Representing Aspects of Language

Robert F. Port
Departments of Linguistics and Computer Science
port@ucs.indiana.edu

Timothy van Gelder
Department of Philosophy
Indiana University, Bloomington, Indiana 47405
tgelder@ucs.indiana.edu

Abstract

We provide a conceptual framework for understanding similarities and differences among various schemes of compositional representation, emphasizing problems that arise in modelling aspects of human language. We propose six abstract dimensions that suggest a space of possible compositional schemes. Temporality turns out to play a key role in defining several of these dimensions. From studying how schemes fall into this space, it is apparent that there is no single crucial difference between AI and connectionist approaches to representation. Large regions of the space of compositional schemes remain unexplored, such as the entire class of active, dynamic models that do composition in time. These models offer the possibility of parsing real-time input into useful segments, and thus potentially into linguistic units like words and phrases.

Introduction

What is the relationship between the kinds of symbolic representations deployed in "classical" cognitive models and representations in their connectionist counterparts? The study of human linguistic capability is a particularly appropriate domain in which to discuss this issue. But we must not let theoretical linguistics define the problem. Linguistics presumes the separability of linguistic knowledge and language use, the distinction of competence from performance. There is much more to the linguistic cognition than syntax. Still, language implies a system of complex representational structures for various purposes. And one of these is *orthography*, the level of an author or editor producing text in a "received" form. The discrete, static, concatenated nature of printed language does imply a model for the underlying cognitive architecture. The intuitive appeal of this view of language has inhibited consideration of the many problems in accounting for speech perception and production, the lexicon, slips of the tongue, etc, that imply consideration of other kinds of models. Both AI theorists and linguists have regarded natural language as basically consisting of concatenated structures of discrete, arbitrary symbols, and have constructed models of language processing that postulate internal representations of the same general kind. Connectionists working on natural language have been developing alternative distributed representational schemes, but tend to agree with traditionalists regarding certain crucial ingredients of the traditional picture — such as the need for compound representations systematically constructed out of standard parts. So what, if anything, really differentiates traditional symbolic and connectionist forms of compositional representation?

There is a widespread tendency to suppose that there must be some one crucial issue which neatly separates traditional symbolic representations from connectionist representations. We believe, however, that no such magic key exists. Rather, there is a variety of ways in which connectionist schemes of representation differ, to a greater or lesser extent, from traditional schemes of representation. This is true even when we restrict attention to specifically compositional schemes.

We consider a representational scheme to be compositional if it systematically constructs complex representations from basic compoundable units, such that the semantic and causal significance of the compound whole is a function of the significances of its basic parts. We describe individual representations belonging to a compositional scheme as *symboloids* in order to stress that the category of representational schemes we are interested in includes, but is considerably more general than, traditional symbolic representations. There must also be included various kinds of fuzzy and dynamic distributed patterns that serve a representational function for nervous systems.

To understand this situation we find it useful to view traditional and connectionist schemes as varying along a number of abstract dimensions defining a

[0] To appear in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, L. Erlbaum, Hillsdale, NJ., 1991.

space of possible schemes of compositional representation. Classical systems occupy one relatively restricted region (centered on prototypical symbolic schemes such as the structures in LISP or formal logic) while various connectionist schemes tend to be widely scattered across other regions. Even different instantiations of natural language, such as printed English and spoken utterances, are found at quite distinct points. The kinds of neural representations that underlie natural language capacities most likely occupy several quite different regions in the space depending on which linguistic skills one looks at (e.g., interpreting speech directed at you, editing a text, learning a new word, mimicking someone's dialect, etc.).

Our research is aimed at isolating and clarifying the most important dimensions of this space — to distill out the conceptual "principal components" (or so we hope). We seek a picture of compositional representation in cognitive systems that is more general and inclusive than has previously been available. An important result of this search is that, from the perspective afforded by an understanding of these various dimensions, one can see that standard AI-style schemes of representation have no monopoly on compositionality. Instead, they merely occupy one relatively narrow corner of the space of possible systems that allow composition of tokens. Further, we think that exploring other regions of this space will be more fruitful for cognitive science research than developing variants within this narrow region. The following is a tentative list of the most general dimensions of the space of symboloidal representational schemes. The issue should not be which scheme is the correct one, but rather which scheme is most appropriate for which phenomena?

