

*To appear in Artificial Intelligence Review, Volume 8, 1995*

## **Sandy ideas and coloured days: Some computational implications of embodiment**

Ronan Reilly  
Department of Computer Science  
University College Dublin  
Belfield, Dublin 4  
Ireland  
`rreilly@nova.ucd.ie`

**keywords:** language evolution, metaphor, synesthesia, cortical computation

### **Abstract**

This paper is an exploration of the relationship between language and vision from the perspective of language evolution on the one hand, and metaphor on the other. Recent research has suggested that the origins of human language capacity can be traced to the evolution of a region in the brain that permits the interaction of information from sensory and motor cortices. In light of this, it is hypothesised that the computational mechanisms of language are derived from those of the sensory-motor domain, and that the pervasiveness of metaphor is one manifestation of language's computational antecedents. A variety of cognitive and computational implications are drawn from these hypotheses.

# Sandy ideas and coloured days: Some computational implications of embodiment

Ronan Reilly  
Department of Computer Science  
University College Dublin  
Belfield, Dublin 4  
Ireland  
rreilly@nova.ucd.ie

## Introduction

When I was an undergraduate I remember reading a book by R. L. Gregory (Gregory, 1974) which contained a chapter entitled "The Grammar of Vision." It discussed the possibility that language might have its roots in visual perception, and described parallels that existed between vision and grammar. Visual perception, Gregory maintained, involved "parsing" the sensory input using a lexicon of features into meaningful "sentences". The process was driven by a set of "grammatical" rules which could be violated in the case of so-called impossible objects, such as those devised by the artist M. C. Escher. Gregory suggested that this rule-based parsing aspect of vision may have been exploited during the evolution of language. I found Gregory's paper very exciting, having previously been puzzled by the evolutionary recency of language and not persuaded by those who argued that it had arisen from a serendipitous mutation. Here was a proposal that the computational machinery underlying language could also be found in a more ancient perceptual system. Nonetheless, as I will argue later, by seeking a connection between language and vision, and indeed between language and other sense modalities, we are forced to look at the computational mechanisms<sup>1</sup> underlying language in a rather different light than that initially envisaged by Gregory. His view of this relationship was very much language-centred, heavily influenced by the work of Chomsky and particularly by the latter's distinction between deep and surface structure (Chomsky, 1967). I hope to show in this paper, however, that an exploration of the possible sensory and motor roots of language allows us to see its underlying mechanisms in a way that is fundamentally at odds with the view of language articulated by Chomsky.

---

<sup>1</sup>When I refer to computational mechanisms, I intend no commitment, at least initially, to any specific computational style, be it classical information processing or connectionist.

## Evolution of Language

If Gregory's hypothesis is at least broadly correct, then how and why did the machinery of sensation and perception suddenly become utilised in this radically new way? Wilkins and Wakefield (in press) argue that the neural preconditions for language evolved slowly and that language emerged abruptly only when the brain achieved the appropriate internal configuration. They argue that language did not evolve directly from communicative-based precursors. They disagree with the contention of, for example, Lieberman (1992), that language arose as a result of selective pressures for improved speech capacity. In fact, Wilkins and Wakefield maintain that language capacity emerged 2 million years before the development of a modern vocal articulatory system. Instead, they propose that the relevant internal configuration giving rise to language was the emergence of regions of cortex in which information could be represented amodally, independent of any particular sense modality. Such regions came about, they argue, because of the need to coordinate information from different sense modalities in, for example, the performance of complex eye-hand coordination tasks. In particular, they arose as a result of the hominid shift to bipedalism and the resultant potential for the development of novel manipulative abilities. Wilkins and Wakefield identify one region in particular, the parieto-occipito-temporal (POT) junction, which Geschwind (1965) has called the "association area of association areas," as being implicated in the emergence of language. A central argument in Wilkins and Wakefield's thesis is that the POT region did not evolve to serve as the processing base for language, but rather was reappropriated for language, having evolved for visual-somasthetic-motor coordination. It will be argued in the rest of the paper that a better way of conceptualising the role of the POT is as a region of multi-modal rather than amodal interaction, and that this multi-modality is a core computational mechanism for language processing. The amodal/multi-modal distinction may seem a little artificial, since for modalities to interact there needs to be some form of common basis for interaction. What I am trying to emphasise here is that the modalities interact without sacrificing their essential sensory nature. Sensory and motor areas project to the POT, but there is no explicit translation of the sensory-motor information into some kind of neural *lingua franca*. In fact, the very form that this information takes, neuronal firing patterns, is itself a common cortical "language."

## Categorisation

One of the main points of contact between perception and language is in the area of categorisation and naming. Objects or their attributes, their colour for example, tend to have names in a language. But the relationship between language and perception is not a clear cut one. In the case of colours, for example, the Dani language of New Guinea has just two names for all colours: one for dark and cool colours, and one for light and warm colours (Rosch, 1973). In a study aimed at refuting Whorf's hypothesis that language determines an individual's conceptual system, Rosch demonstrated that Dani speakers had the same ability as speakers of languages with richer colour vocabularies to manipulate colour concepts. Berlin and Kay (1969) showed that there is a set of focal colours that for biological reasons people find more salient than others. Different languages may place the boundaries of colour terms in different places, but when speakers of these languages are asked to indicate the colour that is most typical of a particular category, they will choose a focal colour rather than a colour at some point in the middle of the region of the spectrum encompassed by the category. In the study of Dani carried out by Rosch (1973), she found that Dani speakers could, for example, learn artificial names for focal colours more readily than non-focal colours. The picture one gets from this and other studies is of a conceptual system rooted in perception and subject to the biological constraints of the perceptual system, and of a language system that neither determines, nor is determined by, the nature of this conceptual system. So, for example, the perceptual saliency of focal colours does not automatically entail their lexicalisation; with the corollary that the failure to have a name for a colour does not make it conceptually unusable or inaccessible.

