativity directly link their research to BVSR and relate activity in default mode hubs to periods of blind variant generation ([9], [63], [75], [76]). However, whether blindness is a necessary feature of the creative process remains unclear. The argument against blindness contends that (i) BVSR is modeled on Darwinian natural selection and (ii) creativity and scientific discovery are not phenomena akin to environmental adaptation through natural selection ([30], [32]) and are not suitably modeled as such. In particular, Gabora [43] argues that Darwinian natural selection explains a paradox not faced by creativity researchers: “how change accumulates when acquired traits are not inherited”. Since acquired cumulative change in science, for instance, is preserved transgenerationally, scientific discovery is not driven by nor does it benefit from blind processes.

This claim, however, represents a pronounced reformulation of the type of phenomena that BVSR is designed to model. BVSR concerns itself with questions of fit and order, not accumulation of change. The fit and order that initially inspired Campbell was that between the cognitive activity of scientific research and the environmental phenomena such activity comes to describe and explain. To Camp-

sibly mundane activity. For example, Poincaré [96] provides the following description of discovering the Fuchsian functions: “The incidents of the travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or the other. At the moment when I put my foot on the step, the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already commenced, but I felt a perfect certainty.”

bell, fitness and order can be seen throughout the natural world, and BVSr is his hypothesized undercurrent for this scale invariant property.

May it first be noted that many of us see in crystal formation a chance variation and selective retention process. For example, in a saturated salt solution of intermediate temperature, most of the adjacencies between one salt molecule and another are as easily moved out of as entered into, and the thermal noise, Brownian movement or whatever produces a continuing change of adjacencies. But while this is true of most adjacencies, there are a few which result in a particular fit, in which the force fields of the two molecules summate to produce an adjacency exceptionally hard to dislodge. These particular adjacencies require less energy to enter than to disrupt, and thus, though they are rare, they are selectively retained and accumulated, forming the orderly crystal pattern. In this process the three essentials to the model are present: a system producing variations, a systematic selection of certain variations after they have happened to occur, and a preservation of the variations. Crystal formation is limited to the rare combination of these three requirements. Extreme heat, such as to liquefy or vaporize salt, will increase the variations component, but destroy the retention system by continually producing energy inputs that exceed the disruption threshold. Extreme cold will remove the variation component. It is only when these two are in a compromised balance that the selective-retention negentropic process of crystal formation can take place. This model is also applicable to other levels of order and structure ... from atoms on up.

As in crystal formation, creativity and scientific discovery can only occur in a condition of compromised balance between entropy and negentropy, noisy variation and static invariance. In this dissertation, I investigate the BVSr hypothesis by constructing a model of the CALU that is mathematically equivalent to the spin models used in statistical physics to explain magnetism and other critical point

phenomena. In general, spin models capture the patterns of local pairwise interactions that are predictive of the global propagation of order. Ising models are spin models of systems with exactly two possible states (e.g., up or down). Heisenberg or Potts models are spin models of systems with more than two possible states. The two-state Ising spin model and its generalization to the continuous state Heisenberg or Potts spin model were developed in the framework of statistical physics in order to investigate the organizing principles of the collective behavior of complex physical systems. The Ising spin model developed by Ernst Ising in the 1920s was originally designed to explain the physics of spontaneous magnetization, a type of phase transition. Phase transitions occur when a small change in a parameter such as temperature or pressure results in a large-scale qualitative change in the state of a system. Spin models have a combinatorial interpretation that make them particularly suitable for complex systems. One purpose of spin models in complex systems research is to investigate how short-range interactions between constituents of a system give rise to long-range correlative behavior, and to predict the potential phase transitions. Beyond physical systems, spin models have proven effective in characterizing order in biological systems at various levels of description. Ising models, which formed the basis of Hopfield’s recurrent artificial neural network, have recently been used to examine synchronized firing patterns in real brain networks [101]. The flocking behavior of birds has been described in terms of a Potts spin model [111]. Evolutionary sequence patterns in proteins [73] and in eukaryotic organisms [6] have been modeled as fitness landscapes using Ising and Potts methods.

In a similar vein, sequence variation in orthographic sequencing of four-letter words in English have been modeled as energy landscapes [117]. Words in this approach are networks of interacting letters. Given a large corpus of text (in this study the Jane Austen corpus and a portion of the American National Corpus), an Ising spin model is fit to give the full joint distribution of letter co-occurrence frequencies across the space of possible combinations for the set $P(l_1, l_2, l_3, l_4)$. A further model is inferred from sampling the same corpora to capture a “dictionary” distribution, i.e., one that assigns an equal probability to all words in the corpus. From

these models, an energy landscape $E(l_1, l_2, l_3, l_4) = \sum_{i>j} V_{ij}(l_i, l_j)$ is constructed on the space of possible words, where V_{ij} can be thought of as the interaction energy between two letters. To find the local minima of the landscape, the change in energy is computed for all single letter replacements and unique local minima are identified as the words for which any replacement results in an increase in energy. Ordering these ground states according to decreasing probabilities produces basins, which can then be populated through excitation of the ground states. Interestingly, not only do the models produce accurate profiles of current words in the dictionary, they also discover new words (i.e., words that were not in the training corpora), 1/5 of which are legitimate words (e.g., mast, tome, welt) and the remaining majority of which represent pronounceable sequences of orthographic English (e.g., fent, hove, and liss).

Sequence variation in HIV-1 strains is a mechanism of viral immune evasion that is particularly pernicious in the highly mutable HIV-1 polyprotein. Predicting the fitness landscape of HIV-1 viral mutations is crucial to the development of prophylactic and therapeutic vaccines [38]. The fitness landscape is used extensively in evolutionary and medical biology as a conceptual tool for investigating genotype to adaptive success mapping. It can be thought of as a three dimensional surface in which genotypes are organized on the x-y plane and fitness is plotted on the z axis. Evolution in this framework consists of walks and adaptation is a climb to a higher position on the landscape. A fitness landscape of HIV-1 protein variants will show which mutant strains are most viable.

The HIV-1 protein consists of three dimensional amino acid chains, segments of which can be represented as strings of letters (e.g., QYRLKHVW). Viral mutants occur when an alternative amino acid comes to occupy one or more of the sites in the amino acid chain. A fitness landscape that explores the space of all possible mutant sequences can be constructed by fitting a spin model to existing sequence data. In this approach, mutants are networks of interacting sites in the amino acid chain. The inferred energy of a mutant corresponds to its fitness (the lower the energy of a mutant strain, the higher its fitness [102]). The ability of this method to robustly predict mutational pathways to escape has been demonstrated in multiple

recent studies ([38], [81], [103]).

Spiking activity across a population of neurons can be characterized as a series of ordered sequence vectors. For each neuron in the network considered over a brief window of time, the neuron either spikes or does not. The activity of each neuron i can be recorded for each brief window of time as $\sigma_i = 1$ if the neuron spiked or $\sigma_i = 0$ if it did not. For a network of N neurons, a binary string or “spike word” $\sigma = (\sigma_1, \dots, \sigma_N)$ can be constructed for each brief window of time. A vocabulary of spike words can be collected by recording network activity over multiple windows of time. The probability distribution $P(\sigma)$ over all 2^N spiking states gives the correlation structure of the network and defines the state space available to the network for representing sensations, actions, and concepts. Ising models have been shown to accurately model spiking patterns across large networks of neurons collected from recordings of salamander retinal ganglion cells and the visual cortex of cats and macaque monkeys.

Mitchell et al., present a model that uses semantic feature vectors to predict neural fMRI activation at every voxel in the brain in response to reading the word for a concrete noun [90]. The success of their model in predicting patterns of activation associated with 60 nouns lends support to the distributional hypothesis regarding lexical semantics and the simulation hypothesis regarding cognitive semantics.

1.3 CURRENT CONTRIBUTION

In this dissertation, I implement a dynamical systems model of the CALU by deriving semantic field vectors for an ensemble of nouns, which I treat as spiking patterns across a network of neurons; approximating the probability distribution of states assumed by the network and modeling the energy landscape using Ising and Potts methods; and selecting low probability, high fitness variants from the energy landscape. I developed this approach to modeling the CALU, as opposed to developing an approach based on other statistical methods used in linguistics research such as loglinear/logistic regression or feature-based maximum entropy models, due to it being more consistent with findings in the creative cognition literature.

The approach based on spin glass methods allows my model to describe energy landscapes over the space of all possible interaction terms (i.e., semantic fields in my model). The creative expressions emerge from these landscapes with minimal input from global resources. This is commensurate with a cognitive process that operates with minimal conscious control or awareness, which is the current consensus view in the cognitive science literature of the incubation stage during creative cognitive processes. Other approaches to modeling linguistic data require the specification of interaction terms, often based on linguistic intuitions [82], which corresponds to a cognitive process guided or controlled through conscious awareness. The spin model approach is consequently more ecologically valid for investigating creative cognitive processes. It is also parsimonious in that it has been found to be a robustly accurate approach for modeling other biological, neurobiological, and behavioral phenomena, which points to a mechanism operating across scales of resolution.

The dissertation is organized as follows. In Chapter 2, I derive the semantic field vectors for an ensemble of nouns. Given the centrality of metaphor to the CALU, I populate the ensemble with abstract nouns. First, I create a semantic space of word embeddings for the 2500 most frequent contentful nouns in a large corpus of text, where the features are the verbs with which they are associated in a dependency relationship. Then I use affinity propagation clustering to identify ensembles of abstract nouns related by metaphor, i.e., they tend to be contextualized by the same verb dependencies.⁴ From these, I select the most diverse ensemble for modeling the network of interacting semantic fields. In Chapter 3, I approximate the maximum entropy distribution consistent with pairwise correlations by fitting inverse Potts models to the sequence data using Monte Carlo simulations. I populate energy landscapes through zero-temperature Monte Carlo simulations. I identify metastable states of local minima in the resulting energy landscapes and select from these the low probability states as creative variants. Chapter 4 assesses the output of the models *in silico* and *in feras*. First, I create a new semantic space

⁴In essence, this step is an approximation of the preparation stage of creative activity in that the system is learning how nouns are used in metaphors.

consisting of word embeddings for the attested variants and the creative variants. I use the phrase-based distributional semantic measures of vector length and neighborhood density to develop a creativity metric that can be used to test whether the creative variants are more likely to obtain a higher score according to the metric. Next, I present the results of a human subjects experiment on the effects of creative metaphors on semantic convergence. Chapter 5 contains summaries, conclusions, and proposals for future work. The Appendix provides the code for implementing the models and analyses.

This dissertation contributes to the mathematical and computational modeling of language use and cognition. The major contributions are demonstrating that the CALU can be robustly predicted based simply on contextual variation with no contribution from recursion or any other form of infinitude mapping mechanisms; that a specific aspect of creative cognition, the CALU, can be implemented by a model that approximates a BVSR process; and, that the model achieves optimal performance when the system is poised at criticality. Further contributions include an original examination of the relationship of metaphoric language to semantic field sequences and semantic convergence.

It isn't ideas I'm short of... I've got too many.

Degas

My dear Degas, you can't make a poem with ideas, you make it with words.

Mallarme

2

Casting the Players

2.1 DISTRIBUTIONAL SEMANTIC MODELS

Distributional semantic models of language construct high-dimensional semantic spaces in which words are represented by numerical vectors that encode the lexical environments where they are likely to be found. Words that are similar in meaning tend to occur in similar environments and will consequently be represented by similarly configured numerical vectors. Distributional semantic models thus intrinsically incorporate the distributional hypothesis that linguistic context is an integral aspect of lexical meaning. The distributional hypothesis has its roots in the structural linguistics of, for example, Harris [58] and Firth [40] and in Wittgenstein's philosophy of language [132]. Early attempts to quantify meaning via a distributional approach relied on hand annotation and speaker intuition to characterize the features underscoring lexical meaning. For instance, in the semantic

differential approach, native speakers rate words along a set of dimensions (e.g., *is animate, is concrete, is countable, is divisible, is existent, is imageable* and so forth) using a likert-scale type instrument to collect the measurements [119].

Early computational implementations of distributional semantic models include the hyperspace analog to language model (HAL) [77] and latent semantic analysis (LSA) [71]. HAL derives a distributional semantic model by passing through a large corpus of language word-by-word and counting the number of times each word co-occurs with each of its neighboring words within a context window of a given number of words. The co-occurrence count for each context word is then weighted to be inversely proportional to the distance between the word in focus and the context word. The weighted counts form the distributional semantic vector for the focal word, representing its meaning as the likelihood of it appearing in the same window as each other word in the vocabulary of the corpus. LSA similarly uses co-occurrence counts to construct a high-dimensional semantic vector space but represents word meaning as a word's frequency distribution across documents rather than context windows of words. To account for cross-document word importance, a lexical association function is applied to each word's semantic vector that dampens the strength of associations in proportion to its entropy across documents. Singular value decomposition on the entirety of the semantic space is then performed to reduce its dimensionality. The distributional semantic spaces of HAL and LSA produce accurate models of word context but include little information about word order. Two recent approaches to semantic space construction, the bound encoding of aggregate language environment (BEAGLE) and the random permutation model (RPM), integrate information about word order by creating semantic space from the bottom up starting with randomly generated signal vectors to represent words in a corpus. BEAGLE updates the randomly generated signal vectors using recursive circular convolution to bind increasingly large n-gram chunks. For example, in the phrase *music lifts spirits*, *music* can be represented as a distributional vector with values for *music->lifts*, *music-lifts->spirits*. RPM uses random permutation to update its signal vectors [62]. In the framework of cognitive science and artificial intelligence, language models based on self organizing

maps [64] and simple recurrent networks [36] are also informed by contextual representations.

Distributional semantic models have been found to capture a broad spectrum of semantic phenomena [7] and exhibit interesting parallels to human language acquisition ([52], [134]), semantic priming ([53], [71], [77]), word categorization [72], reading times [53], and judgments of semantic similarity [85] and association [77].

2.2 CLUSTERING BY ASSOCIATION

Building on the distributional hypothesis, Shutova et al., [105] introduce the hypothesis of “clustering by association”. The hypothesis posits that in distributional semantic space clusters of concrete concepts tend to form around words with similar meanings whereas clusters of abstract concepts tend to form around words with similar metaphorically associated source domains. The critical insight of this approach is that, due to the high frequency and high systematicity of metaphor in language, metaphoric projections in addition to semantic similarities structure contour distributional semantic space [107]. This property of distributional semantic space can be leveraged to automatically identify metaphoric expressions in large corpora of language using statistical clustering methods, as demonstrated in a number of studies. Two main approaches to the automatic identification of linguistic metaphors via clustering need to be considered: clustering by sentences and clustering by words.

The first approach operates under the assumption that sentences have a semantic signature. The semantic signature of a sentence is a function of the nouns and verbs in the sentence. The contribution of any given verb to the semantic signature of a sentence, crucially, remains constant across sentences with metaphoric meaning and those with non-metaphoric meaning. Hence, the contribution of the verb *pour* to the semantic signature of Sentence 1 and Sentence 2 is the same.

- (1) She poured bourbon into his cup.

(2) She poured scorn on his idea.

Sentences with a metaphoric meaning can then be partitioned from sentences without a metaphoric meaning by clustering sentences by semantic signature. The difference in the semantic signatures of Sentence 1 and Sentence 2, according to this approach, is due to the difference in the semantics of the nouns. Birke and Sarkar [15] adopt this approach using hand-annotated seed sentences to initiate semantic signature clustering. Semantic signatures are represented in [92] using clusters of related words extracted from WordNet. Strzalkowski et al., [118] begins with data that has been labeled with topical structure identifiers and imageability scores with the expectation that the difference in the semantic signatures of sentences such as Sentence 1 and Sentence 2 arises from *cups* and *ideas* being dissimilar topics and from *cup* having higher imageability than *idea*.¹

Word based approaches to metaphor identification via clustering assume that grammatical relationships between words can serve as a proxy for conceptual links or mappings. Mason [83], for instance, clusters words according to selectional preferences to discover metaphoric expressions. Selectional preference is a type of semantic association whereby a predicate exhibits an associative behavior towards a class of arguments. Metaphor as a violation of selectional preferences [131] continues to be influential in metaphor research computational linguistics [125]. In Sentence 1, for example, *poured scorn* can be identified as a metaphor by recognizing that *scorn* is not a member of the class of preferred arguments for the verb *pour*. It should be noted, however, that this approach counters findings in cognitive science research that violations of selectional preferences interfere with language processing times in human subjects [113], but metaphoric language is processed as quickly and effortlessly as non-metaphoric language [47].

Shutova et al., [106] identify metaphoric expressions by clustering nouns and verbs according to their contexts of use. First, nouns are clustered according to shared associations with verbs through grammatical relations, i.e., nouns that tend to co-occur with the same verbs in the same argument position will be assigned

¹Imageability is an estimate of the ease with which a mental image can be formed in response to a given word.

to the same cluster. Then, verbs are clustered by frequency distribution to create a source domain verb lexicon. Seed expressions consisting of metaphors from hand-annotated text are used to link source (verb) and target (noun) clusters that exhibit metaphoric associations. New metaphorical expressions are identified by searching large corpora for expressions composed of nouns and verbs from metaphorically associated verb and noun clusters.

In order to create the associational semantic field vectors for modeling the CALU, I use a word-based approach to identifying metaphor clusters. Each vector serving as input into the model represents an abstract noun and consists of the nouns frequency distribution across a semantic field of verbs conditioned by grammatical relations. The word-based approach accommodates the representation of noun meaning as a function of its participation in a configuration of semantic fields. The process of creating a set of semantic field vectors for modeling involves four steps: (i) obtain vectors over a diverse set of nouns and verbs from a large corpus of language; (ii) calculate semantic similarities for each noun vector; (iii) identify metaphor ensembles with clustering over the resulting similarity matrix; and, (iv) select an ensemble or set of ensembles from the identified clusters for modeling.

2.2.1 OBTAIN VECTORS FOR CLUSTERING

As data I use the one billion word benchmark corpus [26], which has been proposed as a benchmark corpus to be used for measuring progress in statistical language modeling. It is available as a standard training and test setup to facilitate consistency across language modeling experiments. The data consist of news and news commentary. I preprocess the data using standard lemmatizing and tokenizing procedures, and then parse it for syntactic dependencies using the SpaCy library for NLP in python². From the parsed corpus, I extract all clauses bound by subject-verb and verb-object dependencies. From this dataset, I extract only those clauses that include both a noun and a verb from the 2500 most frequent contentful nouns and verbs in the 100,000 word frequency list from the Corpus of Con-

²<https://spacy.io/>

temporary American English (COCA) [33]. I build the associational semantic field vectors for each of the 2500 nouns by calculating its frequency distribution across the field of 2500 verbs. Table 2.2.1 presents a sample of the resulting raw counts for noun phrase direct objects and prepositional phrase objects. As a final transformation, I adjust the raw counts for term frequency and inverse document frequency.