Typology of Representations

The major dimensions fall into at least three groups: (1) intrinsic properties of the compoundable tokens, (2) the manner in which composition of tokens takes place and (3) the functional role of symboloids in the operation of the system. We do not claim these are a minimal set of dimensions. (For example, there are probably differing possible assumptions about semantics, that is, about the nature of the world itself.) We have tried to choose dimensions that are independent and which clarify the range of possible solutions to the representation problem. In addition to describing particular computational models, we will make reference both the linguists' models of language and to various aspects of knowledge about language exhibited by skilled users.

Properties of Basic Tokens

The tokens are the most basic entities with representational significance which can be compounded into derived structures. In natural language processing, individual words are assumed (by both linguists and high-school English teachers) to serve as the basic units. As a first-order model, this is satisfactory. Otherwise, orthographies and typographical conventions would be much more problematic than they are. But Bolinger among others has frequently pointed out how difficult it is to nail down the actual list of the particular items that have just the right properties (Bolinger, 1975). Language is not just a set of morphemes plus rules of syntax. Speakers use and 'know' linguistic fragments that come in many sizes: from submorphemic ideophones to words, idioms, and Bolinger's 'collocations', as well as 'cliches' and even entire sentences and paragraphs of boilerplate (as in genres like wills and academic recommendation letters). The nature of these mysteriously constrained yet mysteriously flexible lexicalized 'units' lies well beyond the grasp of current representational schemes employed in linguistic theory. Although an orthography can afford to ignore these data, a model of human linguistic representation must support all of these kinds of linguistic knowledge.

Discrete vs. Continuous Token Set. Basic tokens are physical entities that gain their type-identity (e.g., being an instance of the word *cat*) by exhibiting characteristic variation along certain critical dimensions (e.g., placement of ink on the page). These dimensions can be thought of as delineating a space of possible physical items which might count as tokens (e .g., ink marks on a page or a set of letter strings). This first issue concerns how the basic tokens fall in this space. Are they sparsely or densely packed into it? That is, are there entities which count as basic tokens in the scheme 'between' any two tokens of the scheme? For example: the so-called 'phonemes' or 'phonological segments' used to describe particular languages are assumed to be discrete (that is, located far apart in stimulus space — at least one feature step apart), while the acoustic and articulatory phenomena themselves are clearly densely packed or continuous.

Static vs. Dynamic Tokens. Standard symbolic models invariably assume that basic tokens are static. By this we mean that the tokens exist indefinitely and change only when explicitly altered — like the basic vocabulary of predicate calculus. They provide a conceptual model for standard orthographies of, say, English and French. The inventory of units is relatively static and the objects (the letters and words) endure fairly indefinitely once converted to the printed medium.

These should be contrasted with the representation that presumably must underlie spoken words. This must change continuously over time as the pattern is either produced or recognized. Acoustic changes are essential to a word's being the word that it is, and to its being correctly perceived (Liberman, et al, 1967). Sometimes temporal specifications can be very subtle indeed (Port and Crawford, 1989). It is likely that internal representations of spoken words are also dynamic in this sense, as suggested by some (Browman and Goldstein, 1986; Saltzman and Munhall, 1989).

With dynamic tokens, the type-identity of tokens is based on physical change over time. Another example is the differentiation of smells in the olfactory bulb of the rabbit (Skarda and Freeman, 1987). Familiar smells each exhibit a characteristic limit cycle through the activation state-space of a set of cells. There appears to be no single point in time when the instantaneous state 'represents' the identity of the scent.

Currently one can find connectionist models that employ both kinds of representations. Most networks produce static output vectors, that is, the dynamics of the network leads it toward a point attractor (eg, (Sejnowski and Rosenberg, 1987; Elman and Zipser, 1988)). These two properties of basic tokens, discrete/continuous and static/dynamic seem to be the most important ones,[1] but models also differ in the way composition of basic units takes place.

Manner of Composition

How are representational tokens of whatever form actually combined to make compounds or higher-level representations? This ability is clearly essential to any representational scheme that is to be generally useful (Fodor and Pylyshyn, 1988).

