

Metaphor, Meaning, Computers and Consciousness

Stephen McGregor¹ and Matthew Purver² and Geraint Wiggins³

Abstract. This paper seeks to situate the computational modelling of metaphor within the context of questions about the relationship between the meaning and use of language. The results of this pragmatic assessment are used as the theoretical basis for a proposed computational implementation that seeks metaphor in the geometry of a vector space model of distributional semantics. This statistical approach to the analysis and generation of metaphor is taken as a platform for a consideration of the fraught relationship between computational models of cognitive processes and the study of consciousness.

1 Introduction

Aristotle is commonly credited as the earliest thinker to seriously consider metaphor as a linguistic device, lauding its use as an indication of the highest level of genius [1]. But while a historical account of scholarship about metaphor is a worthwhile topic, and one which will feature throughout this paper, the history of metaphor itself is as convoluted and unobtainable as the history of language. In fact, if it serves any purpose to think about such a remote event as the inception of language, it seems impossible to imagine a clever speaker not immediately taking the agreed definitions of the world's first words and doing something unexpected with them. If anything, a more accurate take on early academic discussion of metaphor might be to consider Aristotle as one of the first philosophers to ponder the question of the relationship between what words mean and what words do.

This paper will seek to evaluate metaphor from a pragmatic point of view, and to situate this evaluation in terms of a framework for the computational analysis and generation of metaphor. This marks a shift from what has become the standard computational approach to metaphor, which considers language in terms of formalisms that are intuitively compatible with symbol manipulating machines. Implicit in these standard approaches is the assumption that words and concepts exist on different levels of abstraction, and that metaphor is a product of a process of transference or mapping that occurs on the conceptual level, with words acting as a kind of index of this process. But the idea that words merely point to concepts runs into trouble in light of certain properties of metaphor that cannot be explained in terms of an abstract conceptual construct of the entities nominated by words. At the root of the approach proposed in this paper is a contention regarding the difficult topic of consciousness: metaphor is often based on the direct experience of perception, and the ease with which a cognitive agent can express the actual quality of one particular percept in terms of the idea of another general percept is rooted in the direct connection between phenomenology

and language. The very relevance of the term “like” to figurative language, manifest when metaphor is translated into simile, suggests that the “likeness” of the conscious experience of qualia is intrinsic in the perpetually unfolding construction of metaphor.

One of the claims made in this paper is that consciousness is always understood metaphorically, and one of the most pervasive and at the same time disputed contemporary metaphors involving consciousness has been the trope that casts the mind as a computer. This particular construct is compelling, in that the mind can be conceived of as having input in the form of perceptual stimulation and output in terms of either conceptualisation of the world or directed action in the world. At the same time, the analogy is disreputable in its relegation of the richness of consciousness to the domain of a rule following, data processing apparatus that is subject to an arbitrary, observer relative interpretation. It seems that a good model of metaphor should explain the appeal of comparing the engine of its own operation – the mind – to a device that is arguably at best just an aid to thought. The model should also account for the perceptual, imagistic aspect of metaphor-making, evident in light of the necessity of comparing one experience to another when trying to describe what it is like to be conscious.

The solution offered here involves turning to high dimensional representations of meaning based on a statistical analysis of the distribution of words in large scale corpora, and, in so doing, embracing the modelling power of the computer, if not the explanatory power of the mind-computer metaphor. The theory behind the system that will be described is based on the idea that a statistical treatment of a large collection of words found in their natural habitat, so to speak, can simulate the construction of a space of meanings. This space, in turn, becomes the linguistic environment in which metaphors are discovered in the process of solving communicative problems: congruences in the geometries of these statistical word-objects suggest ways in which they can be combined in order to construct expressions. The metaphor-making procedure, modelled as a fundamental aspect of ongoing entanglement with a richly informative environment, is finally presented as a key component in the expression of consciousness, a characteristic that may shed some light on the evident propensity for qualia sensing agents to project their own consciousness onto everything else in the world.

2 Consciousness Is a Metaphor

The tension that metaphor has traditionally introduced to the study of language has arisen from the dynamic between words and truth: figurative statements that are clearly contrary to the facts of reality are nonetheless effective at conveying truthful information about the world. This aspect of metaphor poses at least a superficial problem for truth conditional approaches to semantics, which hold that there is either a correspondence between propositions and the world they

¹ Queen Mary University of London, email: s.e.mcgregor@qmul.ac.uk

² Queen Mary University of London, email: m.purver@qmul.ac.uk

³ Queen Mary University of London, email: geraint.wiggins@qmul.ac.uk

portend to describe [36], or a coherence between the set of propositions that collectively constitute a truthful system of beliefs [13]. For Floridi, the imperative of truthfulness means that “semantic information” is necessarily defined in terms of data that remits “veridicality” in relation to the world that it models [19]. Dretske likewise distinguishes between information and the semantic representation indicated by a correct interpretation of that information [16].