Another example of perception interacting with language is in the use of numeral classifiers of the type found in many Asian languages. These are morphemes used in counting which are associated with certain features of the object being counted. In Japanese, for example, the classifier *hon* is used to count objects that are long and thin. So knives might be counted with *hon*. However, if knives are being counted in the context of kitchen implements, the speaker might want to draw attention to this and use *choo* instead. In English, some words are used in a similar fashion. For example, the words *head* and *pairs* in the phrases "two head of cattle" and "three pairs of trousers" perform a similar function to Japanese numeral classifiers.

Numeral classifiers seem, on the surface, a straightforward linguistic system that captures some subset of the perceptual features of the objects being counted.

However, their perceptual characteristics capture only a limited aspect of their use in Japanese. *Hon*, for example, may not only be used to count long thin objects, but also such abstract entities as baseball home runs, telephone calls, and films. The whole of the complex classifier system (there are approximately 150 classifiers, with 27 in common use) is founded on a perceptual base, but undergoes complex elaboration and extension beyond the perceptual domain. The process of elaboration can also be found in the pattern of acquisition of the classifiers by Japanese children; with the more concrete and perceptually anchored usages acquired first, and the more abstract ones later. Furthermore the social context in which the language is acquired has an impact on the pattern of usage. Inoue (1993) has shown, for example, that the pattern of acquisition and usage of numeral classifiers among Japanese children growing up in the US differs from that of Japanese children growing up in Japan. She has shown that these differences emerge as a result of differences in the cultural milieu of the two sets of children. Lakoff (1989) describes an even more dramatic pattern of change in the classifier system of those members of the Australian aboriginal tribe the Dyirbal who have lived in an urban environment for a number of generations. Separated from the cultural and environmental context in which their classifier system evolved, the system atrophies over one or two generations.

To summarise this section, an examination of the interface between perception and language reveals a rather complex picture. On the one hand, the colour naming evidence suggests that the conceptual and language systems, while intermeshed are not interdependent. For example, people are able to manipulate colour concepts without the benefit of a narrowly applied name for a particular colour; in other words, there can be conceptualisation without lexicalisation. On the other hand, the numeral classifier evidence suggests that lexicalisation is no guarantee of conceptual stability, even of a fairly concrete concept such as the class of long thin objects. There is a fluidity and flexibility that is belied by more conventional views of language.

This process of a perceptually anchored system being extended metaphorically or metonymically to other more abstract domains is regularly encountered in a variety of language phenomena, not just numeral classifiers. It seems clear that the mechanism facilitating such extension is central to how we acquire and process language. Yet for a long time it has been treated as a marginal issue. However, a number of influential theorists now argue (notably, Lakoff, 1989) that this process of extension and elaboration is a manifestation of a mechanism that lies at the core of language processing and cognition.

## Cognitive Linguistics

Building on his work in categorisation, Lakoff (1989) has defined the area of *Cognitive Linguistics* as being concerned with a re-evaluation of the traditional approaches to linguistics and cognitive science. Lakoff challenges the now traditional generative linguistic view of language as a formal system, comprising independent syntactic and semantic components, where the semantics is either model-theoretic or involves a translation of a set of symbols into some logical form or semantic representation. In the latter case, the elements of the logical form stand in a correspondence with external reality. This view of language pre-supposes what Lakoff refers to as "the objectivist approach to cognition." Within this framework thought is defined as the manipulation of abstract symbols. Symbols get their meanings via correspondences to entities and categories in the real world. Thought is a process independent from those doing the thinking. Its nature is in no way determined by the organism in which it is embodied. In contrast, Lakoff argues for a cognitive model approach to cognition and a cognitive grammar for language. In fact, he does not see cognition and language as separate processes, but rather as related and interacting in the sense that they share the same basic mechanisms. Lakoff views language and cognition as an organic, embodied system; intimately connected to the physical nature and sensory repertoire of people, and to the cultural and societal context in which they move.

Within cognitive linguistics, metaphor and metonymy play a central role, rather than the peripheral one given them in conventional generative grammar. The basic building blocks of language and cognition are anchored in the perceptual experience of the language user and cogniser; more abstract concepts are constructed by means of metaphorical or metonymic extension to these basic concepts.

Metaphor is much more pervasive than one might at first suspect, spreading far beyond the medium of poetry and prose. A key distinction that conventional linguists make is between *literal* and *figurative* language. Conventional linguists overwhelmingly view literal meaning as primary, and would see the computation of literal meaning as the first step in understanding a sentence. Only if this fails for some reason is the figurative meaning of an utterance examined. However, as Rumelhart (1979) has pointed out, the literal/figurative distinction disappears if we look at it from a psychological point of view. Figurative language is no more burdensome than literal language in terms of the psychological processing demands it makes; whereas, given its "second-order" status, one might have expected its use to be more complex, and for this complexity to manifest itself in psychological data.