Table 2.2.1: A sample of the associational semantic field vector raw counts.

| | abandon | abound | absorb | abstain | abuse | accelerate | accrue | accumulate | ache |
|----------|---------|--------|--------|---------|-------|------------|--------|------------|------|
| water | 2.0 | 0.0 | 71.0 | 0.0 | 0.0 | 1.0 | 0.0 | 3.0 | 0.0 |
| war | 17.0 | 0.0 | 0.0 | 0.0 | 0.0 | 2.0 | 0.0 | 0.0 | 0.0 |
| mother | 4.0 | 0.0 | 1.0 | 0.0 | 4.0 | 0.0 | 0.0 | 0.0 | 1.0 |
| father | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| health | 2.0 | 0.0 | 1.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| child | 20.0 | 0.0 | 0.0 | 0.0 | 17.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| body | 2.0 | 1.0 | 2.0 | 0.0 | 0.0 | 1.0 | 0.0 | 19.0 | 0.0 |
| door | 0.0 | 0.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| music | 4.0 | 0.0 | 3.0 | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 |
| car | 23.0 | 0.0 | 1.0 | 0.0 | 1.0 | 4.0 | 0.0 | 0.0 | 0.0 |
| game | 19.0 | 1.0 | 0.0 | 0.0 | 2.0 | 2.0 | 0.0 | 0.0 | 1.0 |
| food | 3.0 | 0.0 | 13.0 | 19.0 | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| care | 0.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| sense | 8.0 | 0.0 | 4.0 | 0.0 | 1.0 | 0.0 | 1.0 | 0.0 | 0.0 |
| death | 1.0 | 1.0 | 1.0 | 0.0 | 3.0 | 5.0 | 0.0 | 0.0 | 0.0 |
| wife | 3.0 | 0.0 | 0.0 | 0.0 | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| oil | 2.0 | 0.0 | 13.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| photo | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| building | 10.0 | 1.0 | 0.0 | 0.0 | 0.0 | 7.0 | 0.0 | 0.0 | 0.0 |
| mind | 1.0 | 0.0 | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

2.3 CLUSTERING WITH AFFINITY PROPAGATION

The experiments reviewed in the first part of this section above employ traditional algorithms for solving the clustering problem such as k-means clustering, hierarchical agglomerative clustering, and spectral clustering. The solutions involve learning a set of centers or means to minimize the sum of squared errors between data points and the centers. The centers can be thought of as a set of prototypes or

exemplars which, in the case of the above algorithms, are limited to some subset of the given data points. In the affinity propagation algorithm developed in [42], each data point is considered as a potential exemplar for every other point. The clustering problem is solved through the exchange of real-value messages between data points.

The messages consist of input measures of similarity between pairs of data points, and the exchange of messages continues until high-quality clusters emerge that minimize the distance between cluster exemplars and their related data points while maximizing the distance between exemplars and their unrelated data points. To find appropriate exemplars, affinity propagation accumulates evidence of *Responsibility* ($R(i, k)$ from constituent i) for how much it endorses candidate exemplar k as its exemplar and evidence of *Availability* ($A(i, k)$ from candidate exemplar k) for how strongly it qualifies for the position of exemplar to constituent i . The larger the sum of *Responsibility* and *Availability* for any k over all potential constituents i , the more likely candidate k will serve as a cluster exemplar [42]:

$$\sum_{i=N}^n R(k) + A(k), \quad (2.1)$$

where N is the initial number of potential constituents and n is the number of constituents at model convergence. The model converges when every constituent i has chosen a candidate k as its exemplar.

Affinity propagation improves on approaches to clustering such as k-centers clustering, k-means clustering, and expectation maximization algorithms that require specification of a predetermined or estimated number of cluster centers at each step, a requirement that can lead to limitations in accuracy and interpretability. Approaches that initially assume a large number of clusters which are then pruned counter these limitations. Other issues can arise under such approaches, however, due to errors associated with random sampling and underinformed pruning decisions. Affinity propagation avoids the limitations of both approaches through simultaneous search for candidate exemplars and gradual cluster identification, which obviates the need for underinformed initializations or cuts. It has proven ef-

fective for clustering on a variegated collection of data types, for example: images of faces, microarrays of genes, representative sentences in text, and cities efficiently accessed by airline travel [35].

I implement an affinity propagation algorithm on the associational semantic field vectors from Section 2.2.1 to identify metaphor ensembles. As noted, affinity propagation operates by passing messages that are measures of similarity. I calculate pairwise cosine similarities for all semantic field vectors as [82]:

$$similarity = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}, \quad (2.2)$$

where A_i and B_i are components of vectors A and B .

Affinity propagation can be implemented as an instance of the max-sum algorithm in a factor graph [42], which is the approach I adopt here. With the similarity matrix as a factor graph, the two following update rules are computed iteratively until convergence. The first rule updates the *Responsibility* of candidate exemplar k based on similarity to constituent i , the *Availability* scores of rival candidate exemplars, and the similarity measures of constituent i to rival candidate exemplars. The *Availability* for all data points is initially set to zero [42],

$$R(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{A(i, k') + s(i, k')\} \quad (2.3)$$

and then updated with the following rule [42].

$$A(i, k) \leftarrow \min \left\{ 0, R(k, k) + \sum_{i' \notin \{i, k\}} \max\{0, R(i', k)\} \right\} \quad (2.4)$$

Self-*Availability* is updated with a different rule, as follows [42].

$$A(k, k) \leftarrow \sum_{i' \notin \{i, k\}} \max\{0, R(i', k)\} \quad (2.5)$$

2.3.1 CLUSTERING RESULTS

A total of 44 clusters were identified by affinity propagation. From these, I select the five clusters most densely populated with metaphorically linked abstract and concrete nouns. Figures 2.3.1 through 2.3.5 show the chosen networks with key components labeled. The associational nature of the link between the abstract and concrete nouns through shared verb semantic fields is consistent with the hypothesis of clustering by association.

Taking only the abstract noun vectors from the selected clusters results in a dataset of five metaphor ensembles representing a total of 120 unique metaphor types. To preserve noun frequency information, I return to the COCA frequency corpus that provided the original 2500 nouns and replicate each noun vector in proportion to its frequency, yielding a total of 3480 metaphor tokens. This dataset of 3480 vectors of length 2500 representing five ensembles of 114 types forms the input to the spin models discussed in Chapter 3.

2.4 SUMMARY

In Chapter 2 of this dissertation I use natural language processing techniques to clean and parse the one billion word language benchmark corpus. From this parsed corpus, I construct a multi-million word corpus of verb phrases with direct object noun phrases and verb phrases with prepositional phrase objects. I create associational semantic field vectors for all the nouns from this corpus. I calculate the pairwise cosine similarities for all of the vectors. Using a clustering technique that improves upon previous approaches, I discover ensembles of metaphoric expressions based on these similarities.

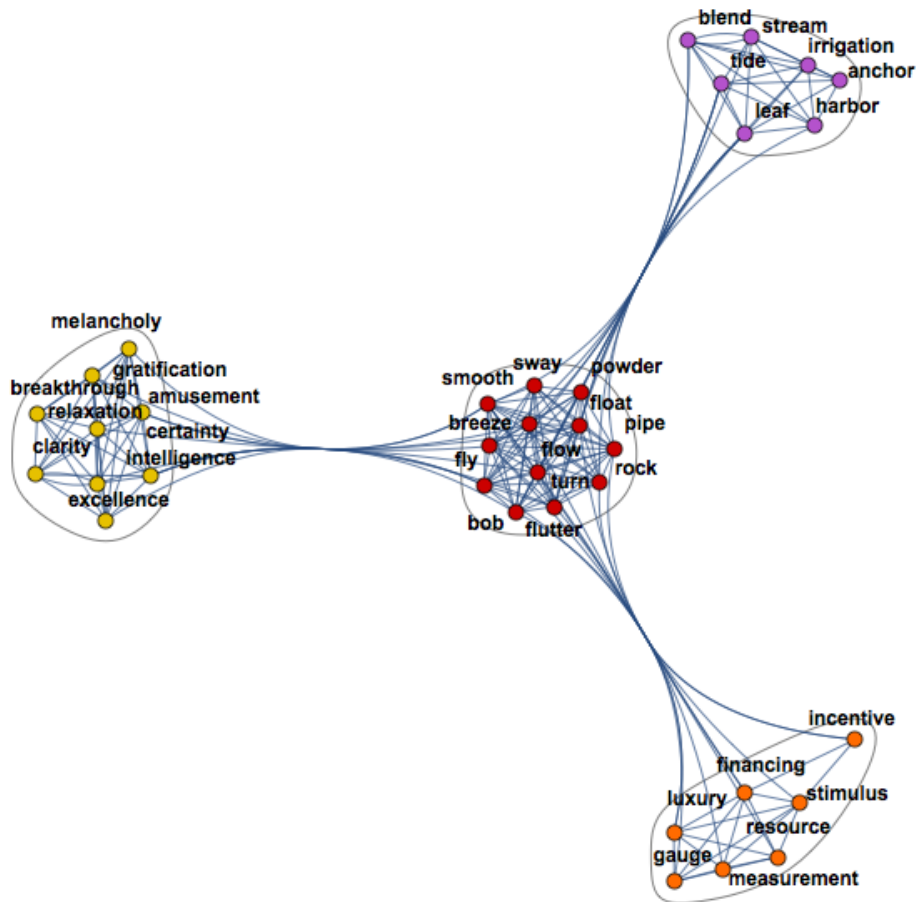


Figure 2.3.1: Cluster 11. Terms for things that flow and float are associated through verb semantic fields with terms used to discuss economics and terms used to discuss intelligence and mild emotions.

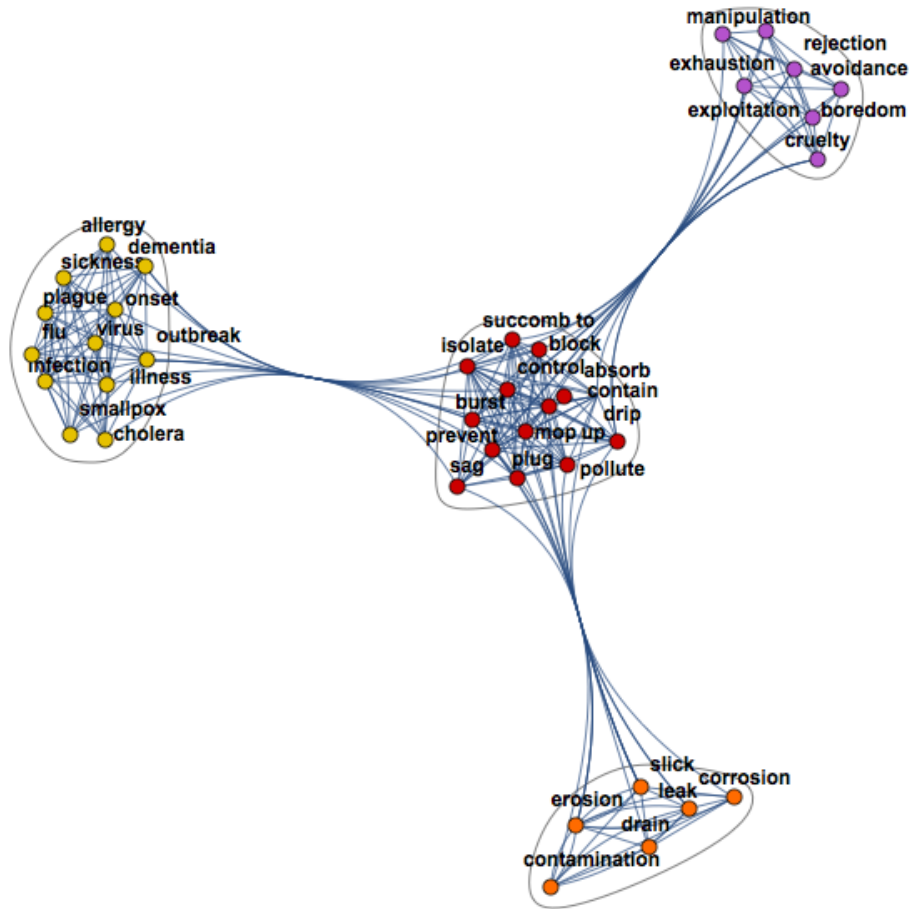


Figure 2.3.2: Cluster 16. Terms for spilling and seeping are associated through verb semantic fields with terms used to discuss illness and terms used to discuss emotions that deplete.

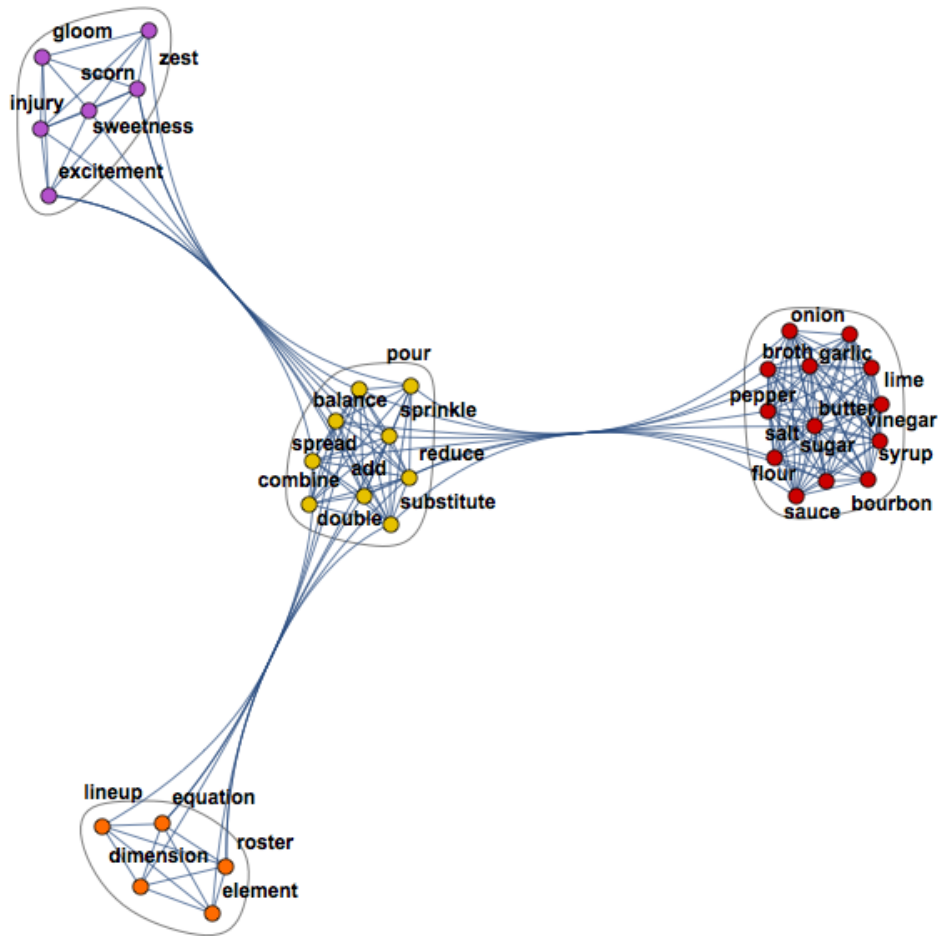


Figure 2.3.3: Cluster 29. Terms for cooking ingredients are associated through verb semantic fields with terms used to discuss math and terms used to discuss communicable emotions.

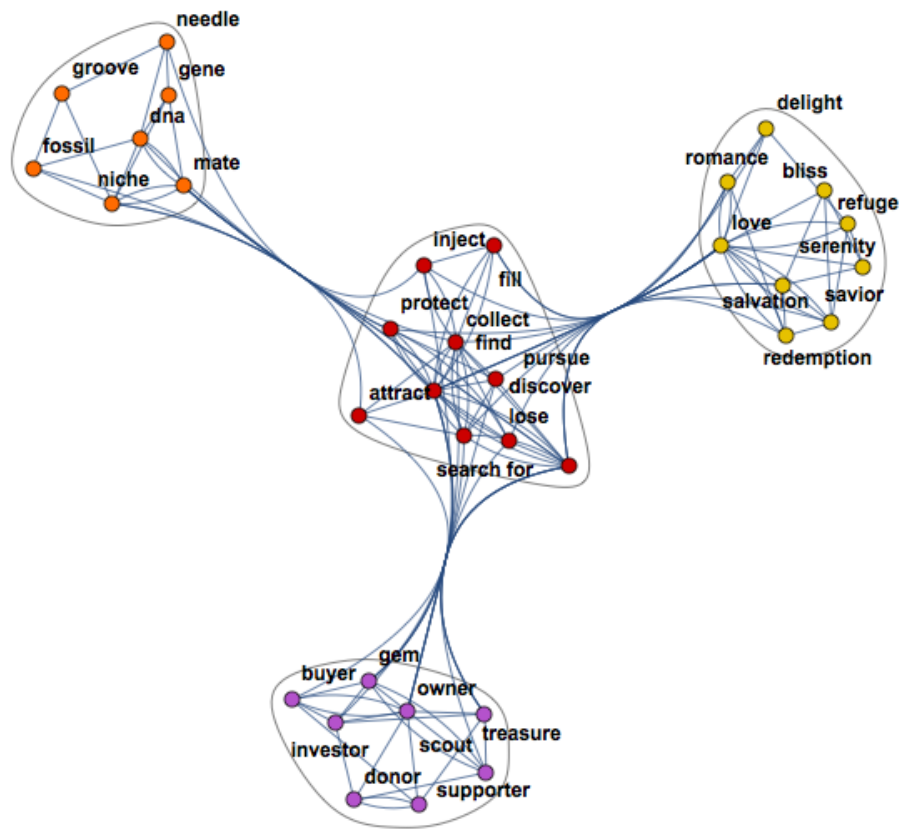


Figure 2.3.4: Cluster 38. Terms for valuable things and people who deal with valuable things are associated through verb semantic fields with terms used to discuss heredity and terms used to discuss eternal emotions.