Taking Dretske’s ideas about indication and interpretation as a point of departure, it is possible to formulate a theory whereby the truthfulness of figurative propositions lies in the correct interpretation of the intention behind a non-veridical statement. Here metaphor becomes a mechanism for encoding information, with the projection from source to target allowing for the transference of a set of intensions from a general case of the source to a specific instance of the target. If this is the case, then a metaphor can be deciphered into a more extensive array of literal propositions. The well studied metaphor “that surgeon is a butcher”, for instance, takes the bloodiness and brutality stereotypically associated with the profession of a butcher and efficiently applies them to the behaviour of some disreputable surgeon. This packaging of literal information sits well with Searle’s approach to metaphor, which sees non-literal language as an invitation to interpretation based on propositional knowledge of the world shared between two interlocutors [35]. Ortony, in his “reconstructivist” theory of metaphor, has even suggested that there must be some sort of mental imagery involved in the interpretation of figurative language: a metaphor evokes a non-literal scene which effects the vivid transference of intension in a way that invites logical inference [31]. This move introduces conscious perception to the explication of metaphor, with the experience of a mental state playing a direct role in the transmission of richly detailed information.

But how can consciousness ever be discussed in a way that is literal or veridical? If qualia, with their intrinsically subjective character, are the substance of conscious experiences, then it seems impossible to describe such phenomenological conditions in terms of truthful propositions about situations in the world. Chalmers has made much of this divide between subjective conscious experience and objective physical reality, focusing in particular on the difficulty of determining the truth conditions of a report of a phenomenological perception [9]. From a phenomenological perspective, the defining characteristic of consciousness is that there is something it is like to experience qualia, and this very “likeness” of the experience immediately suggests the application of analogical conceptualisation and correspondingly metaphoric expression. While a mutually agreed description such as “red thing” might allow two interlocutors to pick out a set of objects with some shared characteristic, it is not clear that there is any way to know that the actual phenomenology of the red experience is similarly shared. Since there is no way to expressively project the actual conscious experience of perceiving an object, a descriptive speaker who wishes to convey something phenomenological is left with no choice but to resort to an act of analogy, giving the world such poetic turns of phrase as “lips as red as blood” or “eyes as blue as the sky”.

Along these lines, Everett has highlighted the absence of abstractly quantifiable colour terms in the language used by the Pirahã people of Brazil, who instead employ standardised expressions that are fundamentally figurative: the color term corresponding to what an English speaker would describe as “red”, for instance, transliterates to the expression “bloodlike”, and “black” becomes the phrase “blood is dirty” [17]. Levinson reports similar findings in his analysis of the Yéfi Dnye language spoken by the inhabitants of an isolated island near Australia, who use the terms for various birds and plants

to describe other similarly coloured objects [29]. Even if, as Kay and Maffi claim, the lack of fixed absolute colour partitions in a language is anomalous [27], the admission of chromatic descriptions such as “chartreuse”, “coral”, or “eggplant” in English illustrates the ease with which a perceptual experience of one thing can be converted into a classification of something else. There is an inherent process of analogising occurring when cognitive agents turn to language to express the subjective characteristics of their perceptual existence.

This perpetual trafficking of intension from one perceptual or conceptual domain to another extends especially into more general descriptions of consciousness. The difficulty of discussing qualia in objective and material terms has compelled philosophers to resort time and again to thought experiments involving components fantastically removed from reality – beetles in boxes, homunculi in theatres, deceptive demons – in order to allude circumspcctly to what it is like to be conscious. Even Dennett, who has questioned the efficacy and indeed the existence of qualia [14], acknowledges that it is generally necessary to employ analogical reasoning when dealing with descriptions of mental processes [15].