Furthermore, child language acquisition data indicates that figurative language is used from a very early stage. Again, given its perceived complexity from a conventional linguistics point of view, one would expect it to appear developmentally later.

The ability to process information metaphorically is surely one of the facilities that the POT of Wilkins and Wakefield (in press) provided to the early hominids. Essential to metaphor is an ability to compare the structure of concepts across modalities. To permit this, some form of inter-modal representation is essential. I will argue in the following sections, however, that far from providing an amodal arena in which sensory information can interact, as Wilkins and Wakefield have suggested, the POT provides a mechanism for multimodal interaction. Thus allowing sensory and non-sensory information to interact, and permitting the computational algorithms of perception to be pressed into the service of language processing.

### **Sensory-Motor Ideas**

The notion of a computational relationship between sensory and motor processes on the one hand and language and cognition on the other is not new. It formed the centrepiece of the developmental theories of the Swiss psychologist Jean Piaget. He argued for a developmental path from sensory motor behaviour in infants through to the internalisation of action-schema, and ultimately to the development of what he referred to as formal operations, which can best be described as abstract logical cognition. Each stage in the psychological development of the child involved the incorporation and utilisation of operations developed at an earlier stage. Thus adult cognition and language are rooted in the sensory-motor (Piaget, 1954).

David McNeill (1975) in his book *The conceptual basis of language*, took Piaget's idea of sensory-motor based cognition and used it to develop a theory of language understanding, or more precisely, of the relationship between thought and language. He proposed that the sensory-motor ideas that form the basis of a child's early behaviour become elaborated by means of what he calls *semiotic extension* to other less directly sensory and increasingly more abstract domains. Some of his argument is based on the pervasiveness of sensory-motor concepts in language that have been semiotically extended to new domains; such concepts as events, temporal sequences, causation, and so on. For example, the sentence *The men stopped their cars* is an event that can be represented by an action scheme which is an abstraction of the actual motor program that might be involved in executing the event. So a sensory motor idea can be simultaneously both the meaning and part of the action.

On the other hand, the sentence *The men placed their trust in government* is a less concrete event, since governments are not containers, and trust is not something that can be placed in any container. The transition from a concrete model of the world to a more abstract conception is achieved by what McNeill calls *semiotic extension*. Ontologically, one starts with the concrete action sequences, which then attain a level of abstraction which can be referred to as a sensory-motor idea. This in turn can then be extended to non-sensory motor concepts by means of semiotic extension. A key element of McNeill's theory is the notion of action schemes as control hierarchies. Sensory-motor ideas are at the top of such a control hierarchy, at a remove from the motor sequences that enervate muscles.

Interestingly, McNeill states quite explicitly that semiotic extension is not metaphorical extension, claiming that metaphor involves the relating of two semantically unrelated concepts, whereas semiotic extension preserves the semantics of the underlying sensory motor idea. This definition of metaphor is somewhat more restrictive than Lakoff's and, indeed, that of other theorists (Black, 1979). In fact, much of what McNeill refers to as semiotic extension can be accommodated within Lakoff's view of metaphor and metaphorical extension. A metaphor involves the projection of concepts from one domain onto those of another (Black, 1979). For it to be successful there needs to be some degree of isomorphism between the two domains. Such isomorphism can, and often does, entail a semantic overlap. For example, the phrase *That man is a wolf* entails a semantic overlap at the level of the behavioural repertoire of both mammals, specifically their predatory behaviour.

McNeill demonstrated the usefulness of his framework by showing how it could account for speech dysfluencies and the morphology and timing of gestures during speech. His basic argument was that the fundamental building blocks of language understanding were sensory motor actions sequences or *syntagma*; these dictated the nature and shape of conceptual representation in language. With regard to gesture, McNeill sees the gestural accompaniment of speech as a manifestation of the sensory-motor foundation of language. McNeill sees language as action, and gestures that emphasise speech, as opposed to those that have a significance outside speech, are rooted in the same sensory-motor base as the conceptual structures of the utterance itself.

In more recent work, McNeill (1992) has explored the relationship between gesture and language in considerably more depth. He argues that gesture and spoken language are intimately connected to each other. He makes the case for the notion of a gesture/language dialectic where gesture and language interact and complement



**Comment:** Page: 8  
Experiment idea - Is there a connection between use of gestures in speech and abstract sensory associations? Aifric seems to have both well developed.



each other in the communication of a concept. From studies of recorded gesture he has discerned a number of striking phenomena suggesting the significance of gesture in conveying information. In one example, he shows that gesture will often be used to supply information that does not manifest itself in the associated speech stream. In a particularly compelling example, he shows that speech errors are not necessarily paralleled by errors in the simultaneously unfolding gestural sequence. In another case he has found that the gestural sequence can itself serve as a framework for driving language production. As if the gestures serve as an externalised schema upon which the verbal output is constructed. McNeill's view is that:

speech and gesture are elements of a single integrated process of utterance formation in which there is a synthesis of opposite modes of thought -- global-synthetic and instantaneous imagery with linear-segmented temporally extended verbalization. (McNeill, 1992, p. 35)

This is Lakoff's embodiment argument in another guise; if the building blocks of language are indeed sensory-motor, this must give rise to features of language generation and processing that cannot be understood or accounted for if we ignore this aspect of embodiment. McNeill's view of language complements that of Lakoff, with McNeill emphasising the motoric roots of language, and Lakoff emphasising the sensory.