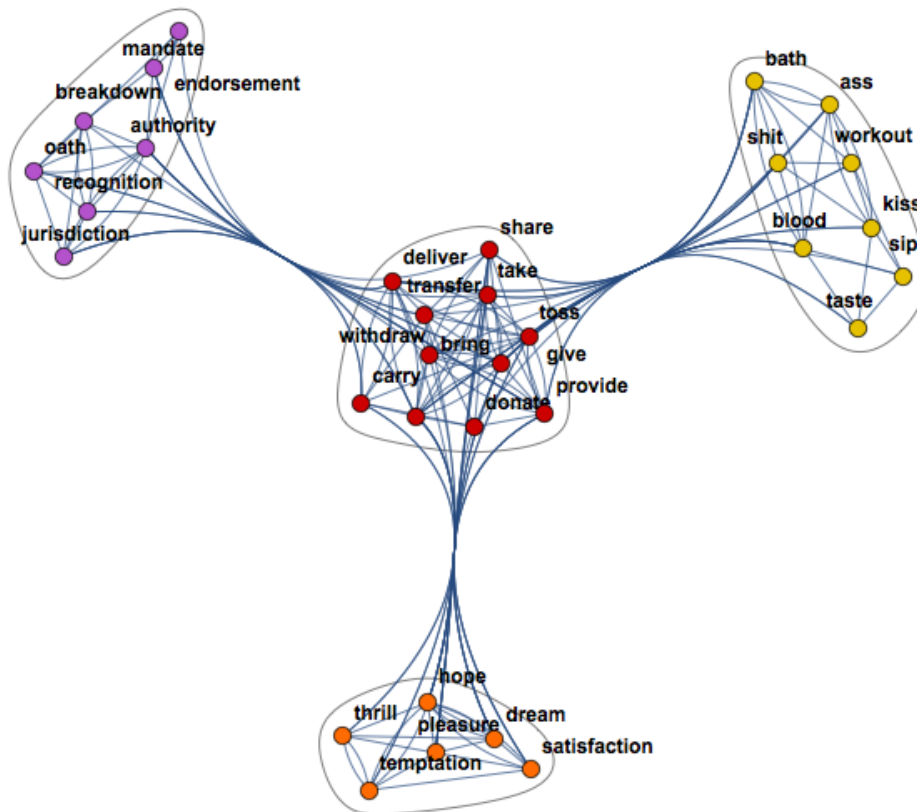


Figure 2.3.5: Cluster 39. Terms for bodily functions and fluids are associated through verb semantic fields with terms used to discuss governance and terms used to discuss emotions of transfer.

We all know that Art is not truth. Art is a lie that makes us realize truth, at least the truth that is given us to understand. The artist must know the manner whereby to convince others of the truthfulness of his lies.

Picasso

3

Spinning the Script

MANY COMPLEX SYSTEMS in nature exhibit an intrinsic order that cannot be recovered by traditional statistical methods. As noted in Chapter 1, research programs across many disciplines have adopted an approach originally developed in statistical physics to model the behavior of spin glasses. The general applicability of this approach arises in that these are the least constrained (i.e., maximum entropy) models capable of reproducing the single variable and pairwise frequencies observed in a set of equilibrium configurations. Figure 3.o.1 depicts a simplified spin system in which an equilibrium state is a stable configuration of spin direction across neighboring sites or nodes.

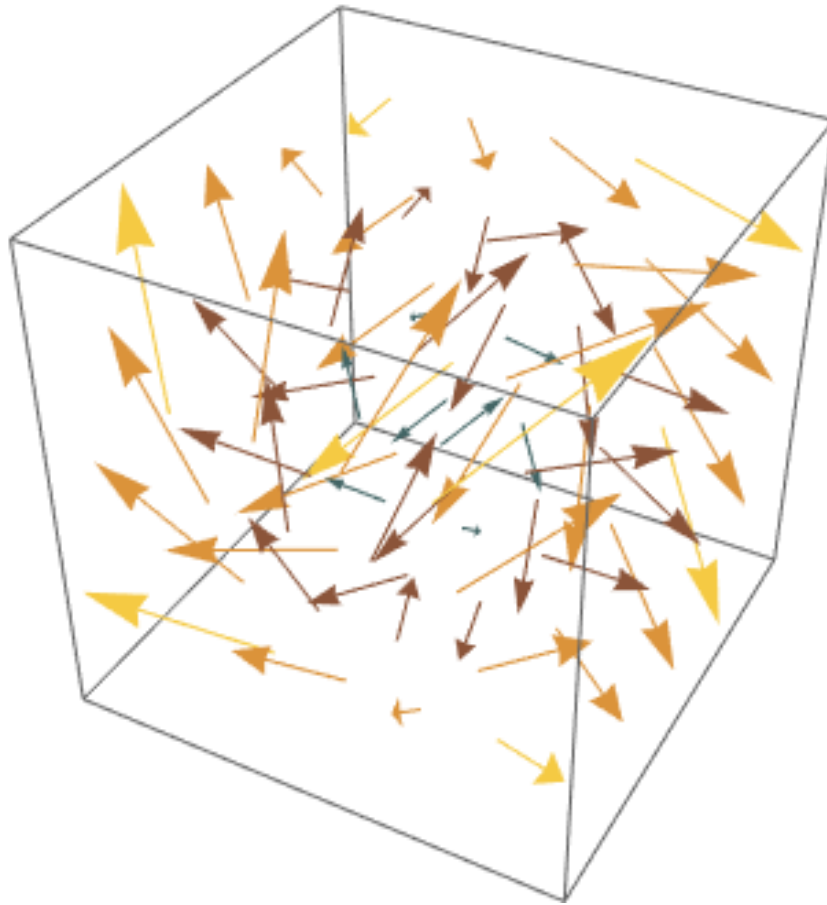


Figure 3.0.1: Spin systems are complex systems that can be modeled as lattices where each node can be in one of any of the possible spin states of the given system. These can be states of magnetization, velocity, chemical composition, or electrical activation (for example). Global characteristics of the system can be modeled as emerging from local interactions of spin states across neighboring sites in equilibrium and non-equilibrium conditions.

Potts spin models find the maximum entropy distribution consistent with single spin frequencies, with pairwise distributions of spins along a sequence, and with a global factor such as temperature that is associated with global states of equilib-

rium. The model takes the form of the Boltzmann distribution [133].

$$P(m) = \frac{1}{\mathbb{Z}} \exp \left[-E(\sigma) \right], \quad (3.1)$$

where \mathbb{Z} , the partition function, is a normalizing constant which ensures that the probability of observing all possible sequences adds up to one and the effective energy of each sequence is [133]:

$$E(\sigma) = -T - \sum_{i=1}^N h_i(\sigma_i) - \sum_{k=1}^K \sum_{i,j} J_k(\sigma_i, \sigma_j). \quad (3.2)$$

For small systems, a series of differential equations can be used to solve the problem of parameter identification exactly, but the method for doing so does not scale to systems beyond $N = 40$ interacting sites [133]. For larger systems, finding a solution reverts to the “inverse Potts problem” or, as it is termed in computer science, Boltzmann machine learning. Inverse Potts models simulate equilibrium states of a system given a vector of site correlations and returns individual energies for each site and coupling energies for interactions across sites.

3.1 NETWORKS OF INTERACTING SEMANTIC FIELDS

The symbol grounding problem has long vexed language scientists and philosophers. Recently it has been posited that language is grounded through the forging of neural firing regularities by routine coupling of experience and brain activity [8]. The wide distribution of this activity across brain network hubs and neuronal ensembles reflects the multifaceted nature of our experience of the environment. Associational semantic field vectors of nouns also exhibit a wide distribution across semantic fields, which in turn has been hypothesized to reflect the grounding of meaning across widely distributed brain networks [90]. Here I model language use in terms of such distributed associational fields and use the model to generate creative metaphoric expressions. The spin at each field site can be thought of as

the proportion of the nodes in the hubs of a network that become active during expression usage. There are in essence an unlimited number of possible spins at each site.

The data being modeled are the associational semantic field vectors of abstract nouns detailed in Chapter 2 comprising (114) types and (3480) attested tokens in the corpus. The vectors consist of $N = 2500$ interacting semantic fields for each expression. For the purpose of testing the validity of this method for modeling language use and controlling for overfitting, the data are split into training (70%) and test (30%) sets. For the purpose of measuring statistical quantities, each semantic field vector is assigned a weight inversely proportional to its observable frequency as follows [14]:

$$\langle \mathbb{O} \rangle = \frac{\sum_{\sigma} w(\sigma) \mathbb{O}(\sigma)}{\sum_{\sigma} w(\sigma)}. \quad (3.3)$$

Based on field correlations, I find the maximum entropy distribution consistent with individual semantic field frequencies ($P_1(\sigma)$), with the pairwise distribution of field frequencies across the sequence ($P_2(\sigma, \sigma'; k)$), and with the observed distribution of lengths of active fields ($P(L)$) in any given sequence. The field energies $h(\sigma)$ represent biases toward activation of some semantic fields over others in the metaphor ensembles included in the data. The exchange couplings J_{ij} describe the interaction among semantic fields across range K . Analogous to temperature in the models of physical systems, $\mu(L)$ in the semantic system investigated by my models acts as a potential for adding semantic structure.

3.2 SOLVING THE MODEL

I solve the inverse Potts problem by combining Monte Carlo simulations with gradient descent, adapting an approach that has been taken in many recent applications [73]. For the set of parameters μ, h, J , the observables $P^{(m)}(L)$, $P_1^{(m)}$, and $P_2^{(m)}$ are estimated by the Metropolis-Hastings algorithm. I assume an initial condition that is independent of field interactions: $\mu(L) = \log P(L)$, $h(\sigma) = \log P_1(\sigma)$, $J_k =$

o and implement the following update rules [44]:

$$\mu(L) \leftarrow \mu(L) + \varepsilon_1 \log \frac{P(L)}{P^{(m)}(L)}, \quad (3.4)$$

$$h(\sigma) \leftarrow h(\sigma) + \varepsilon_1 \log \frac{P_1(\sigma)}{P_1^{(m)}(\sigma)}, \quad (3.5)$$

$$J_k(\sigma, \tau) \leftarrow J_k(\sigma, \tau) + \varepsilon_2 \log \frac{P_2(\sigma, \tau; k)}{P_2^{(m)}(\sigma, \tau; k)}. \quad (3.6)$$

The first two rules are implemented across the first four steps, the third rule is implemented every fifth step, ε_1 is set to 0.005, and ε_2 to 0.01. I experiment with different ranges of interaction energies, including $k = 1, 2, 3, 4$.

3.3 TESTING AND INVESTIGATING THE MODEL

Figure 3.3.1 and Figure 3.3.2 show modeled data versus testing data and observed training data versus testing data, indicating that the model predicts the pairwise correlations with excellent accuracy.

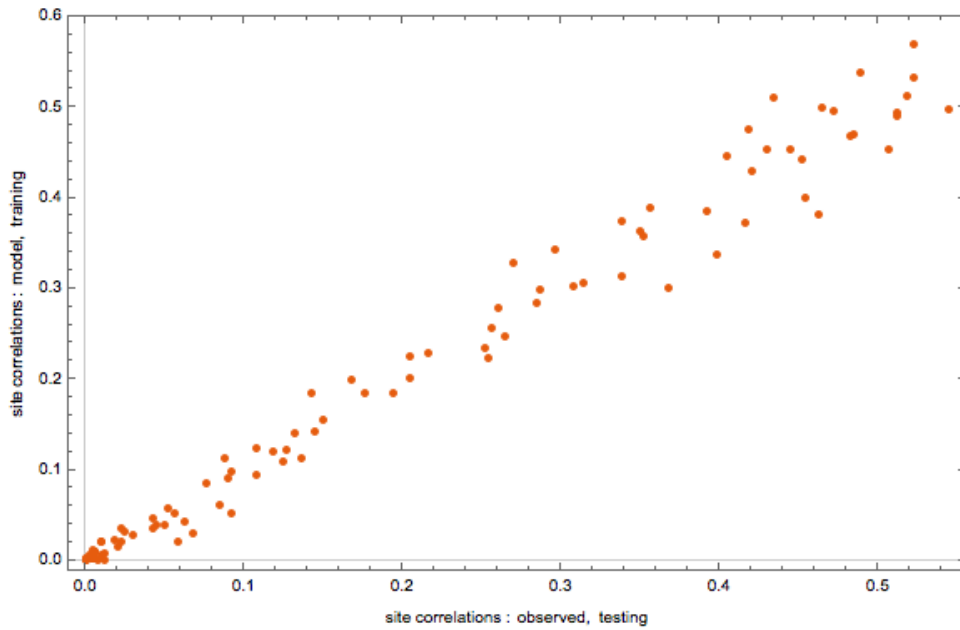


Figure 3.3.1: Probabilities of site correlations of modeled training data plotted against probabilities of site correlations of observed testing data. Correlations are for the $k = 1$ through $k = 4$ nearest semantic fields. The probabilities have been sorted into bins from 0 to 1 where bin-width is 0.01.

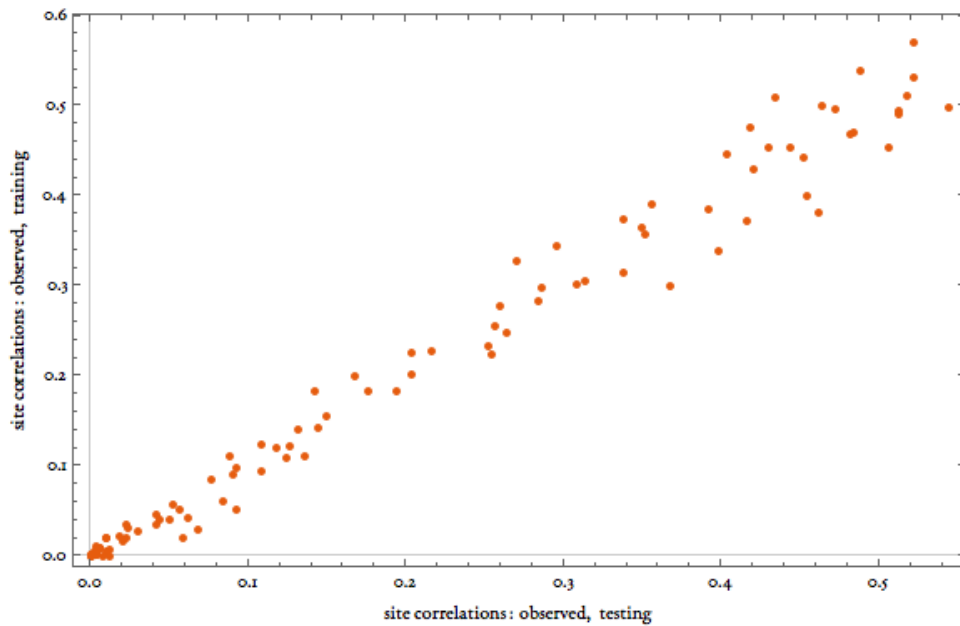


Figure 3.3.2: Probabilities of site correlations of observed training data plotted against probabilities of site correlations of observed testing data. The scatter here is of a commensurate magnitude with that of the model as shown in Figure 3.3.1, which indicates that the model is nearly as accurate as the data allow.

Figure 3.3.3 plots a superimposition of the data from Figure 3.3.2 onto the data from Figure 3.3.1. The plot shows that, while the magnitudes of scatter are commensurate with one another, the magnitude of the modeled data with respect to the observed testing data is greater than that of the observed training data with respect to that of the observed testing data.

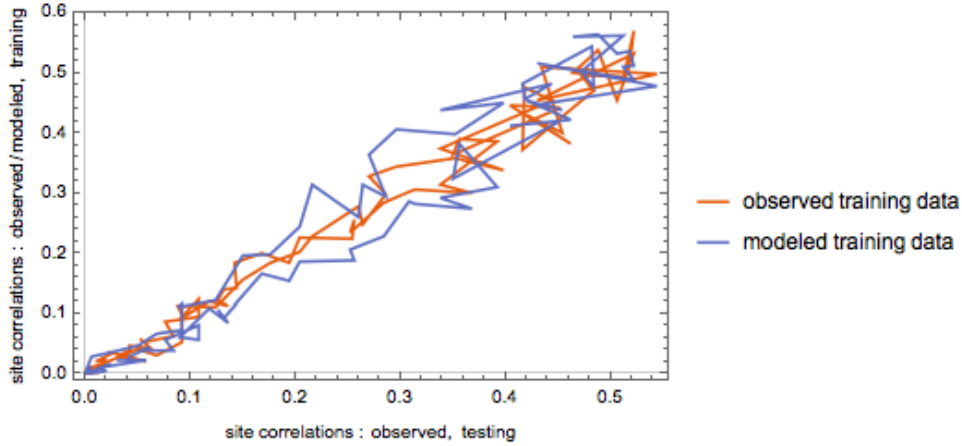


Figure 3.3.3: Comparison of probabilities of site correlations between observed training data and modeled training data plotted against probabilities of site correlations of observed testing data. The comparison indicates that the magnitude of scatter of the modeled data is slightly greater than that of the observed data.

To verify that the pairwise Hamiltonian captures essential features of semantic field interaction, I investigate systematically the model's predictions for measurable regularities that have not been used in determining the model's parameters, namely the probabilities of three-site interactions and of contiguous semantic field activity. I calculate the Jensen-Shannon divergences between the learned and the observed distributions of contiguous semantic fields according to the following definition [82]:

$$JS(P_1, P_2) = \frac{1}{2} \sum P_1(x) \log_2 \frac{P_1(x)}{M(x)} + \frac{1}{2} \sum P_2(x) \log_2 \frac{P_2(x)}{M(x)}, \quad (3.7)$$

where $M(x)$ is the mean distribution. The Jensen-Shannon divergence quantifies the amount of information a single sample x provides about its distribution of origin and can be used to determine the similarity of two distributions. It returns a numerical value between 0 and 1 with 0 corresponding to identity and 1 corresponding to anti-identity. Table 3.3.1 shows the Jensen-Shannon divergences across

interaction ranges explored in the model, as well as the Jensen-Shannon divergence between the training and testing datasets for a gold standard comparison.

| $K = 0$ | $K = 1$ | $K = 2$ | $K = 3$ | $K = 4$ | Testing set |
|---------|---------|---------|---------|---------|-------------|
| 0.413 | 0.252 | 0.081 | 0.076 | 0.070 | 0.016 |

Table 3.3.1: The Jensen-Shannon divergences between model and data.