There is a temptation to take the necessity of analogy in discussion of consciousness one step further by way of construing consciousness itself as a process of metaphor-making. In the 1970s, Jaynes proposed his bold “bicameral” theory of mind based on the idea that pre-literate humans had perceived their own consciousness as a mentally external expression of instructions and proclamations experienced as ongoing auditory hallucinations [26]. To a mind sundered in such a way, the modern experience of self as realised through subjective phenomenology was supposedly replaced with a personal fictive narrative that cast the consciously feeling component of the mind in the role of a god or a commanding spirit. This controversial theory has received some recent support, at least implicitly, in Carruthers’ formulation of “interpretive sensory-access” based on the mindreading faculties that facilitate the acts of interpretation at the centre of consciousness [6]. In a propositional reversal that nonetheless maintains some of the core tenets of Jaynes’ bicameral mind, mindreading capacities can be applied not only to introspection, but also to the interpretation of the mental states of other people and even as the projection of mind-like faculties on objects that are obviously actually inanimate. So, for instance, it seems quite reasonable to metaphorically discuss the temperament of things like computers, cars, appliances, or the weather without the presumption that these types of objects actually have minds.

If these projective theories of mind are to be taken seriously, then the essential role of metaphor in consciousness must be considered. There certainly seems to be a case to be made for the idea that consciousness necessarily involves a transgression of literal conceptualisation of the world, a transference of a feature from one mental object to another that results in an expression of the experience of a thing as something other than what it actually is. There are three propositions at stake here. The first is that the only feasible mechanism for communicating about the experience of consciousness is to cast the description of that experience out onto some universally accessible entity with qualitative attributes that will hopefully simulate the experience. The second is that the mind can only be understood in terms of things other than minds, things that have mind-like properties and therefore analogically corroborate an explanation of what it is like to have a mind. The third is that having a conscious mind necessarily involves the projection of phenomenological characteristics onto external entities, some that presumably are likewise conscious and others that almost certainly are not. In each of these cases, through experiential transference, through analogical descrip-

tion, and through projection of the self onto another, an essentially metaphoric process is at play: knowledge of the mind seems to consist of a network of proxies and equivalences that trace the outline of the thing that they don't quite touch.

3 Words Are Objects

The recent history of theoretical approaches to metaphor has been characterised by an intellectually productive tension, with both sides notably departing from any notion that the figurative use of language should somehow be treated as an exceptional case. On the one hand, there are those who would describe metaphor as a transference or projection of intensionality from the conceptual space of a source to the similarly oriented space of a target, a view that found an early champion in Black and his "interactionist" theory of metaphor [3, 4]. By this account, metaphor involves conceptual mappings that place a non-literal source at the centre of the "implicative complex" of a targeted conceptual system, so that characteristics of the way the source does things are projected onto similar activities undertaken by the target. On the other hand, a dissenting contingent of theorists have argued that the metaphoric use of language stands entirely outside the realm of conceptualisation, and that the meaning of any sentence can only be interpreted literally—an idea originally expounded by Davidson [12], with early support coming from Rorty [34].

In the early 1960s, Hesse argued for the importance of analogy as a tool for scientific understanding [25]. At the root of her argument was the idea that all theories are ultimately models of the world, and that, in terms of the extreme scales involved in, for instance, the study of physics, these models could only be grasped in terms of metaphors: so, for instance, a distributed gas bears an analogy with a space full of colliding and rebounding balls. The study of metaphor subsequently underwent a Renaissance of sorts, with a flurry of research throughout the 1970s (see [32] for a compendium of exemplars), culminating in Lakoff and Johnson's case for an understanding of metaphor as a mapping between isomorphic conceptual schemes [28]. This theory presented metaphorical language in terms of its relationship to an embodied cognitive experience of the world, so, for instance, the analogy which maps the conceptual situation between "up" and "down" to the situation between "happy" and "sad" is a product of the actual culturally loaded experience of orientation in the real world. A lattice of networked spaces, extending from the world through perception and conceptualisation into language, allowing for the transference of entire isomorphic conceptual complexes: if a surgeon is a butcher, then hospitals become abattoirs and patients become animals.

Davidson, however, offered a dissenting interpretation of metaphor, springing from his rejection of the idea that language should be talked about as a system for conceptual representation in the first place [11]. Instead, he proposed that the meaning of a metaphor could only be considered in terms of the literal proposition made by a metaphorical statement, and that the operation of a metaphor in the process of communication must be considered as something altogether outside the realm of meaning [12]. This stance has met with considerable resistance, finding an early opponent in Bergmann, who argued that Davidson's critique only applied to decontextualised encounters with metaphor; once the metaphor is put into the context of a situation involving a speaker with an intention, it can be clearly seen to have a meaning [2]. Hesse also revisited her case for metaphor as a fundamental cognitive operation, arguing that all language is metaphoric in that all language plays a protean role in a nebulous network of meaning [24]. Rorty, on the other hand, came

to Davidson's defence, interpreting his approach as placing metaphor actually in the world of natural events rather than consigning it to an essential role in an interplay of symbols that is ancillary to reality [34]. By this reading, language is not to be considered as a model or representation of reality, but rather as a component directly in reality, existing on the same level of abstraction as impressions and ideas.