### **Negotiated syntax**

Much of what has been discussed so far deals with what might be considered the vaguer aspects of language processing, such as the relationship between language and conceptual structure, the dynamics of meaning extension and change, and so on. What of that paradigmatically hard-nosed area of linguistics: syntax? How can it be dealt with by a metaphor-based language system?

Sandra Thompson and her co-workers (Tsuyoshi & Thompson, 1993; Fox & Thompson, 1990) have carried out a number of studies of conversational speech primarily from the perspective of syntax, and it is clear from their work that the syntactic structure of conversation bears little relationship to the those structures derived from the study of conventional prose narrative, in other words the kind of syntax that one finds described in standard linguistic textbooks. Thompson's conversational data show that the realisation of syntax in conversation is dynamic and negotiated. The analysis of conversational syntax has been eschewed by many in conventional linguistics because of the perception of it as a "messy", performance-

contaminated version of legislated syntax. It appears from the conversational analysis work of Thompson and others (Schlegoff, 1979; Chafe, 1987), that the major unit of analysis in conversation is the *constructional schema*. This is the key element of Thompson's conversational analytic framework, and refers to a structured inventory of clauses distilled from a large number of speech events. The more frequent such patterns are, the more entrenched in the "grammar" of the language they become. Their most important feature is that they are loose frameworks, differing in their degree of grammaticalisation, and not rigidly adhered to by speakers. Significantly, from the point of view of the argument being put forward in this paper, Thompson has suggested that constructional schemas are interrelated in a complex radial fashion, in much the same way as the various category structures elucidated by Lakoff (1989). Examples of constructional schemas are the SVO declarative clause, prepositional phrases, WH-questions, and so on. This focus on what, in effect, are clauses reflects the general view among conversation analysts that the notion of sentence is an artificial construct, with little existence outside the domain of conventional prose narrative.

Related to the issue of negotiated syntax is the broader one of so-called "ill-formedness" in dialogue. In a study involving the development of a natural language dialogue interface to a database, I along with some colleagues collected a corpus of task-related keyboard dialogue (Reilly, Egan, Harper, & Harris 1986). An analysis of this dialogue indicated some rather strange patterns. In particular there seemed to be something akin to the phenomenon identified by Barbara Grosz (1977) in her pump-assembly dialogues, that of the structure of the task imposing a structure on the dialogue. In our case we found that the task structure seemed to affect the syntactic structure of the sentence. Take the sentence: *How many students male and female in the degree course?* Now, according to the syntactic norms of conventional prose narrative, this is not a grammatical sentence. Apart from the absence of a verb, the adjectives following the noun are not in the conventionally accepted position for adjectives. Yet if we look at this utterance in the context of the task being carried out, there is a striking isomorphism; the task involved transcribing information from the computer screen into tables in which the category "student" was superordinate to the categories "male" and "female". Here again we have an example of what has been conventionally viewed as a rigid rule-based system manifesting a fluidity that cannot be attributed simply to contamination by performance factors.

## Synesthesia

The previous sections have presented some psychological and linguistic evidence for an intimate connection between sensory-motor and language computation. In this and subsequent sections, the focus will be on neuropsychological evidence that points to such a link.

There is an extremely rare condition called *synesthesia*, in which individuals experience cross-modal sensory interference: Tastes are experienced as shapes, sounds as colours, and so on. Cytowic and Wood (1982a, 1982b) described an individual who experienced the tastes of various foods as geometric shapes. It is important to emphasise that this experience is not an abstractly metaphorical one, but as far as can be ascertained, a direct sensation. Furthermore, the sensory patterns are consistent over time; the same tastes will yield the same shapes, even after a substantial period of time has elapsed between testing. Cytowic (1989) has suggested that the roots of synesthesia are in the relationship between the limbic system and the neo-cortex. In particular the hippocampus seems to be implicated. This is the part of the brain that appears to be involved in the mediation of episodic memory, where episodic memory can be defined as memory for objects and events in their spatial, temporal, and sensory context.

Now, while the cases that Cytowic et al. described are dramatic and rare, Cytowic has argued elsewhere (Cytowic, 1992) that synesthesia is a phenomenon that can manifest itself in a wider range of people under special circumstances, such as when induced by hallucinogenic drugs, or in people suffering from migraine. In the latter case, prior to a migraine attack many sufferers experience cross-modal sensory distortions. Cytowic argues that cross-modal interaction occurs in all of us, but that it is available to consciousness in only a few, namely those who experience synesthesia.

In his book *The mind of a mnemonist*, Luria (1968) explored the nature of the prodigious memory of S, one of his patients, who worked as a professional mnemonist. S described the relationship between his sensory experiences and language as follows:

I recognize a word not only by the image it evokes, but by a whole complex of feelings that image arouses. It's hard to express... it's not a matter of vision or hearing but some overall sense I get. Usually I experience a word's taste and weight, and I don't have to make an effort to remember it -- the word seems to recall itself. (Luria, 1968; p.28).