3.3.1 ZIPF'S LAW AND CRITICALITY OF THE SYSTEM

Zipf's law predicts that the relationship between the frequency and the rank of an item in a self-organizing dynamical system will follow a distribution described by the power law ($P \propto 1/r$). Zipf's law has been used to model various aspects of language [80] and recent efforts to understand the origins of the scaling have been moderately successful [61]. A formalization of Zipf's law employing a mutation-driven genetic algorithm draws attention to the prevalence of Zipfian distributions in processes of adaptation by mutation [78].

A Zipf plot of rank and probability for the actual data and for model predictions in Figure 3.3.4 demonstrates that both sets obey a Zipfian distribution. For the system of semantic fields modeled here, the rank of a particular state or configuration σ is given by the number of states with lower energy as calculated in Eq. 3. The Zipfian distribution in the present case indicates a linear relationship between energy and system entropy, which is characteristic of thermodynamic systems approaching a critical point phase transition [94]. Returning to Campbell's description of BVSR processes, the system is poised between entropy and negentropy, allowing sequence variation while constraining the set of grammatical sequences.

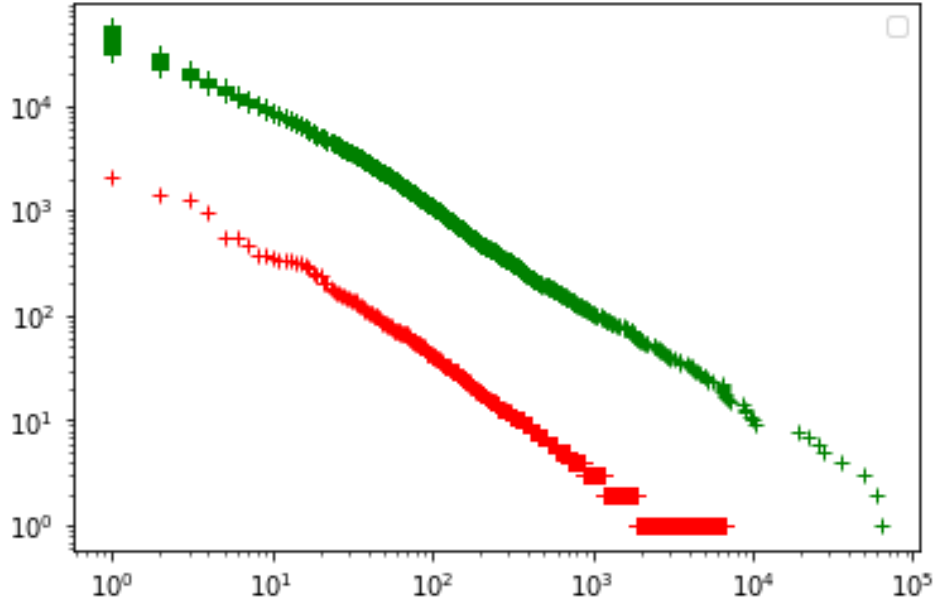


Figure 3.3.4: The observed data (the red line) and the model data (the green line) both follow a Zipfian distribution where FREQUENCY is plotted over RANK. Emergent global behavior where microcanonical entropy is inversely proportional to energy are in proximity to a critical point.

3.3.2 COMPARING METAPHOR ENSEMBLES

Given that the associational semantic field vectors used here for modeling are transformations of those used for clustering, we expect sequence overlap within metaphor ensembles. To investigate with more specificity sequence overlap within ensembles and to gain insight into sequence overlap between ensembles, I calculate the mutual information within and across ensembles. Mutual information measures the degree of correlation between two sites as follows [14]:

$$I_{ij} = \sum_{\sigma, \sigma'} P_{ij}(\sigma, \sigma') \log_2 \frac{P_{ij}(\sigma, \sigma')}{P_i(\sigma) P_j(\sigma')}, \quad (3.8)$$

where $P_i(\sigma)$ is the probability of having spin σ at position i and $P_{ij}(\sigma, \sigma')$ is the probability of having spin σ at position i and spin σ' at position j . Figures 3.3.5

and 3.3.6 show that mutual information is much weaker when measured across all ensembles than when it is measured within a single ensemble. This indicates that, given a single ensemble member, more is known about other members within the ensemble but not necessarily about members of other ensembles.

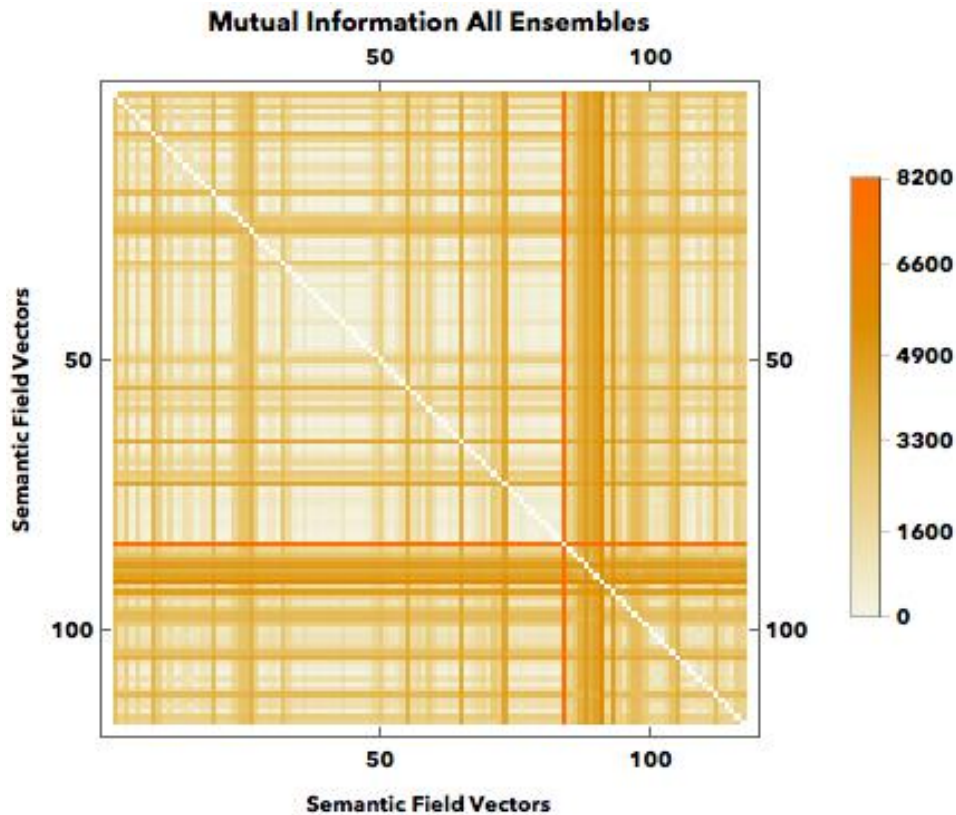


Figure 3.3.5: Mutual information across semantic field sites is relatively weak when calculated across metaphor ensembles. With regard to the *CALU* and creative cognitive processes, this has implications relevant to the theories that emphasize the importance of divergent thinking for creative ideation. If it could be shown that creative metaphors were more likely to be found by including more numerous and more diverse metaphor ensembles, such a finding would support those theories.

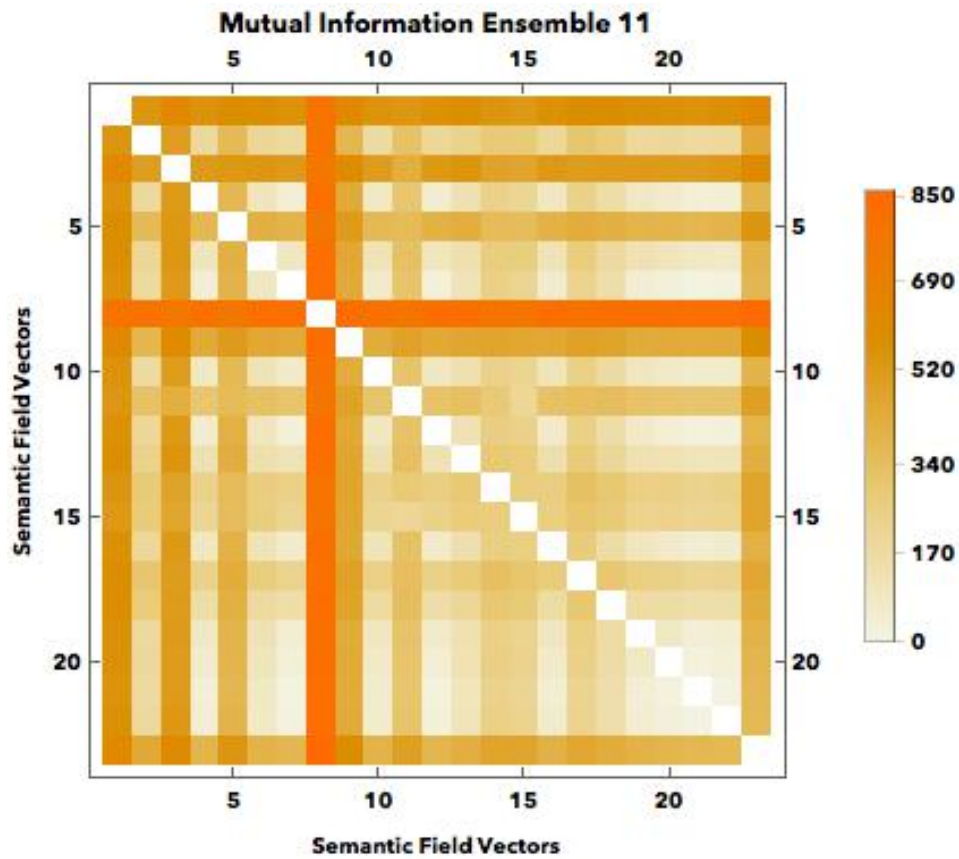


Figure 3.3.6: Within single metaphor ensembles, the mutual information between semantic fields strengthens.

3.4 LANDSCAPING WITH BASINS OF ATTRACTION

Growing entropy deficits between an independent model of a complex system and models taking into account increasing levels of interaction suggests “frustration” due to the existence of many metastable states. The entropy deficit, or multi-information, of a complex system is defined as [14]:

$$I(N) = S_o(N) - S(N), \quad (3.9)$$

where the first operator in the equation represents the entropy of the system

without accounting for local interactions and the second that of the system when interactions are accounted for. Figure 3.4.1 illustrates that the multi-information of the system of interacting semantic fields grows in proportion to the number of pairs at a rate of roughly N^2 .

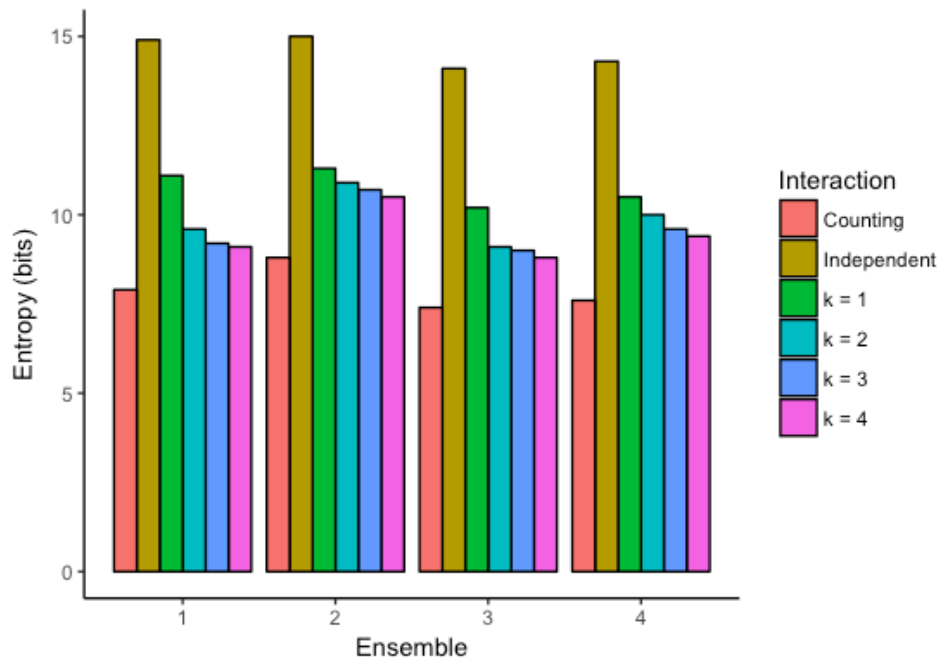


Figure 3.4.1: The entropy of all the metaphor ensembles from frequency counting, from the independent model, and all K interacting semantic field models indicate that the local interactions are having a strong effect on the global features of the system.

In an energy landscape, a metastable state is a local minimum or basin of attraction. I identify basins of attraction in the model's energy landscape using zero-temperature Monte Carlo sampling, which partitions the spin configurations into correlated configurations grounded by one for which a spin flip at any single site results in an increase in energy. For this, a triangular lattice of spin configurations is constructed and the following algorithm applied [44]:

- pick a random site i ;

- check if flipping spin σ_i decreases the energy of the system;
- if flipping spin σ_i decreases the energy of the system, flip it;
- otherwise, do not flip it, move on to the next adjacent site.

The Potts models of semantic field vectors finds 31 basins of attraction. To find whether any of these minima represent an unattested expression (or one unseen by the model), I compare the vectors at the minima to those in the training set. Any sequence of semantic field spins appearing in the model but not in the training set is equivalent to an unattested expression. However, since the abstract expressions are defined by their sequence of semantic field spins, another step must be taken to recover which abstract expression is associated with this mutant (or deviant) sequence of semantic fields. This can be accomplished by gradually adding heat or energy to the system to find the next lowest energy state for which a single flip at any single site results in a match to an attested vector. The abstract noun represented by the attested vector becomes the noun in a creative metaphoric expression and the semantic fields exhibiting deviant activity in the model's mutant minima become the verbs.

To make this process more concrete, consider a basin of attraction grounded by an unattested vector that a minimum increase in energy resolves to the semantic field vector for the abstract word *bliss*. The sites whose spins had to be flipped to achieve this resolution were $\sigma_1, \sigma_{62}, \sigma_{77}$ corresponding to the semantic fields of BOB, FRAGMENT, RUPTURE. The resulting compositional expressions can be construed as verb phrase with noun-headed direct object phrases or preposition headed indirect object phrases, such as – *bob in bliss, fragment bliss, rupture bliss* represent the creative metaphors generated by the Potts spin model. Figure 3.4.2 presents the complete set.

| Intelligence | | Scorn | | | | | Bliss | | | |
|--------------|----------------|--------------|-------------|-------------|--------------|--------------|---------------|-------------|-------------|------------|
| arrange | live in | absent | consider | fight | keep | quit | adopt | cure | overcome by | spark |
| avoid | move | acknowledge | contain | fill with | lay | raise | agree to | deal with | pick | speak of |
| bypass | oppose | | | | | | alter | declare | pinpoint | speculate |
| carry | participate in | add | continue | find | lead | recover from | announce | deserve | play | spread |
| change | preserve | assemble | contribute | fit | leave | reflect | appear | destroy | pour | start |
| close | reel | beat | convey | fix | lift | regain | apply | develop | pray for | stick |
| combat | remove | begin | cope with | flavor with | lose | rehabilitate | ask | devise | promise | support |
| comfort with | resource | blame | counter | form | manage | rejoin | attract | disclose | prove | switch |
| commit to | retire | | | | | | break | discover | pursue | tap into |
| complement | return | boast of | cover | generate | match | remove | burst | don | put | target |
| complete | ride | bolster | crack | give | meet | repair | buy | fall in | question | teach |
| draw | sacrifice | boost | create | grate | miss | replace | capitalize on | feel | raise | tell of |
| eliminate | spread | break | cure | heal | need | require | carry | hide | receive | treat |
| extend | stabilize | | | | | | cash in on | hold | rediscover | turn |
| finance | start | build | cut | hit | nurse | resolve | celebrate | indulge | reflect | use |
| forgo | stay | capture | dampen | hurt by | offer | rest | choose | investigate | rekindle | wait for |
| guarantee | stop | carry | develop | identify | pay | restore | comment on | join | release | welcome |
| issue | wait for | change | ease | improve | pick | return | conceal | keep | reveal | win |
| leave | wear for | check | emerge from | include | place | revamp | confirm | mate | run | work |
| limit | wind | choose | enhance | increase | play | run | consider | meet | search for | write with |
| line | | combat | enter into | inflict | prevent | save | contain | move | serve | |
| | | combine | evaluate | inject | produce | shake | cope with | need | shower with | |
| | | compensate | feature | introduce | promote | spark | create | offer | sing of | |
| | | compete with | field | join | protect from | stir | | | | |

Figure 3.4.2: Creative metaphors can be retrieved by percolating the system to recover the attested vectors in closest proximity in energy space to the unattested vectors. Discrepancies in the active semantic fields yield creative metaphors.

3.5 DISCUSSION & SUMMARY

In Chapter 3, I fit an inverse Potts spin-glass model to a linguistic dataset and use zero-temperature Monte Carlo sampling to simulate a blind variation and selective retention process. Boltzmann machine learning used for the former correspond relatively closely to processes of blind variation while the latter is a form of selective retention. The dataset used for modeling represented 70% of my original dataset allowing me to test how well the model performed in capturing essential features of the data. The Jensen-Shannon divergences in Section 3.3 indicate that the Potts spin model captures the structure of semantic field system very well, and that the model performs with excellent accuracy as the number of interactions accounted

for increases. An analysis of data distribution reveals the data are in a Zipfian distribution, which in a maximum entropy energy landscape amounts to an inverse linear relationship between microcanonical entropy and energy. Complex systems exhibiting such a relationship are said to be poised at criticality. The finding that the system is in such a position hearkens back to Campbell's original conceptualization of BVSR processes, linking the approach developed in this dissertation directly to other active areas of research on creative cognition.

An analysis of the mutual information within and across metaphor ensembles shows that amount of mutual information within ensembles is high, as is to be expected, but does not remain high across ensembles. This amounts to a deficit in the overlap of semantic field information contributing to the meaning of metaphors in different ensembles. The deficit translates into a benefit for the purposes of generating creative metaphors since it represents a larger territory for expansion in semantic space, which in turn loops back into definitions of creativity in the literature that require value and use. Expansion in semantic space is both valuable and useful.