The debate over metaphor in subsequent years has involved a back and forth between those who see metaphor as by-product of an essential cognitive operation and those who claim that language plays a more fundamental role in perception of the world, though Davidson has arguably been broadly misinterpreted. In an expansive consideration of metaphor as evidence of "the poetic structure of mind", Gibbs suggests that Davidson places emphasis on first determining the literal meaning of a metaphor and then accepting that the potential non-literal meanings of the phrase are somehow infinite and unknowable [22], perhaps a misreading of Davidson's contention that "there are no unsuccessful metaphors". As a recent proponent of the non-cognitive take on metaphor, though, Carston has recast Davidson's rejection of cognitive content in terms of a more fundamental "imagistic" feature of language [7]. In particular Carston considers the metaphor "Bill is a bulldozer": the interpretation of this phrase as a description of a man who is grossly aggressive and inconsiderate is clear, but upon further analysis there is no literal property of a piece of equipment such as a bulldozer that bears the inherently human intensions being drawn out in Bill [8]. At best there might be an argument that a double metaphor is being employed here, with a bulldozer standing in for something aggressive and then Bill being described as one of those things, but this introduces a combinatorial explosion of ways to frame all but the simplest metaphors and in so doing seems to miss the point of the cogency of figurative language. Instead, it seems reasonable to say that the metaphor evokes something that is not purely in the realm of language, a direct perception of Bill as a potentially destructive machine.

In this analysis, Davidson and his acolytes emerge as something of the arch-pragmatists. Rather than keeping the construction and interpretation of metaphor on a symbolic level, where language models the world it describes, here the very meanings of the words employed in a metaphor become implements to be handled and used to accomplish communicative goals in the same *ad hoc* way that a more overtly physical object might be picked up and used. Meanings exist, but as the features of elements of language that suggest their functionality: in fact, the meanings of words themselves become the intensions of those words, suggesting potential uses of language in the way that, for instance, the solidness and heaviness of an object might recommend it as a weapon to an attuned perceiver in need of such a device. Just as a shoe might present itself as a hammer under the right circumstances, or a stick or rock as a writing instrument, the word "bulldozer" offers itself as the right term to convey Bill's comportment in the same grasping process of perception and cognition, because language is actually happening on exactly the same level as the rest of existence, not in an abstract secondary space.

At this point, language can be situated in the context of Gibson's theory of affordances, which holds that cognition arises in the process of the perception of opportunities for action in an environment [23]. Clark has worked towards expanding environmentally situated approaches to cognition into the domain of linguistics, describing the "persisting but never stationary material scaffolding" of language [10]. A picture emerges of language use as a process of scavenging a shifting space of meaning for the words that can be used to accomplish some expressive task. These meanings are not representational models that stand in a relationship of signification to perceptions and

conceptions of the world; they are the cognitive detritus of entanglement in an environment that involves communication with other linguistic agents, sitting right alongside other mental experiences of reality.

So an alternative approach to modelling metaphor emerges, one that does not involve considering the language involved in metaphor-making as simply a corollary to mappings between isomorphic conceptual spaces. Instead, metaphor can be envisioned as a process of searching a space of linguistic percepts for the sounds or symbols that can be arranged to fulfil some communicative requirement. The challenge then becomes defining this space of meanings and understanding how word-objects are selected from it. This theory does not refute the descriptive power of Lakoff and Johnson's ideas about conceptual metaphors; in fact, it seems clear that there must be some discernible aspect of meaningful entities that allows them to be cobbled into a pragmatically efficacious structure, and it seems likewise reasonable to construe this act of construction as an aligning of mental objects. As an explanatory device, though, the idea that metaphoric language simply corresponds to congruent concepts seems, upon closer analysis, insufficient.

Hesse's quip about all language being metaphoric also follows from this revised approach: all language use involves grabbing meanings that present themselves as functionally appropriate for the communicative act at hand, and, while some constructions may challenge interpretation more than others, there is no clear reason to draw a definitive line between the literal and the figurative use of meaning. The ubiquity of metaphor takes on a more distinctly Peircean character, though, when word-objects are recognised as existing in the same cognitive space as other percepts. Peirce's claim that all thought is realised through signs [33] seems of a piece with Davidson's pragmatic approach to metaphor once the difference between considering objects as symbols of the mind versus considering symbols as objects of cognition becomes a relatively minor point of contention. To Peirce, reality was a lattice of ubiquitous signification, with meaning manifesting itself through a "life in signs", by which all thought results from the inherently interpretable interplay between things, and all physical interactions are characterised by this kind of life. The perpetual life cycle of event, perception, and interpretation means that signs are always exploding outward from the thing that they signify, becoming themselves the object of a further signification in the instant of their interpretation, even as the interpretation becomes a sign of the thing it interpreted. This endless sequence of becoming something else, accomplished by means of the transformative faculties of symbols, points to a fundamental and enduring process of metaphor-making in the experience of existence.