Concatenation vs. Superimposition. Connectionist work has revealed the importance of distinguishing several ways of combining tokens into complex wholes (van Gelder, 1990). In traditional logic and AI, there is, of course, no change whatever in tokens themselves when composed, as when LISP atoms are strung together to form a complex statement. At the other extreme is superimposition, or simultaneous combination, as found, for example, in a Recursive Auto-Associative Memory or RAAM (Pollack, 1991; Blank et al., 1992). Such systems can be trained to push and pop the elements of hierarchical trees from a single distributed representation in a fixed-size group of nodes. Input tokens are systematically combined into completely distributed compounds by a sequence of learning processes. The result is a static representation from which the entire tree can be constructed even though no specific physical characteristics of the constituents are present. It is even possible to demonstrate structure-sensitive transformations on distributed representations using a RAAM (Chalmers, 1990). These demonstrations suggest that symbolic composition is not the only kind possible.

Between simple physical concatenation and superimposition lies Context Sensitive Concatenation (Smolensky, 1988). Tokens with distinct physical identities are concatenated but they are affected by their combination with particular other tokens. These are illustrated by Elman's recent models (Elman, 1989) and the dynamic memory of Anderson and Port.

Static vs. Temporal Combination. A crucial issue that cuts across the previous one is whether the actual act of combining basic units to form compound representations occurs in time or only statically. (The temporality of combination should not to be confused with the temporal vs. static nature of basic tokens themselves that we discussed in the previous section.) Both standard computer languages and linguistic models (Chomsky, 1965) depend on static hierarchical structures.

Of course, if modelling of time as a sequence of symbols proves inadequate, one can always model time as 'just another parameter' and deposit some kind of time measurements within the representation itself (Klatt, 1976; Port, 1981). But eventually, before addressing the real world, the artificiality of this maneuver must be faced by any model of representation in a nervous system (Port, 1990).

Certain 'recurrent' connectionist models (eg, (Anderson and Port, 1990; Elman, 1989)) deploy complex representations that are essentially temporally extended. The basic components, attractor states, are appended in time, so the structure of a sequence is encoded in characteristic trajectories through state space. Systems along this line can be extended to exhibit limit-cycle dynamics. It is likely that a dynamic analog of concatenation in time can be achieved by passing across the saddle points that separate one attractor basin from the next for each sequential component. We suspect that complex representations formed by temporal combination must underlie a wide range of real-time cognitive processes in speech and language.

Thus, there appear to be several possible kinds of tokens. And each may support composition in several ways.

Functional Role of Representations

Symboloidal schemes of representation exhibit certain kinds of properties that only become apparent when we consider how representations in that scheme are handled within the context of processing. It is to be expected that the radical differences in the nature of representational elements we are proposing will have consequences for their use.

[1] There is at least one other property of tokens as well that may be relevant. How are basic tokens paired with what they represent? Are basic tokens with similar meaning also similar physically? Natural and formal languages use arbitrary symbols. That is, tokens the typically have no intrinsic relation to what they represent. This lack of meaning is a key ingredient in the notorious "grounding problem". Connectionist models often use arbitrary patterns as the basic compoundable units but not always. On the other hand, the stimulation patterns on a sensory surface of an animal (like the finger tips, retina and basilar membrane) illustrate highly non-arbitrary pairing of a stimulus pattern with meaning.

Passive vs. Active. Representations are normally seen as inert structures that are distinct from the processes that manipulate them. This powerful assumption underlies the notion of a Universal Turing Machine. The embedded Turing machine is described as a set of rules, which the universal Turing machine can execute at its convenience. The representations of the embedded model sit on the tape waiting to be acted upon by the host machine.

Another possibility is that representations are self-executing processes. That is, a symboloid might propel itself toward new representational stabilities without need of an external executive that reads the states and executes the steps specified in a static rule table. Self-execution can be driven either by external input or by internal dynamic representations from elsewhere in the system. As a primitive example, the Interactive Activation models (McClelland & Rumelhart,1981; Elman and McClelland, 1986) for speech recognition have representations that are (weakly) active in this sense. As input feature vectors become active during word presentation, they excite phoneme-level units and eventually word-level units. Information flows up, down and laterally within the model such that evolving representations at one level drive the development of representations at other levels. No external observer (or universal Turing machine) reads the activation levels and to decide what to do next. The control of the model has been decentralized or distributed. The Anderson-Port model for auditory pattern recognition that will be described below is also active in this sense.