This is clearly another example of synesthesia, but this time involving aspects of language, where we have vivid, arbitrary, yet unlearned<sup>2</sup> connections made between a words and a variety of sensations. A recent study by Baron-Cohen, Harrison, Goldstein, & Wyke (1993) examined the cases of nine female subjects who showed colour-word synesthesia; they experienced colour sensations when listening to speech. Interestingly, there was some degree of consistency across subjects in the colours evoked by hearing the names of certain letters.

While all of these cases are very unusual, there may be evidence of more mundane forms of synesthesia in the developmental area. There may be, for example, a transient form of synesthesia present in young children that disappears as they develop. Some evidence for this comes from the quite common phenomenon of children having vivid and unlearned sensory associations to abstract words and concepts. For example, children will often report that they associate particular colours with numbers or the names of days of the week. At a more abstract level, my 4 year-old daughter recently described my unsatisfactory explanation (in her view) of how the Earth evolved as "a sandy idea", whereas she would describe a more satisfactory explanation as "smooth and glassy." It also seems that even in the stronghold of the abstract, the domain of mathematics, sensation intrudes. Indeed, there have been a number of well known mathematical prodigies who spoke of their skill in sensory terms, particularly visual, and who maintained that this mapping to the sensory was essential to the accomplishment of their complex computational feats (Penrose, 1989). Einstein (Hadamard, 1945) even used the term "muscular" to describe some of the imagery he used in solving mathematical problems.

Further anecdotal evidence comes from the experience of dozing off during a talk or lecture (I'm sure I'm not alone in experiencing this). At some stage in the transition between wakefulness and a partial sleeping state there is a point at which the speaker's words evoke arbitrary visual images. These have a dream-like quality, yet are "driven" by the speech signal. I should emphasise that these images are not images of the concepts being spoken about, but are dynamic semi-abstract images that evolve in synchrony with the speech. Perhaps these are traces of the sensory-motor roots of language processing algorithms, made available by the defocusing of attention when crossing the borderline between sleep and wakefulness.

It is my contention that a harnessing of the sensory domain for higher-order cognitive and language processing occurs in all of us, but is not amenable in normal

---

<sup>2</sup>It may not be strictly true to say that these connections were unlearned, rather that they were not explicitly taught.

circumstances to direct introspection. Indeed I would argue that the fundamental engine of cognition and language is a sensory based (somasthetic, auditory, visual) mechanism that utilises the computational algorithms of the sense domains, at one remove from their role of processing sense data, yet still maintaining their intrinsically sensory nature. The underlying sensory nature of this computation becomes evident in the rare cases of synesthesia or under some of the more mundane circumstances mentioned above.

### **A possible neural basis**

Are there features of neural architecture that might support the kind of sensory-motor/language interaction that has been argued for in this paper? At a very coarse level of analysis it is possible to identify three dominant features of the central nervous system, and specifically of the cortex: Hierarchical processing, modularisation, and reciprocal interaction. The brain is hierarchical in the sense that there is a process of ever more abstract encoding of sensory information as one moves from the sensory surface inwards and upwards through the central nervous system. It is modular in that there appear to be regions of the brain specialised for very specific aspects of the sensory input. In vision, for example, separate regions of the visual cortex are specialised for form, colour, and motion perception (Livingstone & Hubel, 1988). Complementing this modularisation is a strong pattern of inter-module connectivity. Large areas of the cortex are heavily interconnected via approximately topology-preserving feed-forward and feedback cortico-cortical connections. While the complexity of this circuitry has been elucidated, its function is still far from clear. There have been suggestions that it provides top-down feedback to sensory areas of the brain, fine tuning the sensitivity of the lower-level sensory systems (Rolls, 1989). Whatever, their role, these type of connections are pervasive in the cortex, and it is reasonable to assume that they carry out key role in cortical computation.

In general terms, cortical computation can be conceived of as involving the traversal of a trajectory through high-dimensional space, where the dimensions of the space represent the functional units of the cortex; possibly individual neurons, possibly columns of neurons (Sheperd & Koch, 1990). Computation proceeds from some initial state through a series of intervening states to an end state. While the cortex as a whole can be considered one very high-dimensional space, at any given time only certain regions of this space are being traversed<sup>3</sup>. Now if we consider the reciprocally projecting regions to be sub-spaces of this large cortical hyperspace,

---

<sup>3</sup>Traversal of a region is assumed to involve levels of unit activity above some average resting state.

connected to each other by a subset of their respective dimensions, we have a mechanism whereby a particular trajectory traversal in one region can initiate an "analogous" traversal in another region. The trajectories can be thought of as analogous if they have the same shape, in other words if they follow more or less the same route in the overlapping part of the subspace. We cannot think of the axes of cortical hyperspace to be entirely independent or orthogonal. The reason for this is the essentially two-dimensional planar arrangement of the functional units of the cortex, where activity in one of these units may give rise to activity in neighbouring units.