That the model increases in accuracy as the number of site interactions included increases suggests the existence of multiple metastable states. Metastable states are low-energy states exhibiting what Campbell referred to as “adjacenc[ies] ... hard to dislodge” [22]. They are equivalent to basins of attraction. By identifying all such basins of attraction in the model, I am able to discover unattested semantic field vectors that can be transformed into compositional constituents of creative metaphoric expressions.

Creativity is more than just being different. Anybody can plan weird; that's easy. What's hard is to be as simple as Bach. Making the simple, awesomely simple, that's creativity.

Mingus

4

Gauging the Performance

A HALLMARK of human language is its capacity to facilitate knowledge creation through verbal innovation. Novel combinations of words and morphemes can provide a new perspective on stagnant or unclear concepts. Figurative language, in particular, allows for the dynamic exploration of one parcel of semantic space through its integration with select structure from a separate and distinct parcel of semantic space. In Chapter 3, I develop a model that extends this dynamic exploration to include new territories of semantic space in the exploration process. In the present Chapter, I assess the in-the-wild success of this extension through computational and human subject experiments.

4.1 SP[I]CALU IN SILICO

Distinguishing between unattested but acceptable linguistic expressions and unattested and semantically deviant linguistic expressions is an area of research related to investigations of the CALU. A large dataset of acceptability judgments on unattested adjective-noun compositional expressions demonstrates a clear convergence of individual intuition on what is and what is not an acceptable adjective-noun pairing [126]. Within the same study, it was also found that a number of distributional semantic measures accurately predict the intuitions of the human subjects. I base my computational experiments testing the validity of my model’s output on two of these measures: semantic vector length and the density of the territory surrounding the area occupied by the novel metaphor’s vector in semantic space. To use these measures, it is necessary for me to populate a new semantic space based on a neural language model.

The classic neural language model proposed by [13] consists of a one-hidden layer feed-forward neural network that predicts the next word in a sequence. The product of the model is a word embedding, a real-valued word feature vector in \mathbb{R} . The general architecture of this model includes an embedding layer that generates word embeddings by multiplying an index vector with a word embedding matrix, intermediate fully-connected layers that apply a non-linearity to the concatenation of word embeddings of the n -previous words, and a softmax layer giving a probability distribution over words in the vocabulary.

The neural language model I use to populate the distributional semantic space here is based on the skip-gram architecture proposed in [87]. Instead of predicting a focal word given a set of context words, skip-gram predicts the set of context words given the focal word. Skip-gram sums the log probabilities of the n -word contexts to the left and to the right of the focal word w_f with the objective function [87]:

$$J_{\theta} = \frac{1}{F} \sum_{f=1}^F \sum_{-n \leq j \leq n} \log p(w_{f+j} | w_f). \quad (4.1)$$

Predicting on the focal word removes the need for a hidden layer since intermediate state vector is the word embedding v_{w_f} of the input word w_f [87]:

$$P(w_f + j | w_f) = \frac{\exp(v_{w_f}^T v'_{w_{f+j}})}{\sum_{w_i \in V} \exp(v_{w_f}^T v'_{w_i})}. \quad (4.2)$$

From the neural language model, I create compositional semantic vectors of the creative metaphoric expressions generated by the model. I use dilation for this purpose, which is defined as [87]:

$$p = (u \cdot u)v + (\lambda - 1)(u \cdot v)u. \quad (4.3)$$

The joint semantic vector is composed by stretching vector \mathbf{v} by a factor of λ in the direction of the of vector \mathbf{u} . The resulting compositional vector is expected to correspond more closely to \mathbf{u} than to \mathbf{v} , but some of the semantic components of \mathbf{v} are expected to have seeped into those of \mathbf{u} [89]. For my experiments here, I consider \mathbf{u} to be the noun in the metaphoric expression and \mathbf{v} to be the verb.

I first assess the compositional vectors in terms of vector length. Vectors with very few associations in distributional semantic space are less likely to be meaningful than those that have a greater number of connections. In terms of creativity, a less meaningful expression is less likely to be intelligible and consequently be less valuable.

I next consider the neighborhood density of the compositional vectors [126]. Neighborhood density is measured as the average of the cosines of the compositional vector and its nearest neighbors. For the purpose of assessing creativity, I consider less dense neighborhoods to correspond to more highly creative metaphors. Since a dense neighborhood suggests an area of semantic space that is already well covered, introducing a new occupant provides little added value.

4.2 SP[**I**]CALU IN FERAS

A metaphorical expression such as *unleash anger*, for instance, affords the inferences that anger can be wildly destructive and is best kept under restraint.

Throughout the world's languages, these naturally occurring linguistic metaphors exhibit a pronounced tendency towards conventionalization. Via the process of conventionalization, particular lexical terms from a given semantic domain are used much more frequently to invoke a conceptual integration through metaphor than other equally appropriate terms [100]. English speakers, for example, are more likely to use the term *illuminate* when discussing an idea than they are to use an equivalent term such as *brighten*. The term *illuminate* also exhibits greater metaphoric productivity, leading to expressions such as *illuminate grievances* and *illuminate the causes*.¹

Metaphor is more than an isolated instance of linguistic flare, though, as several decades worth of accumulated evidence can attest [48]. Much of this research stems from Lakoff & Johnson's [69] argument that metaphor is a fundamental cognitive process made overt through the generative capacity of language in the coherent and systematic use of lexical items. From this perspective, metaphor is a system of concepts, which incorporates productive schemata for the elaboration of conceptual domains from multiple vantage points.

These conceptual metaphors also exhibit patterns of conventionalization. Many of these patterns of conventionalization correspond to spatial relationships in our physical world. Metaphors for various sorts of hierarchical systems, for example, invoke a conceptual topology whereby "higher" outranks "lower". This conceptual topology reflects physical experience that is shared across cultures (e.g., the force of gravity). Consequently, many such conceptual metaphors are consistently found cross-linguistically [65]. However, the vocabulary conventionally used to express shared conceptual metaphors frequently diverge across cultures and languages [116]. Here, I exploit this kind of divergence to establish a metric for quantifying the intelligibility of creative linguistic metaphors.

The chapter is structured as follows. Section 2 launches the discussion of metaphor intelligibility with a case study identifying differences in conventionalization pat-

¹For instance, in the Corpus of Contemporary American English (COCA), *illuminate* occurs in seven different types of metaphoric expressions whereas *brighten* occurs in only four different types; and, *illuminate* is three more times likely to occur in a metaphoric expression than *brighten*.

terns in English and Russian. This case study provides a platform for the experimental paradigm for measuring metaphor intelligibility, the development of which is presented in Section 4. Section 5 presents an analysis of the data from three implementations of the paradigm. In Section 6, I introduce relevant related studies on the topic of frequency effects and metaphor comprehension, which serve to illuminate the present findings. The chapter concludes with Section 6, which returns the research covered in this chapter to its position in the context of modelling creative metaphor generation.

4.3 CROSS-LINGUISTIC SEMANTIC DISSONANCE

One of the fundamental challenges in translating foreign text into clear and accessible prose is to render figurative language such that it accurately reflects the intended meaning of the source text and expresses that meaning aptly in the target language. This is not a trivial task, in large part due to language-specific conventionalization patterns and the apparently narrow margin for error. A translation equivalent of the foreign figurative expression often results in a rare or unattested phrase in the target language, leading to a type of semantic dissonance. Semantic dissonance in this context refers to a situation where two content sources share overlapping structural similarities but diverge in terms of usage patterns. Semantic dissonance can interrupt the flow of comprehension as a reader or listener searches for an interpretation of an unexpected or unfamiliar sequence of expressions.

Consider the following similarities and differences between English and Russian in the linguistic expressions used to discuss the concept of ATTENTION/ВНИМАНИЕ. Both languages share a generic metaphor whereby understanding of the more abstract domain of PERCEPTION/ВОСПРИЯТИЕ is grounded in terms of the less abstract domain of PHYSICAL CONTACT/ФИЗИЧЕСКИЙ КОНТАКТ. Lakoff [69] refers to this general system of metaphors as PERCEIVING IS CONTACT BETWEEN PERCEIVER AND PERCEIVED. This mapping manifests itself linguistically in English with examples such as the following.

- (3) When Ruth Harkness became the first foreigner to *capture* a live *panda*

eight years later, she named it Su Lin.

- (4) Last week, Hungarian artist Dandolf *captured* the world's *attention* when he released an appallingly difficult puzzle featuring a hidden panda.

The italicized expressions in sentences (3) and (4) illustrate that the same lexeme (*capture*) can be used in English to characterize events of actual physical contact as well as events in which the physical contact is completely fictive. In Russian, the lexeme (*поймать*) can also occur when referring to both types of events, as seen in sentences (5) and (6) below:

- (5) Городские службы и охотнадзор пытались *поймать животное*, но хитрый лис ускользнул. (tr. Urban services and wildlife management attempted to capture the animal, but the wily fox slipped away.)
- (6) Но мы знаем, что Державину очень важно *поймать внимание* императрицы и как-то реабилитировать себя в ее глазах. (tr. But we know that for Derzhavin it was very important to capture the attention of the Empress and somehow rehabilitate himself in her eyes.)

The metaphors in sentences (5) and (6) can be viewed as special instances of the general metaphor system referenced above, in which the generic source domain of physical contact is realized through the more specific domain of physical captivity, and the generic target domain of perception through the more specific domain of attention. Moreover, sentences (3) through (6) show that the English lexeme *capture* and the Russian lexeme *поймать* exhibit a degree of semantic consonance in that they can both occur (a) when the topic is the more concrete domain of physical captivity, as well as (b) when the topic is the more abstract domain of attention. Given this semantic consonance, translations into Russian of English sentences such as (3) and (4) should yield consistently comprehensible meanings for Russian speakers who know no English. A similar expectation persists for translations into English of Russian sentences such as (5) and (6) for English speakers who know no Russian.

Such semantic consonance is not always exhibited, however, due to variability in conventionalization patterns across the two languages. For instance, in Russian the conceptual link between ФИЗИЧЕСКАЯ ПЛЕНА (PHYSICAL CAPTIVITY) and ВНИМАНИЕ (ATTENTION) is also realized linguistically through the lexeme *приковать*.

- (7) Племянница маршала СССР *приковала себя* к ограде Кремля в знак протеста против захвата ее жилья. (tr. The niece of a Marshall of the USSR chained herself to the Kremlin fence in protest against the seizure of her residence.)
- (8) Сильная красная линия в политических предпочтениях Омской области может *приковать* к региону *внимание* перед федеральным избирательным циклом. (tr. The strong red line in political preferences in Omsk can chain the attention of the national United Russia party to the region in advance of the federal election cycle.)

The italicized expressions in sentences (7) and (8) demonstrate that in Russian the lexeme *приковать* is semantically plausible in the context of events involving physical as well as attentional constraints, in parallel with the lexeme *поймать* as exemplified in sentences (7) and (8). The cross-linguistic semantic consonance of sentences (1) through (4) disintegrates in this case. That is to say, to translate the Russian sentences (7) and (8) into plausible English sentences two different lexemes must be used.

- (9) The niece of a Marshall of the USSR *chained herself* to the Kremlin fence in protest against the seizure of her residence.
- (10) The strong red line in political preferences in Omsk can *rivet the attention* of the national United Russia party to the region in advance of the federal election cycle.

In this instance, the underlying conceptual metaphor is the same for speakers of both languages, but the linguistic experience is not. The expression *chain the attention* is not a felicitous metaphor in English not because speakers of English lack the

conceptual infrastructure supporting an integration between the source domain of PHYSICAL CAPTIVITY and the target domain of ATTENTION. It is a plausible expression given that it has been conventionalized in another related language, and grammatically it is a perfectly acceptable English construction. Instead, it is not a felicitous metaphor in that it is unattested in English. This type of infelicity raises important questions about meaning, language productivity, and language processing. In the current study, I address in particular the question of whether such novel expressions are meaningful or intelligible – i.e., do native speakers converge on a specific interpretation.

In the experimental paradigm described in 4.4 below, multiple such cases of semantic dissonance are exploited to examine novel metaphor intelligibility. In developing the experimental paradigm, instantiations of semantic dissonance in naturally produced English and Russian texts were identified. This process included the hand annotation of Russian and English texts and the examination of output from automated systems for the identification of metaphor in Russian and English, as detailed in [106].

Several of the metaphoric expressions used in identifying examples of cross-linguistic differences come from hand annotation of text. The set of English hand-curated expressions are taken from [104]. The Russian expressions derive from hand annotation of a diverse collection of naturally occurring Russian texts by the author and a native Russian speaker who is also a metaphor researcher. Hand annotation of the Russian texts followed the procedure developed in [115] and outlined below.

Annotation procedure:

1. For each verb in a sentence, establish its meaning in context and compare it to its basic meaning when used in other contexts. As defined in the metaphor identification process of the Pragglejaz Group [115], basic meanings are:
 - (a) more concrete, i.e., easier to access via mental imagery
 - (b) related to sensory motor experience
 - (c) more precise

(d) historically older

2. If the verb has a more basic meaning then the meaning in the present context, mark the verb as metaphoric.
3. For each argument of a metaphoric verb, identify the target domain as represented by the argument and the source domain as represented by the verb using a subset of categories from the Master Metaphor List [115].
4. Label the verb-argument pair with a source-target mapping.

Consider the following sentence.

- (11) Грусть и злость охватывает всю вашу душу, когда великую идею, вами давно уже и свято чтимую, подхватят неумелые и вытащат к таким же дуракам на улицу. (tr. Sadness and anger covers your entire soul when the inept pick up a great idea, long and sacredly revered by you, and pull it out onto the streets for other such fools.)

In this sentence, the meaning of the three underlined verbs must be established and compared to potentially more basic meanings. For instance, the verb *охватывает* does have a more basic sense of “to be spread over or on top of (something)”. In the context of the sentence, the materials being spread over something are EMOTIONS rather than PHYSICAL MATERIALS. Consequently, the expression *Грусть и злость охватывает* would be annotated as a metaphor with an interconceptual mapping of EMOTIONS: PHYSICAL MATERIALS.

The system for automatic identification of metaphor is based on the hypothesis of clustering by association detailed in Chapter 2. Metaphorical patterns are learned by means of hard clustering of verbs and nouns at one level of generality employing a spectral clustering algorithm, that has proven to be effective in lexical acquisition tasks and is suitable for high-dimensional data [120]. The identification of metaphoric expressions is boosted from a small number of linguistic example seed expressions. The seed expressions are in the form of verb:subject and verb:direct object constructions in which the verb metaphorically frames the

noun, as in the above examples. The seed expressions establish a link between the verb cluster that contains source domain vocabulary and the noun cluster that contains diverse target concepts associated with that source domain. This link allows the system to identify a large number of new metaphorical expressions in a text corpus. In summary, the system (i) performs noun clustering in order to harvest target concepts associated with the same source domain; (ii) creates a source domain verb lexicon by means of verb clustering; (iii) uses seed expressions to connect source (verb) and target (noun) clusters, between which metaphoric associations hold; (iv) searches the corpus for metaphoric expressions describing the target domain concepts using the verbs from the source domain lexicon.

The English verb and noun datasets used for clustering contain the 2000 most frequent verbs and the 2000 most frequent nouns in the British National Corpus (BNC)². The BNC is balanced with respect to topic and genre, which makes it appropriate for the selection of a dataset of most common source and target concepts and their linguistic realizations. The features for clustering were, however, extracted from the English Gigaword corpus, which is more suitable for feature extraction due to its large size. The Gigaword corpus was parsed using the RASP parser [1].

The Russian verb and noun datasets used for clustering contain the 2000 most frequent verbs and the 2000 most frequent nouns in the Russian Web as Corpus (RuWaC). The RuWaC is a two billion-word representative collection of text from the Russian Web. The corpus was parsed using the Malt dependency parser for Russian [95].

The above datasets were analysed for examples of semantic dissonance – i.e., Russian expressions that employ a metaphoric framing verb which is not used in semantically equivalent English expressions.³ A total of 118 expressions were identified, which were then further harvested for the purposes of the present study. The crosslinguistic framework allows for the findings from this study to be generalized

²<http://www.natcorp.ox.ac.uk/>

³This step was conducted by a native speaker of English who is fluent in Russian (the author) and a native speaker of Russian who is fluent in English.

for addressing questions with respect to crosslinguistic differences in metaphor use and machine translation of metaphoric expressions.

4.4 INTELLIGIBILITY EXPERIMENTS

The following experiments were designed to measure the intelligibility of novel metaphoric expressions, with intelligibility defined as convergence of interpretation across native speakers. For each experiment, participants were shown a number of expressions (not all of which were metaphors) and asked to choose the best paraphrase of the expression. The level of contextual embedding varied across the three experiments. Stimuli for the experimental instruments were created according to the following procedure.

The 118 semantically dissonant expressions from the Russian dataset were harvested such that those most consistent with the following criteria were retained for inclusion in the experimental stimuli.

1. The expression represents a conventional metaphoric expression in Russian.
2. The verb in the Russian expression is productive with respect to an existing metaphor schema. For example, the Russian verb *приковать* is used to mark a CAPTIVITY source frame in schemas associated with emotions, ideas, and social progress.
3. An equivalent schema exists in English, for example, a CAPTIVITY source frame is regularly recruited to conceptualize abstractions such as emotions, ideas, and social progress.
4. The set of lexemes regularly used to recruit the equivalent schema in English does not include any lexemes in the set of English verbs regularly used as a direct translational equivalent of the Russian verb in the original metaphoric expression.

The forty-five metaphoric expressions that best fulfilled the above criteria served as the bases for the experimental stimuli. Each expression was translated into (i) a nonmetaphoric phrase expressing in English a meaning equivalent to that expressed by the original Russian phrase, (ii) a conventional metaphoric phrase expressing in English a meaning equivalent to that expressed by the original Russian phrase, and (iii) an unattested phrase in English that is a direct translational equivalent of the original Russian phrase. The resulting 135 phrases form an array of forty-five sets (one set per original Russian expression), with each set containing three members (i.e., the nonmetaphoric translation, the conventional translation, the unattested translation).