And here consciousness re-enters the consideration of metaphoric language: consciousness as the thing that can only be objectively grasped through metaphor, or metaphor as the mechanism that facilitates the subjective experience of consciousness. By Peirce's account, the world is conscious, an audacious asseveration that nonetheless lines up well with the idea that being conscious involves the perpetual invocation of the fundamental metaphor that everything else is conscious, as well. If the Peircean variety of panpsychism is perhaps a bit strong, a consideration of the metaphoric nature of individual consciousness at least offers an explanation of why the rest of reality would seem that way, as well. In fact, in accepting that language is wrapped up in a pragmatic process of meaning-grasping, and that all use of word-objects is essentially a ready-to-hand encounter with linguistic percepts, the experience of perpetual metaphor and therefore of imminent and ubiquitous consciousness becomes a less alarming outcome.

4 Meaning Is Geometric

Computational models of metaphor have tended to embrace symbolic approaches that treat language as a representation of cognitive content. As a first approximation, this is not unreasonable, given that computers are symbol manipulating machines: a semantic formalism is precisely the kind of information processing model that is tractable to such a machine. Consequently, van Genabith has found success programming computers to analyse metaphors using type theoretical constructs where source and target both belong to a supertype by virtue of their joint properties, and the intensions transferred by the metaphor are categories specific to the type of the source [38]. Similarly, Veale has built a computational system that handles analogies in terms of "dynamic types" culled from prefabricated conceptual networks such as WordNet [39]. And Gargett and Barnden have described a metaphor generating system that applies information processing instructions to conceptual structures [21], in this case implemented through the contextually sensitive typed schema of Feldman's embodied construction grammar [18].

These kinds of systems treat words as indices to concepts, where the logical structure of concepts can be aligned so as to indicate the affiliated linguistic expression that conveys the projection of properties from source to target. In this way, they are implementations of the conceptual approach to metaphor outlined by Lakoff and Johnson: their success stems from their recourse to abstract representations of concepts, and language is treated as a kind map of the mappings inherent in the dynamics of the conceptual space, metaphoric precisely because of the analogical aspect of cognitive content. In the case of Veale's system, the conceptual schema are, compellingly, built in an *ad hoc* way, even if this ongoing construction is based on a pre-established network. With Gargett and Barnden's system, the underlying formalisms are specifically designed to contextualise conceptual representations in terms of the physical world. By the same token, though, these models are intrinsically committed to the cognitive-content approach to metaphor, treating language as a secondary feature merely pointing to the world model of a conceptual space.

It is not clear how such a system could, for instance, model the direct imagistic experience of perceiving an aggressive person as a bulldozer. The inescapable figurativeness of consciousness, that property by which there is a bulldozer-like quale in the encounter with this unpleasant individual, is lost to a system that depends on conceptual constructs removed from encounters with the percepts – the language and the imagery – that become the symbolic index to those concepts. If the project of computationally modelling metaphor is to be pursued further, it seems necessary to formulate a way in which a space of meanings can be constructed directly from an encounter with language in the world, based on the actual statistical features of the language rather than on predetermined rules regarding the processing of symbols. But how can a computer go about realising this kind of language model?

In fact, symbol manipulating machines seem like exactly the right tools for engaging with this task, and a viable methodology already exists in the form of ongoing work on vector space models of distributional semantics. This approach to language modelling involves the geometric representation of words as points in a high dimensional space [40]. Words are construed as vectors, with the dimensions of these vectors corresponding to the contexts in which a word is likely to occur: in the most straightforward implementation, a dimension of a word-vector corresponds to a term, and the scalar value of that dimension indicates the likelihood of the word co-occurring with

that particular term. When the co-occurrences of the words found throughout a large scale corpus are computed, the result is a space in which the proximity of word-vectors to one another corresponds to the similarity of the contexts in which those words have been found. The intuition behind work in this direction has been that words that are found in a similar context will naturally be likewise semantically similar [37].