Now let us assume that there are two regions of the cortex that reciprocally project to each other:  $L$  and  $M$  (cf. Figure 1). These regions will have neighbourhoods,  $L^n$  and  $M^n$ , which will be influenced by, and will influence, the activity in  $L$  and  $M$ . Assume also that there are other regions that project to  $L$  and  $M$  in much the same way that  $L$  and  $M$  project to each other; we will call these  $L^P$  and  $M^P$ . Let us further assume that region  $L$  has responsibility for some aspect of language processing, and region  $M$  deals with the organisation of some form of motor activity. We now have a mechanism whereby the repertoire of trajectories in one region (indeed, in one domain) can influence those in another, in a somewhat arbitrary way<sup>4</sup>. Influence is mediated in three ways: first, directly, by the reciprocal connections and the associated pattern of resonances that are set up; second, indirectly, by the influence of the neighbourhood regions  $L^n$  and  $M^n$ ; and third indirectly by the other regions projecting into  $L$ ,  $M$ ,  $L^n$  and  $M^n$ , that is regions  $L^P$  and  $M^P$ . The key feature of this influence is that processing in  $L$  can be driven by activity in  $M$ , but this processing may be experienced, purely in terms of  $L$ , that is, in linguistic terms. Patterns of activity in  $M$  that have a motoric function can thus manifest themselves linguistically, in much the same way as McNeill's sensory-motor ideas. Similarly, activity in other regions, the visual region for example, can influence and constrain activity in  $L$ . In this way, algorithms of a sensory-motor nature can be exploited in language processing.

**<Insert Figure 1 about here>**

In this process of trajectory harmonising, we have a mechanism that can provide a basis for the type of metaphorical computation that has been argued here to be central to cognitive processing. It provides a repertoire of algorithms developed for sensory-motor applications that can be pressed into use for many more tasks. The

---

<sup>4</sup>Obviously, the associations between one set of trajectories and another are not entirely arbitrary, since it is the similarity of trajectories established in both regions that drives this process.

very nature of the reciprocal cortico-cortical connections means that the algorithms are content-addressable; similar trajectories in one area of the brain will evoke similar trajectories in another. Obviously some algorithms will prove more useful than others, and these will be reinforced or preferentially selected as an individual develops, possibly in the way that Edelman (1989) has proposed. Consequently, the details of such algorithms will vary from person-to-person.

### Some simulations

It should be obvious at this stage that connectionist networks provide a sympathetic medium for exploring some of the theoretical issue raised in the preceding sections. What follows is not intended to be an exhaustive exploration of the trajectory-harmonising hypothesis, but a simple demonstration using connectionist networks that the process can be usefully exploited. The example is based on a study carried out by Servan-Schreiber, Cleeremans, and McClelland (1991) on the memory capacity of simple recurrent networks (SRNs). Servan-Schreiber *et al.* set out to explore the conditions under which an SRN can carry information over a long sequence of inputs. The set of simulations described here involves two learning experiments. In the first, an SRN is trained on a corpus of letter sequences generated by a finite state automaton (FSA). The task of the network is to anticipate the next letter in the sequence given the current letter and some form of memory of the preceding letters in the sequence. This is a straightforward replication of Servan-Schreiber *et al.*'s study.

The FSA used is given in Figure 2. It consists of a sequence of eight states, where a state transition entails the generation of a letter from the set {B, E, P, S, T, V, X}. Because of the way in which the FSA is organised, only a limited number of sequences are allowable or "grammatical".

**<Insert Figure 2 about here>**

Figure 3 is a diagram of the SRNs used in the first experiment. It contains seven input units, seven output units, and three hidden units. The context units were used to hold a copy of the hidden unit vector from the previous time step, thus providing a memory of the preceding elements in the sequence. In all of the experiments described here, the context units were reset at the end of an input sequence to prevent information from one sequence affecting another. A set of 30,000 sequences were generated from the grammar. The average sequence length was 8.2, with a minimum length of 5 and a maximum of 40. A typical set of

sequences from this corpus would be: BTSSSXSE, BPTTTVVE, and so on. Using different initial weight settings, several networks of the type shown in Figure 3 were trained for one pass through the 30,000 sequences. The first one to achieve 100% accuracy on a test set of 10,000 sequences sampled from the training set was selected for use in the second experiment. In all cases, the learning rate was set to 0.02 and the momentum to 0.7.

**<Insert Figure 3 about here>**

Note that networks of this sort are not able to predict precisely the next element in the sequence, what they can do is predict the set of legal continuations. So, for example, following the element B, the two next legal continuations are T or P. If the network is behaving correctly, the units representing these two elements should be active, that is they should have activation values above some criterion level. Those units representing illegal continuations should be inactive, having values within a certain criterion of zero.

As the network processes an input sequence and generates an appropriate sequence of outputs, the hidden units of the network follow a specific trajectory through state space (in this case a three-dimensional space). It turns out that this trajectory can, in turn, be used by other networks to solve more effectively similar problems. This was the focus of the second experiment. A new SRN was built (cf. Figure 4, network (a)) and given the same training task, but this time it was provided with additional input in the form of the hidden unit vector from the network trained in the first experiment (Figure 4, network (b)). Note that network (b) receives no further training during this experiment.

**<Insert Figure 4 about here>**

Network (a) was trained on the same corpus of 30,000 sequences under two conditions: (1) Where additional input was provided by the hidden unit vector from the previously trained net<sup>5</sup> shown in Figure 3; and (2) where the additional information was a randomly re-ordered version of this vector. The purpose of this study was to compare performance on the task while controlling for everything except the provision of potentially useful information. The question was, of course, would such information indeed be useful. To ascertain this, 20 training replications involving a single pass through the 30,000 sequence corpus were carried out under

---

<sup>5</sup>The trained net had performed with 100% accuracy on a subset of 10,000 of the training corpus. An output vector was deemed accurate if each of its elements was within 0.3 of the target output. This was the same criterion used by Servan-Schreiber, Cleeremans, & McClelland (1991).



the two experimental conditions. The same set of 20 initial weight settings were used in both conditions. There was a statistically significant difference between the normal and re-ordered conditions, with a mean of 98% and 72% correct for the normal and random replications respectively ( $t=11.25$ ,  $df=19$ ,  $p<0.001$ ).