4.4.1 EXPERIMENT 1

For Experiment 1, the phrases were embedded within naturally occurring English sentences. The sentences were collected through web-based searches on the conventional metaphoric expressions in each set from the forty-five set array of expressions. Selected sentences were those that best accommodated the entire three-member set in terms of non-semantic linguistic criteria including grammaticality and phonological fit. The following sentence, for example, was chosen for the set (*teach skills, nurture skills, graft skills*).

- (12) Once we start developing awareness of the use of space as part of the language of drama we are *teaching* theater skills.
- (13) Once we start developing awareness of the use of space as part of the language of drama we are *nurturing* theater skills.
- (14) Once we start developing awareness of the use of space as part of the language of drama we are *grafting* theater skills.

The experimental task required participants to choose the best paraphrase for the highlighted phrase from a multiple choice list of five expressions. The multiple choice items were identified according to the following criteria.

1. Each item must appear only once in the text of the experiment.
2. All items must have similar co-occurrence frequencies.⁴
3. All items must have similar concreteness ratings.⁵

The multiple choice items for sentences (12), (13), and (4.5) included *deploy a skill*, *improve a skill*, *foster a skill*, *integrate a skill*, *offer a skill*.

Three versions of the experimental instrument were generated based on the alternations of embedded phrase within the stimuli. Thus, every participant saw a mix of nonmetaphoric expressions, conventional metaphoric expressions, and novel metaphoric expressions; and, a set of judgments for every novel metaphoric expression could be compared to the set of judgments for conventional metaphoric expressions and nonmetaphoric expressions within the same sentence and with the same multiple choice paraphrases. Figure 4.4.1 provides a screen shot of the Instructions for the experiment.

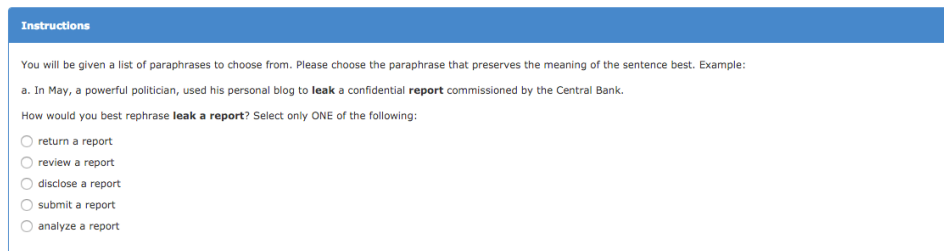


Figure 4.4.1: Instructions seen by participants in Experiment 1.

PARTICIPANTS

Two-hundred and seventy (270) workers on Amazon’s Mechanical Turk took part in Experiment 1. All were native speakers of English with the status of Master worker. For each of the three versions of the experimental instrument, ninety (90) unique workers participated.

⁴Frequency statistics were drawn from the Corpus of Contemporary American English (COCA) [33].

⁵Concreteness ratings were drawn from Brysbaert et al., [19].

RESULTS

Participant responses were compiled according to item and experimental condition. Agreement rates per item by condition were calculated using Gwet's AC₁ chance corrected agreement coefficient [54]. Gwet's method has achieved improved results in modeling agreement rates in studies of medical diagnoses and educational assessment [55] by reducing sensitivity to prevalence and eliminating extreme expected values of marginal probabilities. The agreement coefficients (Gwet's a 's) for this experiment were 0.66 given a nonmetaphoric expression, 0.62 given a conventional metaphoric expression, and 0.40 given a novel metaphoric expression.

Gwet [54] proposes a benchmarking model for agreement coefficients based on precision, taking into account statistical error, and the Interval Membership Probability for intervals within a given benchmark scale. Table 4.4.1 describes the Landis and Koch Kappa Benchmark Scale used in the present analysis.

Table 4.4.1: Landis and Koch Kappa Benchmark Scale.

| Kappa Statistic | Strength of Agreement |
|------------------------|------------------------------|
| < 0.0 | Poor |
| 0.0 to 0.20 | Slight |
| 0.21 to 0.40 | Fair |
| 0.41 to 0.60 | Moderate |
| 0.61 to 0.80 | Substantial |
| 0.81 to 1.00 | Almost Perfect |

Gwet's method for benchmarking proceeds as follows. For each interval beginning at the highest benchmark level, the Interval Membership Probability, or IMP, (i.e., the cumulative probability that the observed agreement coefficient falls in that interval or those above it) is calculated. The IMP at which the cumulative probability reaches 95% represents the level of agreement. The IMP is calculated

using the following formula [54].

$$IMP = P\left(\frac{COEFF - b}{SE} \leq Z \leq \frac{COEFF - a}{SE}\right) \quad (4.4)$$

Applying this method to the a 's reported above, the level of agreement for both nonmetaphoric and conventional metaphoric paraphrases is *Substantial*, while the level agreement for novel metaphoric paraphrases is *Moderate*.

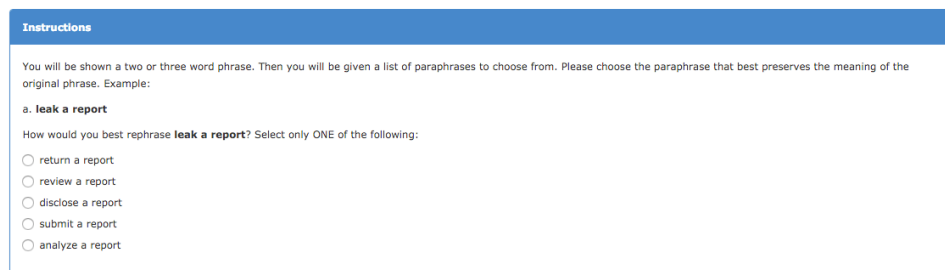
To test the difference between agreement coefficients for statistical significance, a pairwise test based on large-sample linear approximation of the agreement coefficient was conducted indicating that the difference between the agreement coefficient for nonmetaphoric and conventional metaphoric paraphrases was not significant ($p = 0.32$), the difference between the agreement coefficient for nonmetaphoric and novel metaphoric paraphrases was significant ($p = 0.008$), and the difference between the agreement coefficient for conventional metaphoric and novel metaphoric paraphrases was also significant ($p = 0.008$). Effect sizes in the latter cases were 0.81 and 0.77, respectively.

4.4.2 EXPERIMENT 2

In Experiment 2, the target phrases were disassociated from the contextual information provided by the sentences of Experiment 1. The experimental task similarly required participants to choose the best paraphrase, though in this experiment the stimuli consisted of the decontextualized phrase. The same multiple choice list of five expressions developed for Experiment 1 were used as the potential paraphrases.

Again, three versions of the experimental instrument were generated based on the alternations of translational phrase. Every participant again saw a mix of nonmetaphoric expressions, conventional metaphoric expressions, and novel metaphoric expressions; and, a set of judgments for every novel metaphoric expression could be compared to the set of judgments for conventional metaphoric expressions and nonmetaphoric expressions within the same sentence and with the same multiple choice paraphrases. Figure 4.4.2 provides a screen shot of the Instructions for

Experiment 2.



Instructions

You will be shown a two or three word phrase. Then you will be given a list of paraphrases to choose from. Please choose the paraphrase that best preserves the meaning of the original phrase. Example:

a. **leak a report**

How would you best rephrase **leak a report**? Select only ONE of the following:

- return a report
- review a report
- disclose a report
- submit a report
- analyze a report

Figure 4.4.2: Instructions seen by participants in Experiment 2.

PARTICIPANTS

Ninety (90) workers on Amazon’s Mechanical Turk took part in Experiment 1. All were native speakers of English with the status of Master worker. For each of the three versions of the experimental instrument, thirty (30) unique workers participated.

RESULTS

Participant responses were again compiled according to item and experimental condition. Agreement rates per item by condition were calculated using Gwet’s AC1 chance corrected agreement coefficient. Table 4.4.2 provides the agreement coefficient by condition for Experiments 1, 2, and 3. The agreement coefficients (Gwet’s *a*’s) for this experiment were 0.67 given a nonmetaphoric expression, 0.63 given a conventional metaphoric expression, and 0.36 given a novel metaphoric expression.

Applying the benchmarking method as detailed above to the *a*’s reported for this experiment, the level of agreement for both nonmetaphoric and conventional metaphoric paraphrases is *Substantial*, while the level agreement for novel metaphoric paraphrases is *Moderate*.

To test the difference between agreement coefficients for statistical significance, a pairwise test based on large-sample linear approximation of the agreement coef-

ficient was conducted indicating that the difference between the agreement coefficient for nonmetaphoric and conventional metaphoric paraphrases was not significant ($p = 0.32$), the difference between the agreement coefficient for nonmetaphoric and novel metaphoric paraphrases was significant ($p = 0.006$), and the difference between the agreement coefficient for conventional metaphoric and novel metaphoric paraphrases was also significant ($p = 0.006$). Effect sizes in the latter cases were 0.87 and 0.79, respectively.

4.4.3 EXPERIMENT 3

In Experiment 3, the contextual verbal phrases were disassociated from the target domain information indexed by the nouns. The experimental task similarly required participants to choose the best paraphrase, though in this experiment the stimuli consisted of the decontextualized phrase. The same multiple choice list of five expressions developed for Experiment 1 were used as the potential paraphrases.

Again, three versions of the experimental instrument were generated based on the alternations of translational phrase. Every participant again saw a mix of nonmetaphoric expressions, conventional metaphoric expressions, and novel metaphoric expressions; and, a set of judgments for every novel metaphoric expression could be compared to the set of judgments for conventional metaphoric expressions and nonmetaphoric expressions within the same sentence and with the same multiple choice paraphrases. Figure 4.4.3 provides a screen shot of the Instructions for Experiment 3.

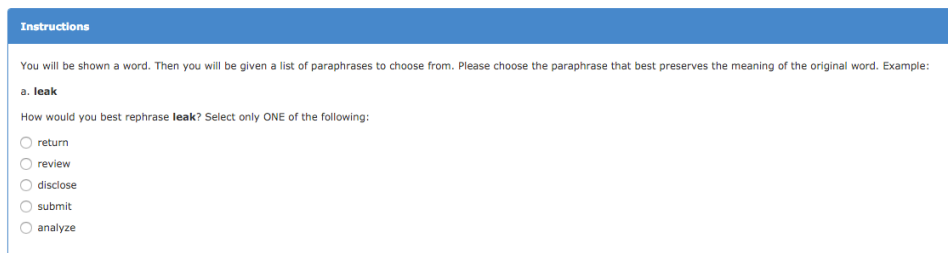


Figure 4.4.3: Instructions seen by participants in Experiment 3.

PARTICIPANTS

Ninety (90) workers on Amazon's Mechanical Turk took part in Experiment 3. All were native speakers of English with the status of Master worker. For each of the three versions of the experimental instrument, thirty (30) unique workers participated.

RESULTS

Participant responses were again compiled according to item and experimental condition. Agreement rates per item by condition were calculated using Gwet's AC₁ chance corrected agreement coefficient. Table 4.4.2 provides the agreement coefficient by condition for Experiments 1, 2, and 3. The agreement coefficients (Gwet's *a*'s) for this experiment were 0.61 given a nonmetaphoric expression, 0.54 given a conventional metaphoric expression, and 0.30 given a novel metaphoric expression.

Applying the benchmarking method as detailed above to the *a*'s reported for this experiment, the level of agreement for nonmetaphoric paraphrases is *Substantial*, for conventional metaphoric paraphrases *Moderate*, and for novel metaphoric paraphrases *Slight*.

To test the difference between agreement coefficients for statistical significance, a pairwise test based on large-sample linear approximation of the agreement coefficient was conducted indicating that the difference between the agreement coefficient for nonmetaphoric and conventional metaphoric paraphrases was not significant ($p = 0.08$), the difference between the agreement coefficient for nonmetaphoric and novel metaphoric paraphrases was significant ($p = 0.004$), and the difference between the agreement coefficient for conventional metaphoric and novel metaphoric paraphrases was also significant ($p = 0.006$). Effect sizes in the latter cases were 0.93 and 0.82, respectively.

Table 4.4.2 shows the agreement coefficients for all three experiments.

Table 4.4.2: Gwet's AC1 agreement coefficient by condition and context.

| Context | Condition | Coeff | Std. Error | 95% CI |
|--------------------|------------------|--------------|-------------------|---------------|
| Sentence | Nonmetaphoric | 0.66 | 0.04 | (0.58, 0.73) |
| Sentence | Conventional | 0.62 | 0.04 | (0.54, 0.70) |
| Sentence | Novel | 0.40 | 0.04 | (0.33, 0.48) |
| Stand-Alone Phrase | Nonmetaphoric | 0.67 | 0.04 | (0.58, 0.75) |
| Stand-Alone Phrase | Conventional | 0.63 | 0.04 | (0.54, 0.72) |
| Stand-Alone Phrase | Novel | 0.36 | 0.04 | (0.28, 0.44) |
| Verb Only | Nonmetaphoric | 0.61 | 0.04 | (0.52, 0.70) |
| Verb Only | Conventional | 0.54 | 0.04 | (0.46, 0.62) |
| Verb Only | Novel | 0.30 | 0.04 | (0.22, 0.37) |

4.5 GENERAL DISCUSSION

Conceptual Metaphor Theory (CMT) as put forth by Lakoff & Johnson [69] in *Metaphors We Live By* established metaphor as a cognitive phenomenon central to human language functionality. Metaphor processing within the human cognitive system consequently became an important operation to be explained by linguists and cognitive scientists. Early research focused on determining whether comprehension of metaphoric language involves qualitatively different processes than those associated with comprehension of nonmetaphoric language. Several important findings have emerged from this work, the chief of these being that the processing of metaphoric language is not dependent on a literal-first interpretation later to be corrected [46], and that the processing of conventional metaphoric language is as fast and automatic as the processing of nonmetaphoric language [51].

Another line of research examines the role of frequency effects on metaphor processing. In continuous dynamic systems such as human cognition, frequency effects act as a force driving the system towards equilibrium, influencing all aspects of human cognition including language [20]. Through several decades of research, frequency effects have been demonstrated to figure prominently in lan-

guage use and understanding. Vitevitch et al., [127] for instance, investigated the effect of phonotactic and stress placement frequency on acceptability judgments and processing times for nonsense words. They found that high probability sound patterns led to higher acceptability ratings and faster reaction times, illustrative of frequency effects at the level of phonological processing. Hare et al., [57] reported similar results showing frequency effects for lexical and morphological levels of linguistic structure. Extending this approach to semantics and the processing of metaphoric language, Blank [16] compared processing times for sentences ending with a word that introduces a nonmetaphoric interpretation versus ending the same sentence with a word conducive to a metaphoric interpretation. Reaction times here were higher in response to metaphoric endings, but the effect was smaller for more frequent metaphoric expressions. Gentner and Wolff [45] examined frequency effects in metaphor from the perspective of psychological priming. They tested whether target domain terms (i.e., words associated with the topic of a sentence) or source domain terms (i.e., words associated with the domain serving to frame the topic) more effectively prime comprehension of metaphoric language. They found that conventional metaphors were primed equally well by target domain and source domain terms, while novel metaphors were better primed by source domain terms. In a cross-modal priming study, Blasko and Connine [18] considered the effect of both familiarity and aptness on the processing of metaphoric language, with results suggesting that aptness decreases processing times only for unfamiliar metaphors. Blasko and Briihl [17] measured reading times and recall for familiar and unfamiliar metaphors. Familiarity affected the former but not the latter.

That frequency effects play an important role in the processing of metaphoric language can be accounted for by three influential theories of metaphor. Giora's [50] revision of Gibbs' [46] processing model incorporates the following assumptions. Saliency, based on contextual frequency effects, determines the interpretation of metaphoric and nonmetaphoric language; novel interpretations rely on rejection of salient interpretations; and, novel interpretations are more difficult and require more contextual support. In the Space Structuring Model of metaphor

comprehension proposed by Coulson and Matlock [29], metaphoric and non-metaphoric language recruit the same set of processes, which can be described in terms of a conceptual network. Integration of concepts and features within a conceptual network involves processes of composition, completion, and elaboration. Routine integrations require less elaboration than infrequently instantiated ones and, consequently, can be expected to resolve into a stable state more quickly. Emergent Metaphor Theory [100], which posits that metaphoric schemata reflect usage patterns in metaphoric language, has recently been tested through a number of experiments which show that higher frequency metaphors elicit higher acceptability ratings and incur less time for comprehension.

Frequency effects and conventionalization patterns form the foundation of Emergent Metaphor Theory (EMT), which posits that metaphor schemata emerge from frequency effects in language use [100]. From the perspective of EMT, the strength of a particular metaphor schema is proportional to the frequencies of the tokens that invoke it and to the family size of the schema. For example, the metaphoric expressions *illuminate the idea*, *illuminate grievances*, *illuminate the causes*, *illustrate the principle*, and *illustrate the absurdity* all reinforce the metaphor schema KNOWING IS SEEING; whereas the expressions *grasp the idea* and *grasp the solution* reinforce the schema KNOWING IS TOUCHING. KNOWING IS SEEING is a stronger metaphoric schema than KNOWING IS TOUCHING since it is instantiated by a greater variety of expression types and the frequency of tokens within these types is greater. EMT claims that the accessibility and ease of processing of a given metaphoric expression depends on the strength of the schema the expression instantiates as well as the frequency of the expression itself.

Sanford [99] tested this claim using accessibility judgments and reaction time measurements. He found that accessibility increases and reaction times decrease as the metaphoric schema invoked by an expression becomes stronger. These results replicate and extend many earlier findings of faster processing times for conventional or familiar metaphors versus novel metaphors [47]. However, processing times and accessibility judgments provide an incomplete picture of semantic processing. To gain a better understanding of the semantic processing of novel

versus conventional expressions, the issue of semantic divergence must also be addressed. Semantic divergence reflects the principle that the more meaningless or nonsensical an expression, the less likely the speakers of a language are to converge on a unified interpretation. In this paper, we introduce a number of novel metaphoric expressions that have been generated through a cross-linguistic analysis of metaphor schema and their terminological instantiations. We then report the results of several experiments designed to measure the semantic divergence of the novel expressions and that of conventional metaphoric expressions relative to literal counterparts. We show that although the novel expressions have greater semantic divergence than the conventional expressions, they are not meaningless or nonsensical. This property persists as contextual information is eliminated as long as the term associated with the target conceptual domain remains intact. Our results support the proposition that metaphor schemata contribute to the processing of metaphoric expressions. They also represent unique evidence for the compositionality of conventional as well as novel metaphoric expressions.