Digital vs. Analog. Do the representations enable operations upon them that are 'positive' and 'reliable' in the terms of (Haugeland, 1985)? In classical symbol systems, representations are digital in the sense that the most basic identifying and transforming processes can always be carried out with complete, unambiguous success (e.g., the executive system can tell that the symbol in the buffer is either *foo* or it is not; there is no question of its being "somewhat *foo*"). This digital character is, of course, supported by other properties of the representational scheme such as discrete basic tokens and the strict concatenation property.

In some connectionist schemes, however, we find representations that are not digital in this sense. For example, if one attempts to use the RAAM architecture to represent many sequences by sequential superimposition, the representations of stack states become so closely packed in the activation space of the relevant units that operations on the representations eventually cannot be carried out positively and reliably (Pollack, 1991). Such a system is essentially analog. As trees are reconstructed from the stack, reliability deteriorates. Things that are different are eventually forced into equivalence classes.

Dynamic Models and Time

We think that these dimensions are useful in revealing both the possible diversity in symboloidal schemes of representation and some of the similarities that exist between the classical symbolic model and connectionist schemes. But the connectionist framework also lends itself to dynamic models (Hopfield, 1984; Grossberg, 1980) and these open up opportunities for a different way of conceptualizing representations (see (van Gelder, 1991) for further discussion). Thus, we now explore that possibility in a little more detail in the context of a model for recognition of auditory patterns. These patterns might be as short and complex as a familiar chair squeak or bird chirp, of intermediate length, like a syllable or a strum on a guitar, or they might be hierarchically structured auditory objects like words, melodies or sentences.

A crucial step in dealing with real-time input to a cognitive model is the step from *milliseconds* to abstract *event categories*. This is very difficult to achieve because abstract events typically occur with variable durations. Thus temporal measurements in milliseconds (even if nervous systems did have a way of measuring them) are simply the wrong kind of description (Rosen, 1978). It appears that dynamic systems provide a way to obtain abstract categories from real-time input.

Dynamic Memory. Processing of temporal patterns as dynamic representations has been demonstrated to a limited extent. For example, Simple Recurrent Networks (Elman, 1990; Elman, 1989) respond to learned sequential patterns by following a characteristic trajectory in activation space. We view these trajectories as dynamic representations, distinct for each input sequence. In this case, states of the system near the end of each sequence distinguish each learned pattern. This can be illustrated by looking a little closer at the model described by Anderson and Port.

A network was trained (by a form of supervised gradient-descent learning) to recognize sequential patterns of tones. The network has 6 input frequency bins of which only one was strongly active on each time frame (since the input was effectively sinusoidal over this frequency range). There were 7 nodes in the fully recurrent Dynamic Memory, as shown in Figure 1. The activation of each node on the next time step is the decayed value from the previous time step plus the squashed sum of weighted activations from all other nodes in the Dynamic Memory clique. In addition, each of these nodes receives inputs from a layer that indicates the energy in the six frequency bands. The system learned to recognize a particular sequence of 8 input spectra comprising a melody-like pattern (Anderson and Port, 1990).

The model was trained to recognize one or more melody-like target patterns by using one of the dynamic memory nodes as an identification node for each

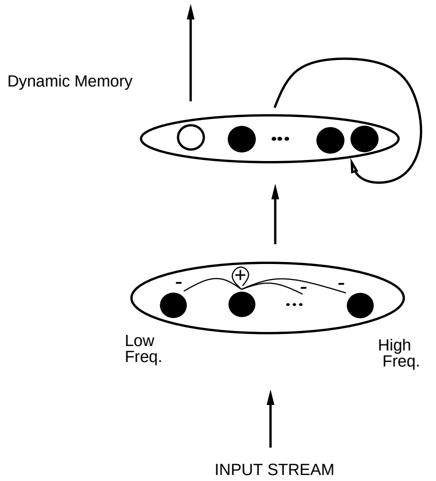

Figure 1: The network that recognizes melody-like 8-step sequential patterns. There are six input nodes, one for each frequency bin, and seven dynamic memory nodes that are fully connected (to itself and all neighbors. One (or more) of these nodes also serves as a category identification node and is trained to reach a target value for just the last 2 time frames of the pattern. The input-feature group learned to sharpen the spectrum by lateral inhibition of its neighbors.