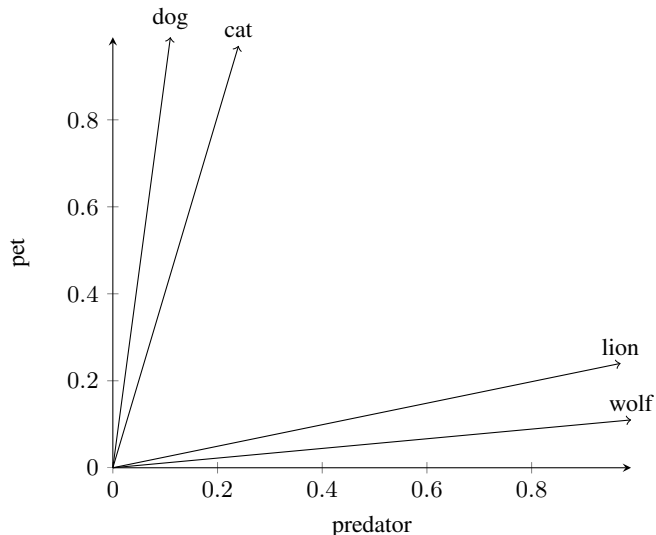


Figure 1: In a highly simplified (two dimensional) vector space model, the words “dog” and “cat” are seen to be semantically relatively similar by virtue of their frequent co-occurrence with the term “pet”, whereas “wolf” and “lion” are more likely to occur in the context of the term “predator”.

Furthermore, the mathematically tractable properties of a geometric space have been exploited in the modelling of compositionality, with linear algebraic operations between word-vectors producing statistical structures corresponding to the meaning of larger segments of language [5, 30]. A similar intuition can be applied to the construction of metaphor, though with the philosophical caveat, informed by Davidson’s take on metaphor, that, where meaning applies to the space of words, the compositions constructed from this space are properly understood only in terms of their use in acts of communication. All the same, it is the geometry of the space of words that suggests ways in which sets of meanings can be pragmatically constructed as metaphors: if proximity corresponds to similarity, then regional clusters of related terms should be discoverable within the vector space. Moreover, the relationship between the terms within such a space indicates a particular geometry, and a congruence in the configuration of terms between two regions might be interpreted as an indication of a potential metaphor. So, for instance, the constellation of word-vectors indicated by the sequence $\{surgeon - patient - hospital - scalpel\}$ would be expected to line up with the shape described by $\{butcher - animal - abattoir - cleaver\}$.

Underwriting this statistical prediction is the theoretical intuition that the way in which a computer encounters symbols in a corpus stands in a synecdochical relationship to the way in which a cognitive agent encounters percepts – including linguistic symbols – in an environment. The hope is that treating large scale corpora as a kind of native habitat for computers serves as a more veridical simulation of the process by which cognitive agents directly grapple with lan-

guage in the physical world than does the construction of abstract conceptual representations. Just as an agent maintains a shifting lexicon of meaning based on a continuous entanglement with language percepts, a computer can establish a network of relationships based on the statistics of its ongoing encounter with symbols in a textual environment. The statistical relationship of words learned by a corpus traversing computer becomes its knowledge base, its space of meanings that can be invoked in a disengaged way when the definition of a particular term is sought, but that at the same time stand ready-to-hand waiting to be grabbed as affordances in the construction of communicatively effective language. When the moment comes for the system to compose an expression, it grasps for the combination of terms that fulfil the required criteria, and these criteria are specifically modelled in terms of the geometric alignment of regions within the space of meanings.

Some preliminary work has been done exploring the relationship between established conceptual metaphors construed in terms of the arrangement of their components within a vector space model, comparing, for instance, the region of butchery to the region of surgery, or the region of orientation (e.g. $\{up - down - in - out\}$) to the region of emotion (correspondingly $\{happy - sad - inclusion - loneliness\}$). Early results have invited cautious optimism: the geometry of the compared vector regions has remitted a high degree of congruence in the anticipated alignments. Future research will have to examine the way in which regions of vectors, corresponding to the construct of conceptual spaces [20], can be defined within a vector space, and this direction of inquiry will in all likelihood motivate a close consideration of the techniques employed in the construction of the vector space itself, as well. The prospective outcome of this project is a system that will use corpus analysis to facilitate a program outputting novel and useful metaphors based on inputs that are perceived as being relatively literal.

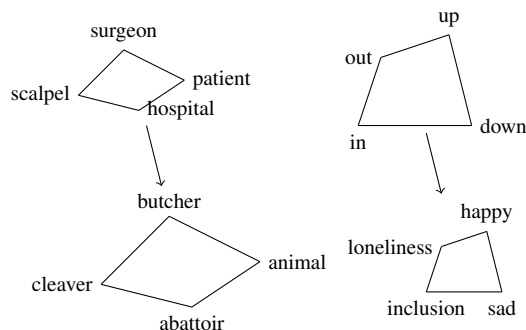


Figure 2: Congruences discovered in subregions of a vector space model suggest metaphoric mappings. The regions do not necessarily have to be of the same scale in order to identify a possible alignment.