While the issue of how frequency affects the time course of metaphor comprehension has been broadly addressed in the metaphor research literature, the question of how frequency influences metaphor interpretation has been less well studied. Framing these questions in terms of locations in semantic space, metaphor research has devoted much effort to understanding the time course, relative to metaphor frequency, for arriving at a stable location in semantic space. The extent to which frequency affects convergence on a single location in semantic space has, in contrast, received little attention.

The experiments discussed in this chapter represent an effort to address this question. The results from the three experiments indicate that people are less likely to converge on a given meaning when confronted with novel metaphoric expressions. The results of Experiment 3 shows in particular that these effects cascade down to the level of the framing verb or source domain lexicon. That is to say, given a verb that is frequently used to invoke a source domain in a metaphoric mapping, people are more likely to converge on a more abstract paraphrase of that verb than when they are given a verb rarely used metaphorically.

To examine the implications more closely, a qualitative analysis of the individual experimental items was performed. Figure 4.5.1 presents the agreement rates across contexts for each item (sentence, phrase, and verb).

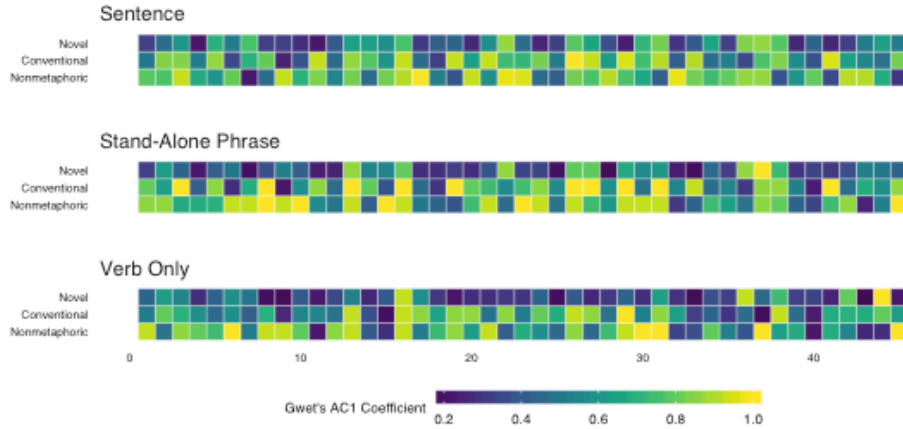


Figure 4.5.1: Agreement rates per item by condition.

The following item, and its equivalents in Experiments 2 and 3, resulted in high agreement rates in the nonmetaphoric and conventional metaphoric conditions but low agreement rates in the novel metaphoric condition.

Some of the younger generation also incorporate/absorb/catch a trend and mix it with a traditional twist.

The choices for the paraphrase were: assimilate (a trend); intensify (a trend); forecast (a trend); observe (a trend); welcome (a trend).

That participants were unlikely to converge on a paraphrase given the novel metaphor *catch a trend*, even though the conceptual mapping INTELLECTUAL ACCEPTANCE IS PHYSICAL CONTACT, suggests that frequency effects influence how metaphors are interpreted. It is not just that people take longer to interpret novel

metaphors – they are also less likely to arrive at the same location in semantic space. This supports a view of metaphor processing in which comprehension relies on a continuously updated prediction of incoming content based on currently available information. In line with theories regarding semantic deviance [2], unexpected expressions can be deviant on the basis of distributional semantics alone. This pertains to metaphoric expressions, which are considered not entirely compositional, as well as expressions that are more straightforwardly compositional.

4.6 CONCLUSION

In this chapter I described the results of a series of experiments testing the intelligibility or interpretability of novel metaphoric expressions as compared with that of nonmetaphoric and conventional metaphoric expressions. The findings from these experiments have implications with respect to theories of metaphor processing and language processing in general. They also shed new light on linguistic theories of compositionality and semantic deviance.

With respect to the dissertation research being presented here, the results of these experiments allow me to establish a metric for intelligibility. Additionally, they allow me to test my hypothesis that novel metaphors introducing a novel mapping are not only more valuable but more intelligible than those recapitulating an old one.

Finality is a law as tyrannical as rhythm; for it too constrains our movements to follow a predetermined order. Only by not conforming too rigorously can we give the impression of grace.

Souriau

5

Curtain Calls and Coming Attractions

The main contributions of this dissertation are within the fields of cognitive science, linguistic semantics, and metaphor research. The dissertation describes new approaches and original findings with respect to blind variation and selective retention processes during creative cognition, with respect to the interactions of semantic fields in lexical semantics, and with respect to the properties of the collective behavior of systems of metaphor ensembles. A brief recapitulation of these is given below.

In Chapter 2 of this dissertation, I build upon well utilized algorithms for identifying metaphors by applying affinity propagation clustering to identify ensembles of metaphoric expressions. With this technique I am able to discover ensembles that are equivalently well-linked to a central exemplar expression. This allows a firm basis for experimenting with associational semantic field vector sequences as input to a creative process of blind variation and selective retention. The principles

behind the method further link to cognitive science research on categorization and the function of prototypes and family resemblances.

In Chapter 3, I develop a spin model of the creative aspect of language use. From the energy landscape described by the model, I identify basins of attraction as sources of novel semantic field sequences that, due to their occupation of a highly stable location in the energy landscape, also have high potential to possess the qualities of useful and valuable. The model thus implements a creative process of blind variation and selective retention. Through a dynamical systems analysis of the observed data and the model data, I find that both systems are in a state poised between negentropy and entropy, a finding consistent with many other complex systems in nature. The maximum entropy formalism shows how the observable correlations between metaphorical expressions at any two sites carry the signatures of collective behavior in the system as a whole. The idea that crucial aspects of semantics should be viewed as emergent, collective phenomena has been discussed for decades. The challenge has been to identify and develop precise mathematical tools for extracting quantitative models of this collective behavior from the abundant linguistic data available. The models I develop in Chapter 3 contribute to the ongoing efforts to meet this challenge.

In Chapter 4, I establish experimental techniques for measuring quantitatively how well the novel expressions generated by the Sp[i]CALU exhibit the qualities of useful and valuable, both of which have been identified in the philosophical, psychological, and computational literature on creativity as criteria for creativity. Using computational experiments, I show that the generated expressions are valuable in that they expand semantic space. With the human subjects experiments, I show that they are useful in that they are nearer to conventional metaphoric expressions and non-metaphorical expressions than to translations of foreign expressions in terms of intelligibility.

Future directions for this research include experimenting with more divergent and less divergent semantic ensembles to understand more precisely how the balance of entropy to negentropy influences semantic systems and creativity within semantic systems. Is generating of creative expressions dulled when the system

consists of one metaphor ensemble, two metaphor ensembles? A related question is what fluctuation, if any, arises in the optimal balance of entropy to negentropy across different types of systems. Is the optimal balance for a semantic system much different from that of a system of sequence alignments from, for example, phonology or prosody? How does the relationship between entropy and negentropy change if information from both streams is included in the model?

A further line of promising future research is to investigate historical linguistics and language change with spin models. In the models I explore in this dissertation, the clusters associated with the metastable states are not completely disconnected from one other. Continuous paths of sequences were observed between most metastable states, which suggests that they could be traces of earlier creative (or deviant, or mutant) developments. The zero-temperature Monte Carlo method used for finding these paths, because it naturally favors low energy barriers, is more likely to find paths where all sequences are present in the data. An extension of the current work could include retracing trajectories between a collection of historical linguistic changes and the corresponding variation in semantic field sequencing. Such an investigation could potentially provide new evidence on the mechanisms of language evolution.

Appendix

```
AffinityProp[s_] :=
  n = Dimensions[s];
  a = ConstantArray[0, {n}];
  r = ConstantArray[0, {n}];
  s = s + ($MachineEpsilon * s + $MinMachineNumber * 100).RandomReal[n];
  lam = 0.7;
  For[i = 1, i < 10000, i++,
    oldR = r;
    as = a + s; (y1, I1) = Max[as, [], 2];
    For[i = 1, i < (n + 1), as[i, I[i]] = -$MaxMachineNumber];
    (y2, I2) = Max[as, [], 2];
    r = s - Array[y1, {1, n}];
    For[i = 1, i < (n + 1), nr[i, I[i]] = s[i, I[i]] - y2[i]];
    r = (1 - lam)*r + lam*oldR;
    oldA = a;
    rp = Max[r, 0];
    For[k = 1, k < (n + 1), rp[k, k] = r[k, k]];
    a = Array[Sum[Rp, 1], [N, 1]] - rp;
    dA = Diagonal[a];
    a = Min[a, 0];
    For[ k = 1, i < (n + 1), a[k, k] = dA[k]];
    a = (1 - lam)*a + lam*oldA;
  ];
  e = r + a;
  I = Diagonal[e] > 0;
  K = Length[I];
  t = Max[s[ : , I], 2];
  idx = I[c]
```



```

SeedRandom[121212];
generateMatrix[nStates_, x_] :=
Module[{i, j, p, allPairs, n, m},
  n = Length[x];
  p = ConstantArray[0, {nStates}, {nStates}];
  allPairs = Partition[x, 2, 1];
  For[i = 1, i <= nStates, i++,
    {
      m = Count[allPairs, {i, y_}];
      For[j = 1, j <= nStates, j++,
        If[m != 0, p[[i, j]] = Count[allPairs, {i, j}]/m, p[[i, j]] = 0]
      ]
    }
  ];
  p
]
spin[pi_, m_, dim_] :=
  beta = Log[1 + Sqrt[2]];
  \[Alpha], u, x, sample, pts, sum, k, found},
  nStates = Length[m];
  n = 1;
  x = ConstantArray[0, {dim}];
  x[[n]] = 1;
  i = 1;
  While[i < maxN,
    {
      y = sampler[q, x[[n]]];
      \[Alpha] = beta[x[[n]], y, pi, q];
      u = RandomReal[];
      n++;
      If[u <= \[Alpha], x[[n]] = y,
        x[[n]] = x[[n - 1]]];
      i++;
    }
  ];
  generateMatrix[nStates, x]
]
cumSum[list_] := Module[{i, sum, s, k},

```

```

sum = 0;
k = Length[list];
s = ConstantArray[0, {k}];
For[i = 1, i <= k, i++,
  {
    sum = sum + list[[i]];
    s[[i]] = sum;
  }
];
s
]

sampler[q_, x_] := Module[{s, found, j, k, sample, y},
  s = Flatten[Position[q[[ x, All]], Except[0], 1, Heads -> False] ];
  sample = q[[ x, s]];
  sample = cumSum[sample];
  y = RandomReal[];
  found = False;
  For[j = 1, j <= Length[sample], j++,
    If[ Not[found], If[y <= sample[[j]], {k = j; found = True}]]
  ];
  y = s[[k]]
]

def sigmoid(x):
    return 1. / (1 + np.exp(-x))

def sigmoid_derivativ(y):
    return y * (1 - y)

def tanh_derivative(y):
    return 1. - y**2

def rand_arr(a, b, *args):
    np.random.seed(0)
    return np.random.rand(*args) * (b - a) + a

class par:
    def __init__(self, n, dimX):

```

```

self.n = n
self.dimX = dimX
len = dimX + n
self.wg = rand_arr(-0.1, 0.1, n, len)
self.wi = rand_arr(-0.1, 0.1, n, len)
self.wf = rand_arr(-0.1, 0.1, n, len)
self.wo = rand_arr(-0.1, 0.1, n, len)
self.bg = rand_arr(-0.1, 0.1, n)
self.bi = rand_arr(-0.1, 0.1, n)
self.bf = rand_arr(-0.1, 0.1, n)
self.bo = rand_arr(-0.1, 0.1, n)
self.wgD = np.zeros((n, len))
self.wiD = np.zeros((n, len))
self.wfD = np.zeros((n, len))
self.woD = np.zeros((n, len))
self.bgD = np.zeros(n)
self.biD = np.zeros(n)
self.bfD = np.zeros(n)
self.boD = np.zeros(n)

def diff(self, lr):
self.wg -= lr * self.wgD
self.wi -= lr * self.wiD
self.wf -= lr * self.wfD
self.wo -= lr * self.woD
self.bg -= lr * self.bgD
self.bi -= lr * self.biD
self.bf -= lr * self.bfD
self.bo -= lr * self.boD
self.wgD = np.zeros_like(self.wg)
self.wiD = np.zeros_like(self.wi)
self.wfD = np.zeros_like(self.wf)
self.woD = np.zeros_like(self.wo)
self.bgD = np.zeros_like(self.bg)
self.biD = np.zeros_like(self.bi)
self.bfD = np.zeros_like(self.bf)
self.boD = np.zeros_like(self.bo)

```

```

class state:
    def __init__(self, n):
        self.g = np.zeros(n)
        self.i = np.zeros(n)
        self.f = np.zeros(n)
        self.o = np.zeros(n)
        self.s = np.zeros(n)
        self.h = np.zeros(n)
        self.lowerH = np.zeros_like(self.h)
        self.lowerS = np.zeros_like(self.s)

class node:
    def __init__(self, par, state):
        self.par = par
        self.state = state
        self.xc = None

    def lower(self, x, sp = None, hp = None):
        if sp == None: sp = np.zeros_like(self.state.s)
        if hp == None: hp = np.zeros_like(self.state.h)
        self.sp = sp
        self.hp = hp
        xc = np.hstack((x, hp))
        self.state.g = np.tanh(np.dot(self.par.wg, xc) + self.par.bg)
        self.state.i = sigmoid(np.dot(self.par.wi, xc) + self.par.bi)
        self.state.f = sigmoid(np.dot(self.par.wf, xc) + self.par.bf)
        self.state.o = sigmoid(np.dot(self.par.wo, xc) + self.par.bo)
        self.state.s = self.state.g * self.state.i + sp * self.state.f
        self.state.h = self.state.s * self.state.o

        self.xc = xc

    def upper(self, uh, us):
        ds = self.state.o * uh + us
        do = self.state.s * uh
        di = self.state.g * ds
        dg = self.state.i * ds
        df = self.s_prev * ds

```

```

di_input = sigmoid_derivative(self.state.i) * di
df_input = sigmoid_derivative(self.state.f) * df
do_input = sigmoid_derivative(self.state.o) * do
dg_input = tanh_derivative(self.state.g) * dg

self.par.wiD += np.outer(di_input, self.xc)
self.par.wfD += np.outer(df_input, self.xc)
self.par.woD += np.outer(do_input, self.xc)
self.par.wgD += np.outer(dg_input, self.xc)
self.par.biD += di_input
self.par.bfD += df_input
self.par.boD += do_input
self.par.bgD += dg_input

dxc = np.zeros_like(self.xc)
dxc += np.dot(self.par.wi.T, di_input)
dxc += np.dot(self.par.wf.T, df_input)
dxc += np.dot(self.par.wo.T, do_input)
dxc += np.dot(self.par.wg.T, dg_input)

self.state.lowerS = ds * self.state.f
self.state.lowerH = dxc[self.par.dimX:]

class net():
    def __init__(self, parL):
        self.parL = parL
        self.listN = []
        self.listX = []

    def listY_is(self, listY, loss):

        assert len(listY) == len(self.listX)
        idx = len(self.listX) - 1
        loss = loss.loss(self.listN[idx].state.h, listY[idx])
        dh = loss.lowerD(self.listN[idx].state.h, listY[idx])
        ds = np.zeros(self.parL.n)
        self.listN[idx].topD_is(dh, ds)
        idx -= 1

```

```

while idx >= 0:
    loss += loss.loss(self.listN[idx].state.h, listY[idx])
    dh = loss.lowerD(self.listN[idx].state.h, listY[idx])
    dh += self.listN[idx + 1].state.lowerD_h
    ds = self.listN[idx + 1].state.lowerD_s
    self.listN[idx].topD_is(dh, ds)
    idx -= 1

return loss

def listXC(self):
    self.listX = []

def listX_add(self, x):
    self.listX.append(x)
    if len(self.listX) > len(self.listN):
        lstmS = state(self.parL.n, self.parL.dimX)
        self.listN.append(LstmNode(self.parL, lstmS))
    idx = len(self.listX) - 1
    if idx == 0:
        self.listN[idx].lower_data_is(x)
    else:
        sp = self.listN[idx - 1].state.s
        hp = self.listN[idx - 1].state.h
        self.listN[idx].lower_data_is(x, s_prev, h_prev)

```

References

- [1] Øistein E Andersen, Julien Nioche, Ted Briscoe, and John A Carroll. The BNC parsed with RASP4UIMA. In *Proceedings of LREC*, 2008.
- [2] Sally Andrews, Brett Miller, and Keith Rayner. Eye movements and morphological segmentation of compound words: There is a mouse in mouse-trap. *European Journal of Cognitive Psychology*, 16(1-2):285–311, 2004.
- [3] Antoine Arnault and Claude Lancelot. *Royal Grammar*. Cambridge University Press, 1975.
- [4] Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. Opinionetit: Understanding the opinions-people network for politically controversial topics. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2481–2484. ACM, 2011.
- [5] Alexander Bain. *The senses and the intellect*. Longman, Green, Longman, Roberts, and Green, 1864.
- [6] Jeff P Barnes and Jack H Freed. Dynamics and ordering in mixed model membranes of dimyristoylphosphatidylcholine and dimyristoylphosphatidylserine: a 250-ghz electron spin resonance study using cholestane. *Biophysical journal*, 75(5):2532–2546, 1998.
- [7] Marco Baroni and Alessandro Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721, 2010.
- [8] Lawrence W Barsalou. Grounded cognition: Past, present, and future. *Topics in cognitive science*, 2(4):716–724, 2010.
- [9] Roger E Beaty, Mathias Benedek, Robin W Wilkins, Emanuel Jauk, Andreas Fink, Paul J Silvia, Donald A Hodges, Karl Koschutnig, and

- Aljoscha C Neubauer. Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia*, 64:92–98, 2014.
- [10] Roger E Beaty, Mathias Benedek, Scott Barry Kaufman, and Paul J Silvia. Default and executive network coupling supports creative idea production. *Scientific reports*, 5:10964, 2015.
- [11] Roger E Beaty, Mathias Benedek, Paul J Silvia, and Daniel L Schacter. Creative cognition and brain network dynamics. *Trends in cognitive sciences*, 20(2):87–95, 2016.
- [12] Roger E Beaty, Paul J Silvia, and Mathias Benedek. Brain networks underlying novel metaphor production. *Brain and Cognition*, 111:163–170, 2017.
- [13] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155, 2003.
- [14] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M Walczak. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*, 109(13):4786–4791, 2012.
- [15] Julia Birke and Anoop Sarkar. A clustering approach for nearly unsupervised recognition of nonliteral language. In *EACL*, 2006.
- [16] Glenn David Blank. Metaphors in the lexicon. *Metaphor and Symbol*, 3(3):21–36, 1988.
- [17] Dawn G Blasko and Deborah S Briihl. Reading and recall of metaphorical sentences: Effects of familiarity and context. *Metaphor and Symbol*, 12(4):261–285, 1997.
- [18] Dawn G Blasko and Cynthia M Connine. Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology Learning Memory and Cognition*, 19:295–295, 1993.
- [19] Marc Brysbaert and Boris New. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction

of a new and improved word frequency measure for american english. *Behavior research methods*, 41(4):977–990, 2009.