Obviously, one cannot draw too many conclusions from this result. The primary goal of the experiments was to demonstrate how a trajectory generated in one network can be used to facilitate the execution of a related task in another; and this was shown quite clearly. Note that although the input letter sequences were identical for the two learning experiments, the overall tasks were not identical, since in the second experiment network *A* had deal with the letter input, context input, and hidden unit input from network *B*. A drawback of this coupled network, is the nature of the coupling; the two sub-networks are in lock step in discrete time. A more realistic test of this framework would involve the use of asynchronous networks operating in continuous time with reciprocal connections, thus approximating more closely the processing style of real neural networks.

## Empirical implications

What empirical implications can one draw from ideas presented here? Well, starting with the developmental domain it should be possible to explore what I have referred to as transient developmental synesthesia among young children. Indeed there is an implicit acknowledgement of its existence in some of the techniques used by teachers of reading: sounding out a word, using coloured letters, using letters that have distinct tactile properties. One prediction from what I have discussed is that this approach could be made more effective by exploiting the *arbitrary* sensory associations that get established within individual children when they are acquiring concepts such as numbers, letters, names for days, and so on. One should find that using these associations, providing that they are accessible, will enhance the learning process. Furthermore, in adults who have clear memories of their developmental synesthetic associations, it should be possible to devise experiments to demonstrate enhanced recognition, say, of numbers in the presence of the relevant synesthetic stimulus.

Another implication of the hypothesis outlined above, is that the language mechanisms of people who have a sensory-motor deficit from birth, in addition to the obvious one of deafness, should be subtly different from people who do not possess such deficits. I am not suggesting that the language of these people should be inadequate in any way, rather that since there will have been a reliance on a different

**Comment:** Page: 18  
Do a literature search on language deficits and the blind.

repertoire of sensory-motor algorithms, these will have given rise to different conceptual and linguistic processing *styles*. It's not clear to me at present how these styles might differ, or if the differences will be large enough to be detected against the background noise of individual differences. There is, however, substantial evidence in the literature that blind and visually impaired children acquire language more slowly than sighted children. Of course, this might be due to a number of factors; the diminution of experience brought about by the impairment, or a difference in the language input to which the child is exposed. Nonetheless, a recent study by Andersen, Dunlea, and Kekellis (1993) has shown that the variation in language development cannot be uniquely ascribed to either of these factors, a finding which tends to support the hypothesis being proposed here.

### **Implications for AI and cognitive science**

We can draw a number of rather general implications from the ideas presented here for AI and cognitive science:

1. Visual, and more broadly, sensory-motor processing should be explored as possible sources of low-level algorithms for language processing;
2. The most suitable computational framework for doing this is connectionism, since it facilitates the interaction of algorithms from different domains more readily than other approaches;
3. The most suitable paradigm for doing this seems to be something along the lines of the  $L_0$  project (Feldman, Lakoff, Bailey, Narayanan, Regier, Stolcke, this volume), though with extensions to deal with motor activity;
4. Architectures most likely to yield interesting results are those that embody some form of temporal dynamics, since it has been proposed here that a key element of the computational style underlying natural language processing is trajectory harmonising between language and sensory-motor areas of the cortex.

### **Conclusions**

It has been argued here that there is an intimate link, both evolutionarily and developmentally, between the computation that underlies language processing and sensory-motor computation. From an evolutionary point of view it seems that the computational infrastructure upon which language depends is an interface between sensory and motor cortices. This suggests that the computational repertoire of these

systems must owe something to the more ancient perceptual and control systems, and perhaps this debt manifests itself in some aspects of the nature and structure of language. One such manifestation is the pervasiveness of metaphor, and moreover of metaphors that are anchored in the structures and functions of the human body. Furthermore, these basic physical metaphors are extended to more abstract concepts in a principled and systematic way. Whereas metaphor and metonymy are primarily semantic and pragmatic in nature, there is also the suggestion from research carried out on conversational dialogue, that syntax has similar organisational characteristics as other conceptual structures. The notion of embodiment was further explored in the context of McNeill's sensory-motor theory of conceptual structure and was illustrated with specific reference to gesture. Evidence was also adduced from neuropsychology and neuroscience of strong computational links between language and the sensory-motor systems. It was contended, therefore, that the structure of language, its fluidity and flexibility arise from the exploitation of extant sensory-motor systems, albeit at an abstract level. The implications for AI and cognitive science are that natural language processing might usefully be constructed on a sensory-motor foundation involving systems that process and coordinate sensory information from several sources and that act on, and react to, their environment.