5 Conclusion

In weighing the merits of considering the use of words as distinct from the meaning of words, it is worthwhile to observe the extreme ease with which people produce and digest figurative language: metaphor is so universal that almost nothing makes sense if it is taken absolutely at face value. Such a linguistic environment might appear particularly hostile to so formal and literal an agent as a computer. It would seem that the relationship between language and the situations described by language is much messier than some semantic

formalisms would suggest, and the role that meaning plays in the process of communication cannot be easily situated in a denotational relationship to some sort of mental content outside of language. In order for a computer to have a chance in a scenario where all language is open to interpretation, it is necessary for the information processing system to have recourse to its own semantic constructs, and these naturally take the form of statistical interpretations of the bearing of words in their compositional contexts.

Using a computer to model the pragmatic dynamics of metaphor reveals nothing about how consciousness works or why consciousness exists. In this regard, the most that can be said about the system described in this paper is that it attempts to simulate a process with which consciousness is concerned—and this much is true of any computer program that presents data in a way that is designed to be interpretable to a conscious user. Nonetheless, the project of constructing a metaphorical framework within a symbol manipulating system takes on added resonance when considered in the scope of the ineluctably analogical modality of the understanding of the conscious mind. Even if the model that has just been proposed doesn't shed any light on the nature of consciousness, it does address some of the questions about the linguistic operation involved in conceptualising consciousness. It is the very ineffability of consciousness that forces a philosopher to resort to analogy and metaphor when discussing this hard topic and indeed when describing the experience of it. In understanding the construction of metaphor as a utilisation of meaning towards the goal of expression, it becomes clear how a cognitive agent must be constantly involved in this operation, always grasping for the combination of meanings that work when put out into the world as the communication of a mental experience. In the process of constructing the sustained sense of self at the core of a conscious experience of the world, a cognitive agent must necessarily cast the idea of the self out into the world to reflect upon it; it is only natural, then, that an essential feature of consciousness should be to imagine that other things are conscious.

So this pragmatic reconsideration of metaphor and the computational implementation of the redesigned model offer at least the beginning of an explanation for the mind's propensity to figuratively project its own consciousness onto the entities that it encounters in the world. This final observation regarding the relationship between metaphor and consciousness can be turned into a possible stance in the debate regarding the controversial construct that reimagines the mind as a computer: if anything, it is the mind that projects consciousness onto the computer, not the computer that stands in as a model for what the mind does. The conceit of the mind as computer seems to easily forget that the operations of a computer are only meaningful by virtue of the values assigned to its inputs and outputs by some agent who is plugged into reality in a deeply intentional way—but then the mysteriousness of consciousness likewise evades the question of what exactly it is that is doing the conscious sensing, leaving only the fanciful notion that all nature of other things can consciously sense, as well. And so in the end, the metaphor of the mind as a computer is perhaps actually just a reversal of the metaphor of a computer as a kind of mind, a lending out of the self which is actually just a specific case of what conscious minds, in their incessant and incurable projecting, do to everything in the world.

ACKNOWLEDGEMENTS

This research has been supported by EPSRC grant EP/L50483X/1.