- [20] Joan Bybee. *Language, usage and cognition*. Cambridge University Press, 2010.
- [21] Donald T Campbell. Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review*, 67(6):380, 1960.
- [22] Donald T Campbell. Unjustified variation and selective retention in scientific discovery. *Studies in the philosophy of biology*, pages 139–61, 1974.
- [23] Donald T Campbell. A general ‘selection theory’, as implemented in biological evolution and in social belief-transmission-with-modification in science. *Biology and Philosophy*, 3(2):171–177, 1988.
- [24] Peter Carruthers. Human creativity: Its cognitive basis, its evolution, and its connections with childhood pretence. *The British journal for the philosophy of science*, 53(2):225–249, 2002.
- [25] Ronald Carter. Common language: corpus, creativity and cognition. *Language and Literature*, 8(3):195–216, 1999.
- [26] Ciprian Chelba, Tomas Mikolov, Mike Schuster, Qi Ge, Thorsten Brants, Phillip Koehn, and Tony Robinson. One billion word benchmark for measuring progress in statistical language modeling. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [27] Noam Chomsky. A note on the creative aspect of language use. *The Philosophical Review*, 91(3):423–434, 1982.
- [28] Herbert H Clark. *Using language*. 1996.
- [29] Seana Coulson and Teenie Matlock. Metaphor and the space structuring model. *Metaphor and Symbol*, 16(3-4):295–316, 2001.
- [30] Arthur Cropley. In praise of convergent thinking. *Creativity research journal*, 18(3):391–404, 2006.
- [31] Mihaly Csikszentmihalyi. Flow and the psychology of discovery and invention. *Harper Collins*, 1996.

- [32] Subrata Dasgupta. On the blind-mindedness of creative thought: Comment on “creative thought as blind-variation and selective-retention: Combinatorial models of exceptional creativity” by dean keith simonton. *Physics of life reviews*, 7(2):188–189, 2010.
- [33] Mark Davies. *The corpus of contemporary American English*. BYE, Brigham Young University, 2008.
- [34] John W Du Bois. Towards a dialogic syntax. *Cognitive Linguistics*, 25(3):359–410, 2014.
- [35] Delbert Dueck and Brendan J Frey. Non-metric affinity propagation for unsupervised image categorization. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.
- [36] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- [37] Gilles Fauconnier and Mark Turner. *The way we think: Conceptual blending and the mind’s hidden complexities*. Basic Books, 2008.
- [38] Andrew L Ferguson, Jaclyn K Mann, Saleha Omarjee, Thumbi Ndung’u, Bruce D Walker, and Arup K Chakraborty. Translating hiv sequences into quantitative fitness landscapes predicts viral vulnerabilities for rational immunogen design. *Immunity*, 38(3):606–617, 2013.
- [39] Kurt Feyaerts. Towards a dynamic account of phraseological meaning: creative variation in headlines and conversational humour. *International Journal of English Studies*, 6(1):57–84, 2009.
- [40] John Rupert Firth. *Papers in Linguistics 1934-1951: Repr.* Oxford University Press, 1961.
- [41] Mark V Flinn, David C Geary, and Carol V Ward. Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior*, 26(1):10–46, 2005.
- [42] Brendan J Frey and Delbert Dueck. Clustering by passing messages between data points. *science*, 315(5814):972–976, 2007.

- [43] Liane Gabora. Commentaries: Why the creative process is not darwinian: Comment on” the creative process in picasso’s guernica sketches: Monotonic improvements versus nonmonotonic variants”. *Creativity Research Journal*, 19(4):361–365, 2007.
- [44] James E Gentle. *Random number generation and Monte Carlo methods*. Springer Science & Business Media, 2006.
- [45] Dedre Gentner and Phillip Wolff. Alignment in the processing of metaphor. *Journal of memory and language*, 37(3):331–355, 1997.
- [46] Raymond W Gibbs. Literal meaning and psychological theory. *Cognitive science*, 8(3):275–304, 1984.
- [47] Raymond W Gibbs. *The poetics of mind*, 1994.
- [48] Raymond W Gibbs Jr. *Conceptual metaphor in thought and social action*. 2014.
- [49] Rachel Giora. On the cognitive aspects of the joke. *Journal of Pragmatics*, 16(5):465–485, 1991.
- [50] Rachel Giora. Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics (includes Cognitive Linguistic Bibliography)*, 8(3):183–206, 1997.
- [51] Rachel Giora. Literal vs. figurative language: Different or equal? *Journal of pragmatics*, 34(4):487–506, 2002.
- [52] Lila R Gleitman. Verbs of a feather flock together. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, pages 209–232, 2002.
- [53] Thomas L Griffiths, Mark Steyvers, and Joshua B Tenenbaum. Topics in semantic representation. *Psychological review*, 114(2):211, 2007.
- [54] Kilem Gwet. Kappa statistic is not satisfactory for assessing the extent of agreement between raters. *Statistical methods for inter-rater reliability assessment*, 1(6):1–6, 2002.
- [55] Kilem Li Gwet. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*, 61(1):29–48, 2008.

- [56] Jacques Hadamard. *The mathematician's mind: The psychology of invention in the mathematical field*. Princeton University Press, 1945.
- [57] Mary Hare, Ken McRae, and Jeffrey L Elman. Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48(2):281–303, 2003.
- [58] Zellig S Harris. Distributional structure. *Word*, 10(2-3):146–162, 1954.
- [59] Werner Heisenberg and James MacLachlan. Physics and philosophy: The revolution of modern science. vol. 19 of world perspectives, 1958.
- [60] Wilhelm von Humboldt. 1988. *On language: The diversity of human language structure and its influence on the mental development of mankind*, pages 1836–1840, 1836.
- [61] Ramon Ferrer i Cancho and Ricard V Solé. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791, 2003.
- [62] Michael N Jones and Douglas JK Mewhort. Representing word meaning and order information in a composite holographic lexicon. *Psychological review*, 114(1):1, 2007.
- [63] Rex Eugene Jung, Brittany S Mead, Jessica Carrasco, and Rane A Flores. The structure of creative cognition in the human brain. *Frontiers in human neuroscience*, 7:330, 2013.
- [64] Teuvo Kohonen and Panu Somervuo. Self-organizing maps of symbol strings with application to speech recognition. In *Proceedings of WSOM*, volume 97, pages 4–6, 1997.
- [65] Zoltán Kövecses. *Metaphor in culture: Universality and variation*. Cambridge University Press, 2005.
- [66] Maria E. Kronfeldner. Creativity naturalized. *The Philosophical Quarterly*, 59(237):577–592, 2009.
- [67] Alexander G Kruglov. Creativity as a determinant of the development of homo sapiens. *International Journal of Biomedicine*, 6(4):298–302, 2016.
- [68] William Labov. Rules for ritual insults. In *Sociolinguistics*, pages 472–486. Springer, 1997.

- [69] George Lakoff and Mark Johnson. *Metaphors we live by*. Chicago University, 1980.
- [70] Mark J Landau, Daniel Sullivan, and Jeff Greenberg. Evidence that self-relevant motives and metaphoric framing interact to influence political and social attitudes. *Psychological Science*, 20(11):1421–1427, 2009.
- [71] Thomas K Landauer and Susan T Dumais. A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.
- [72] Thomas K Landauer, Darrell Laham, and Peter Foltz. Learning human-like knowledge by singular value decomposition: A progress report. *Advances in neural information processing systems*, pages 45–51, 1998.
- [73] Ronald M Levy, Allan Haldane, and William F Flynn. Potts hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Current Opinion in Structural Biology*, 43:55–62, 2017.
- [74] Patricia Lichtenstein and Ekaterina Shutova. Metaphor congruent image schemas shape evaluative judgment: a cross-linguistic study of metaphors for economic change. *Proceedings of the Cognitive Science Society*, 2017.
- [75] Siyuan Liu, Ho Ming Chow, Yisheng Xu, Michael G Erkinen, Katherine E Swett, Michael W Eagle, Daniel A Rizik-Baer, and Allen R Braun. Neural correlates of lyrical improvisation: an fmri study of freestyle rap. *Scientific reports*, 2:834, 2012.
- [76] Siyuan Liu, Michael G Erkinen, Meghan L Healey, Yisheng Xu, Katherine E Swett, Ho Ming Chow, and Allen R Braun. Brain activity and connectivity during poetry composition: toward a multidimensional model of the creative process. *Human brain mapping*, 36(9):3351–3372, 2015.
- [77] Kevin Lund and Curt Burgess. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2):203–208, 1996.
- [78] Nicholas M Luscombe, Jiang Qian, Zhaolei Zhang, Ted Johnson, and Mark Gerstein. The dominance of the population by a selected few: power-law behaviour applies to a wide variety of genomic properties. *Genome biology*, 3(8):research0040–1, 2002.

- [79] Ernst Mach. On the part played by accident in invention and discovery. *The Monist*, 6(2):161–175, 1896.
- [80] Benoit Mandelbrot. An informational theory of the statistical structure of language. *Communication theory*, 84:486–502, 1953.
- [81] Jaclyn K Mann, John P Barton, Andrew L Ferguson, Saleha Omarjee, Bruce D Walker, Arup Chakraborty, and Thumbi Ndung’u. The fitness landscape of hiv-1 gag: advanced modeling approaches and validation of model predictions by in vitro testing. *PLoS Comput Biol*, 10(8):e1003776, 2014.
- [82] Christopher D Manning, Hinrich Schütze, et al. *Foundations of statistical natural language processing*, volume 999. MIT Press, 1999.
- [83] Zachary J Mason. Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational linguistics*, 30(1):23–44, 2004.
- [84] Teenie Matlock. Framing political messages with grammar and metaphor. *American Scientist*, 100(6):478, 2012.
- [85] Scott McDonald and Michael Ramscar. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. In *Proceedings of the 23rd annual conference of the cognitive science society*, pages 611–616, 2001.
- [86] Vinod Menon and Lucina Q Uddin. Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5-6):655–667, 2010.
- [87] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [88] Arthur I Miller. *Insights of genius: Imagery and creativity in science and art*. Springer Science & Business Media, 2012.
- [89] Jeff Mitchell and Mirella Lapata. Composition in distributional models of semantics. *Cognitive science*, 34(8):1388–1429, 2010.

- [90] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *science*, 320 (5880):1191–1195, 2008.
- [91] Steven Mithen. A creative explosion? theory of mind, language, and the disembodied mind of the upper paleolithic. in (s. mithen, ed.) *creativity in human evolution and prehistory*, 1998.
- [92] Michael Mohler, David Bracewell, David Hinote, and Marc Tomlinson. Semantic signatures for example-based linguistic metaphor detection. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 27–35, 2013.
- [93] Brigitte Nerlich and David D Clarke. Ambiguities we live by: Towards a pragmatics of polysemy. *Journal of Pragmatics*, 33(1):1–20, 2001.
- [94] Mark EJ Newman. Power laws, pareto distributions and zipf’s law. *Contemporary physics*, 46(5):323–351, 2005.
- [95] Joakim Nivre, Johan Hall, Jens Nilsson, Atanas Chanev, Gülşen Eryigit, Sandra Kübler, Svetoslav Marinov, and Erwin Marsi. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135, 2007.
- [96] H Poincare. *The foundations of science: Science and hypothesis, the value of science, science and method*, univ. Press of America, Washington, 1982.
- [97] Geoffrey K Pullum and Barbara C Scholz. Recursion and the infinitude claim. *Recursion in human language*, (104):113–138, 2010.
- [98] Eugene Sadler-Smith. Wallas’ four-stage model of the creative process: More than meets the eye? *Creativity Research Journal*, 27(4):342–352, 2015.
- [99] Daniel Sanford. Emergent metaphor theory: Frequency, schematic strength, and the processing of metaphorical utterances. *Journal of Cognitive Science*, 14(1):1–45, 2013.
- [100] Daniel R Sanford. Metaphors are conceptual schemata that are emergent over tokens of use. *Journal of Cognitive Science*, 13(3):211–235, 2012.

- [101] Elad Schneidman, Jason L Puchalla, Ronen Segev, Robert A Harris, William Bialek, and Michael J Berry. Synergy from silence in a combinatorial neural code. *Journal of Neuroscience*, 31(44):15732–15741, 2011.
- [102] Guy Sella and Aaron E Hirsh. The application of statistical physics to evolutionary biology. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9541–9546, 2005.
- [103] Karthik Shekhar, Claire F Ruberman, Andrew L Ferguson, John P Barton, Mehran Kardar, and Arup K Chakraborty. Spin models inferred from patient-derived viral sequence data faithfully describe hiv fitness landscapes. *Physical review E*, 88(6):062705, 2013.
- [104] Ekaterina Shutova. Automatic metaphor interpretation as a paraphrasing task. In *Proceedings of NAACL 2010*, pages 1029–1037, 2010.
- [105] Ekaterina Shutova. Models of metaphor in nlp. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 688–697. Association for Computational Linguistics, 2010.
- [106] Ekaterina Shutova, Lin Sun, Elkin Darío Gutiérrez, Patricia Lichtenstein, and Srinu Narayanan. Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning. *Computational Linguistics*, 2017.
- [107] Ekaterina V Shutova. *Computational approaches to figurative language*. PhD thesis, 2011.
- [108] Dean Keith Simonton. Debating the bvsr theory of creativity: Comments on dasgupta (2011) and gabora (2011). *Creativity Research Journal*, 23(4):381–387, 2011.
- [109] Dean Keith Simonton. Creativity, problem solving, and solution set sightedness: Radically reformulating bvsr. *The Journal of Creative Behavior*, 46(1):48–65, 2012.
- [110] Jonathan Smallwood, Emily Beach, Jonathan W Schooler, and Todd C Handy. Going awol in the brain: Mind wandering reduces cortical analysis of external events. *Journal of cognitive neuroscience*, 20(3):458–469, 2008.

- [111] Alexandre P Solon, Jean-Baptiste Caussin, Denis Bartolo, Hugues Chaté, and Julien Tailleur. Pattern formation in flocking models: A hydrodynamic description. *Physical Review E*, 92(6):062111, 2015.
- [112] Étienne Souriau, Erik Beranek, and Tim Howles. *The Different Modes of Existence*. University of Minnesota Press, 2016.
- [113] Michael J Spivey and Michael K Tanenhaus. Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6):1521, 1998.
- [114] R Nathan Spreng, W Dale Stevens, Jon P Chamberlain, Adrian W Gilmore, and Daniel L Schacter. Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *Neuroimage*, 53(1):303–317, 2010.
- [115] Gerard J Steen, Aletta G Dorst, J Berenike Herrmann, Anna Kaal, Tina Krennmayr, and Trijntje Pasma. *A method for linguistic metaphor identification: From MIP to MIPVU*, volume 14. John Benjamins Publishing, 2010.
- [116] Anatol Stefanowitsch. Happiness in english and german: A metaphorical-pattern analysis. *Language, culture, and mind*, pages 137–149, 2004.
- [117] Greg J Stephens and William Bialek. Statistical mechanics of letters in words. *Physical Review E*, 81(6):066119, 2010.
- [118] Tomek Strzalkowski, George Aaron Broadwell, Sarah Taylor, Laurie Feldman, Boris Yamrom, Samira Shaikh, Ting Liu, Kit Cho, Umit Boz, Ignacio Cases, et al. Robust extraction of metaphors from novel data. In *Proceedings of the First Workshop on Metaphor in NLP*, pages 67–76, 2013.
- [119] George J Suci and Percy H Tannenbaum. The measurement of meaning. *Urbana: University of Illinois Press*, 2, 1957.
- [120] Lin Sun and Anna Korhonen. Hierarchical verb clustering using graph factorization. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1023–1033. Association for Computational Linguistics, 2011.
- [121] Joan Swann, Rob Pope, and Ronald Carter. *Creativity in language and literature: The state of the art*. Palgrave Macmillan, 2011.

- [122] Paul Thibodeau and Stephen Flusberg. Metaphorical accounting: How framing the federal budget like a household's affects voting intentions. *Cognitive Science*, Forthcoming.
- [123] Paul H Thibodeau and Lera Boroditsky. Metaphors we think with: The role of metaphor in reasoning. *PloS one*, 6(2):e16782, 2011.
- [124] Tony Veale, Kurt Feyaerts, and Geert Brône. The cognitive mechanisms of adversarial humor, 2006.
- [125] Tony Veale, Ekaterina Shutova, and Beata Beigman Klebanov. Metaphor: A computational perspective. *Synthesis Lectures on Human Language Technologies*, 9(1):1–160, 2016.
- [126] Eva M Vecchi, Marco Marelli, Roberto Zamparelli, and Marco Baroni. Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. *Cognitive science*, 2016.
- [127] Michael S Vitevitch, Paul A Luce, Jan Charles-Luce, and David Kemmerer. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech*, 40(1):47–62, 1997.
- [128] Elijah Wald. *Talking 'bout Your Mama: The Dozens, Snaps, and the Deep Roots of Rap*. Oxford University Press, 2012.
- [129] Graham Wallas. *The art of thought*. J. Cape, 1926.
- [130] Linda R Waugh. The poetic function in the theory of roman jakobson. *Poetics Today*, 2(1a):57–82, 1980.
- [131] Yorick Wilks. A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence*, 6:53–74, 1975.
- [132] Ludwig Wittgenstein. *Philosophical investigations*. philosophische untersuchungen. 1953.
- [133] Fa-Yueh Wu. The potts model. *Reviews of modern physics*, 54(1):235, 1982.
- [134] D Yarlett and Michael JA Ramscar. Language learning through similarity-based generalization. *Unpublished PhD Thesis, Stanford University*, 2008.