## References

- Andersen, E. S., Dunlea, A., & Kekelis, L. (1993). The impact of input: Language acquisition in the visually impaired. *First Language, 13*, 23-49.
- Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley, California: University of California Press.
- Black, M. (1979). More about metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 19-43). Cambridge, UK: Cambridge University Press
- Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21-51). Amsterdam: Benjamins.
- Chomsky, N. (1967). *Aspects of a theory of syntax*.
- Cytowic, R. E. (1992). *The man who tasted shapes*. London, UK: Abacus.
- Cytowic, R. E., & Wood, F. B. (1982a). Synesthesia I: A review of the theories and their brain basis. *Brain and Cognition, 1*, 23-25.

- Cytowic, R. E., & Wood, F. B. (1982b). Synesthesia II: Psychophysical relationships in the synesthesia of geometrically shaped taste and coloured hearing. *Brain and Cognition*, 1, 36-49.
- Cytowic, R. E. (1989). *Synesthesia: A union of the senses*. New York: Springer-Verlag.
- Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.
- Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T., Stolcke, A. (this volume). L<sub>0</sub> - The first four years. *Artificial Intelligence Review*, 8.
- Fox, B. A., & Thompson, S. (1990). A discourse explanation of the grammar of relative clauses in English conversation. *Language*, 66, 297-316.
- Geschwind, N. (1965). Disconnexion syndromes in animals and man. *Brain*, 88, 237-294, 585-664.
- Gregory, R. L. (1974). *Concepts and mechanisms of perception*. London, UK: Gerald Duckworth & Co. Ltd.
- Grosz, B. (1977). *The representation and use of focus in dialogue understanding*. Unpublished doctoral dissertation, University of California, Berkeley.
- Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Princeton, New Jersey: Princeton University Press.
- Inoue, K. (1993). *Linguistic agreement, conceptual coherence, and linguistic residue in the Japanese numeral classifier system*. Unpublished doctoral dissertation, Department of Anthropology, University of Illinois at Urbana-Champaign.
- Lakoff, G. (1989). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, Illinois: The University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, Illinois: The University of Chicago Press.
- Lieberman, P. (1992). On the evolutionary biology of speech and syntax. In J. Wind, B. Chiarelli, B. Bichakjian, & A. Nocentini (Eds.), *Language origin: A multidisciplinary approach*. Germany: Kluwer Academic Publishers.

- Livingstone, M. & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Nature*, 240, 740-749.
- Luria, A. R. (1968). *The mind of a mnemonist*. New York: Basic Books.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, Illinois: The University of Chicago Press.
- McNeill, D. (1979). *The conceptual basis of language*. Chicago, Illinois: The University of Chicago Press.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Penrose, R. (1989). *The emperor's new mind*. Oxford, UK: Oxford University Press.
- Reilly, R., Egan, O., Harper, J., & Harris, J. (1986). *ESPRIT Project 527. Dialogue studies: Main phase (Deliverable 4)*. Educational Research Centre, St Patrick's College, Dublin.
- Rolls, E. (1989). The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In R. Durbin, C. Miall, & G. Mitchison (Eds.), *The computing neuron* (pp. 125-159). Wokingham, UK: Addison-Wesley.
- Rosch, E. (1973). Natural categories. *Cognitive Psychology*, 4, 328-350.
- Rumelhart, D. E. (1979). Some problems with the notion of literal meanings. In A. Ortony (Ed.), *Metaphor and thought* (pp. 78-90). Cambridge, UK: Cambridge University Press
- Schlegoff, E. (1989). The relevance of repair to syntax-for-conversation. In T. Givón (Ed.) *Syntax and semantics volume 12: Discourse and syntax* (pp. 261-286). New York: Academic Press.
- Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal constraints in simple recurrent networks. *Machine Learning*, 7, 161-193.
- Sheperd, G. M., & Koch, C. (1992). Introduction to synaptic circuits. In G. M. Sheperd (Ed.), *The synaptic organization of the brain* (pp. 3-31). Oxford, UK: Oxford University Press.

Tsuyoshi, O., & Thompson, S. (1993). What can conversation tell us about syntax?  
Invited paper presented to Fifteenth Annual Conference of the Cognitive  
Science Society, University of Colorado at Boulder, Colorado.

Walsh, P. (1990). Imagery as a heuristic in the comprehension of metaphorical  
analogies. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, & G. Erdos (Eds.),  
*Lines of thinking: Reflections on the psychology of thought. Volume 1 -  
Representation, reasoning, analogy, and decision making.* Chichester, UK:  
John Wiley & Sons.

Wilkins, W., & Wakefield, J. (in press). Brain evolution and neurolinguistic  
preconditions. *Brain and Behavioral Sciences.*

## Figure Captions

### Figure 1

A schematic representation of the interaction of language and motor areas via  
reciprocal cortico-cortical projections. Trajectories in the  $L$  and  $M$  regions become  
harmonised, constrained by activity in adjacent cortical regions ( $L^n$  and  $M^n$ ) and by  
areas projecting into them from other regions ( $L^P$  and  $M^P$ ).

### Figure 2

The finite state automaton used to generate letter sequences for the simulation  
experiments described in the text.

### Figure 3

The simple recurrent network (SRN) used in the first simulation experiment.  
Note that the hidden activation values at time  $t$  are copied back to context units at  
time  $t+1$ .

### Figure 4

The coupled SRN used in the second simulation experiment. Learning only  
takes place in network (a). The second experiment involved two conditions, one

where the hidden unit vector from network (b) was fed unchanged into network (a), and the other in which this vector was randomly re-ordered.