REFERENCES

- [1] Aristotle, *The Poetics*, Macmillan and Co, London, 1895.
- [2] Merrie Bergmann, 'Metaphorical assertions', *The Philosophical Review*, **91**, 229–245, (1982).
- [3] Max Black, 'Metaphor', in *Proceedings of the Aristotelian Society*, volume 55, pp. 273–294, (1955).
- [4] Max Black, 'More about metaphor', in *Metaphor and Thought*, ed., Andrew Ortony, 19–41, Cambridge University Press, 2nd edn., (1977).
- [5] William Blacoe and Mirella Lapata, 'A comparison of vector-based representations for semantic composition', in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 546–556, (2012).
- [6] Peter Carruthers, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford University Press, 2011.
- [7] Robyn Carston, 'Metaphor: Ad hoc concepts, literal meaning and mental images', in *Proceedings of the Aristotelian Society*, volume CX, pp. 297–323, (2010).
- [8] Robyn Carston, 'Metaphor and the literal/nonliteral distinction', in *The Cambridge Handbook of Pragmatics*, eds., K. Allan and K. M. Jaszczolt, Cambridge University Press, (2012).
- [9] David J. Chalmers, *The Conscious Mind*, Oxford University Press, 1996.
- [10] Andy Clark, 'Language, embodiment, and the cognitive niche', *Trends in Cognitive Sciences*, **10**(8), (2006).
- [11] Donald Davidson, 'On the very idea of a conceptual scheme', in *Proceedings and Addresses of the American Philosophical Association*, volume 47, pp. 5–20, (1974).
- [12] Donald Davidson, *Inquiries into Truth and Interpretation*, chapter What Metaphors Mean, Clarendon Press, Oxford, 2nd edn., 1978.
- [13] Donald Davidson, 'A coherence theory of truth and knowledge', in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed., Ernest LePore, (1986).
- [14] Daniel C. Dennett, *Consciousness Explained*, The Penguin Press, London, 1991.
- [15] Daniel C. Dennett, *Intuition Pumps and Other Tools for Thinking*, W. W. Norton and Company, 2013.
- [16] Fred I. Dretske, *Knowledge and the Flow of Information*, CSLI Publications, 1981.
- [17] Daniel L. Everett, 'Cultural constraints on grammar and cognition in pirahã: Another look at the design features of human language', *Current Anthropology*, **46**(4), 621–646, (2005).
- [18] Jerome Feldman, 'Embodied language, best-fit analysis, and formal compositionality', *Physics of Life Reviews*, **7**(4), 385–410, (2010).
- [19] Luciano Floridi, *The Philosophy of Information*, Oxford University Press, 2011.
- [20] Peter Gärdenfors, *Conceptual Space: The Geometry of Thought*, The MIT Press, Cambridge, MA, 2000.
- [21] Andrew Gargett and John Barnden, 'Gen-meta: Generating metaphors using a combination of ai reasoning and corpus-based modeling of formulaic expressions', in *Proceedings of TAAI 2013*, (2013).
- [22] Raymond W. Gibbs, Jr., *The Poetics of Mind*, Cambridge University Press, 1994.
- [23] James J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston, 1979.
- [24] Mary Hesse, 'The cognitive claims of metaphor', *The Journal of Speculative Philosophy*, **2**(1), 1–16, (1988).
- [25] Mary B. Hesse, *Models and Analogies in Science*, Sheed and Ward, New York, 1963.
- [26] Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Penguin Books, 1976.
- [27] Paul Kay and Luisa Maffi, 'Color appearances and the emergence and evolution of basic color lexicons', *American Anthropologist*, **101**(4), 743–760, (1999).
- [28] George Lakoff and Mark Johnson, *Metaphors We Live By*, University of Chicago Press, 1980.
- [29] Stephen C. Levinson, 'Yéli dnye and the theory of basic color terms', *Journal of Linguistic Anthropology*, **10**(1), 3–55, (2001).
- [30] Dmitrijs Milajevs, Dimitri Kartsaklis, Mehrnoosh Sadrzadeh, and Matthew Purver, 'Evaluating neural word representations in tensor-based compositional settings', in *Proceedings of EMNLP 2014*, (2014).
- [31] Andrew Ortony, 'Why metaphors are necessary and not just nice', *Educational Theory*, **25**(1), 45–53, (1975).

- [32] *Metaphor and Thought*, ed., Andrew Ortony, Cambridge University Press, 2nd edn., 1993.
- [33] Charles Sanders Peirce, *Peirce on Signs*, The University of North Carolina Press, Chapel Hill, NC, 1991.
- [34] Richard Rorty, 'Unfamiliar noises: Hesse and Davidson on metaphor', *Proceedings of the Aristotelian Society*, **61**, 283–296, (1987).
- [35] John R. Searle, 'Metaphor', in *Metaphor and Thought*, ed., Andrew Ortony, Cambridge University Press, (1979).
- [36] Alfred Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*, Hackett, Indianapolis, IN, 2nd edn., 1983. Translated by J. H. Woodger.
- [37] Peter D. Turney and Patrick Patel, 'From frequency to meaning: Vector space models of semantics', *Journal of Artificial Intelligence Research*, **37**, 141–188, (2010).
- [38] Josef van Genabith, 'Metaphors, logic and type theory', *Metaphor and Symbol*, (2001).
- [39] Tony Veale, 'Dynamic type creation in metaphor interpretation and analogical reasoning: A case-study with wordnet', in *Conceptual Structures for Knowledge Creation and Communication*, 146–159, Springer, (2003).
- [40] Dominic Widdows, *Geometry and Meaning*, CSLI Publications, Stanford, CA, 2004.