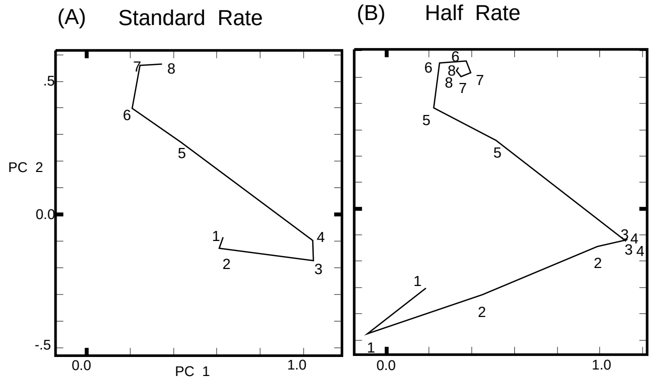

Figure 2: The trajectory of a target pattern plotted on the first two principal components of the 7-D activation space of dynamic memory. The patterns were 8-frame tone sequences (see Anderson and Port, 1990). The left panel shows the response after the presentation at the standard rate (the rate at which training was performed). The right panel shows the same network's response to the target when each spectral frame was presented twice.

target. The weight vector after training enabled it to produce stable trajectories for target sequences ending in a particular corner of the space (where an identification node approached 1) for the last two time cycles of presentation of any target. The model achieved better than 90% correct identification on each of several targets. This contrasts with the response of the system to distractor patterns (that is, to non-trained patterns). In this case, trajectories are not distinctive. As shown in the left hand panel of Figure 2, when a target is presented, the memory followed a characteristic path through activation space (represented here by its two most significant principal components).

What if the rate of presentation is altered? When the patterns are presented at half-tempo, as shown in the right hand panel of Figure 2, essentially the same trajectory is produced. This suggests the possibility that chains of abstract objects could be learned by such a dynamic network. If the system waits to allow the input patterns themselves to determine when the representational state should change, a major problem in control of cognitive models might be solved — the problem of how to parse time into useful pieces.

The critical property of rate invariance is achieved because the representation consists of a sequence of dynamically linked states. These states are not forced to proceed from one to the next at a rate clocked by the machine itself, but are instead controlled by changes in the inputs (in interaction with the internal state). These results show how static fixed points can be combined into a learned basic sequence when driven by external input (which might be sensory or produced by another group of nodes). It is likely such tokens can be combined into larger structures by a process that nests them into higher-level, more slowly changing, representations. The incoming pattern itself entrains the dynamics of the trajectory determining the entire composite representation. The representations of the constituents of these melody-like patterns actively lead toward those states characteristic of the learned pattern as a whole. This is the sense in which the tokens are dynamic and play an active role in the system. They determine, at the appropriate point in time, what will happen next.

Interestingly, when attention is addressed to the dynamics of recurrent networks as a way of accounting for interesting behavior, then one notices that connectionism itself is important in this enterprise particularly for its support of learning and distributed representations. Other, non-connectionist models might also support interesting dynamics.

Conclusions.

One goal of this paper was to suggest that there are many more possible schemes that can serve a representational function than have been well-explored thusfar. Without a clear idea of the degree of independence of the parameters discussed here, it is difficult to estimate the size of this space. We hope that the framework offered here will help fix thoughts on this problem.

A second goal was to emphasize the possibility of dynamic, active representations that are meaningful because of their grounding in sensory surfaces in real-time (Gibson, 1968). Although symbolic models for language assume timeless static symbols *apriori*, systems embedded in the world must obtain appropriate symbolic functions on their own. Dynamic representations in the auditory pattern-recognition model exhibit natural rate invariance by entraining their dynamics to the temporal structure of stimulation. This is an important step toward an understanding of how the ordered, nested, symbol-like representations of language could be learned by a system that functions in real time.

Returning to language, we have suggested that dynamic models have many useful properties – especially for aspects of linguistic skill that are ignored by linguistics. Clearly, further technical developments will be necessary to demonstrate composition beyond the most basic level for dynamic representations. Many aspects of human speech have proven intractable to symbolic models. Perhaps symboloids can do better.

Acknowledgements. Thanks to Sven Anderson for important contributions to the development of these ideas and to David Chalmers as well as Lisa Meeden, Doug Blank and other members of the Mind, AI and Computation group at Indiana for helpful comments on various drafts. This research was funded in part by the Office of Naval Research, grant number N00014-91-J1261.

References

Anderson, S. and Port, R. (1990). A network model of auditory pattern recognition. Technical Report 1, Indiana University, Institute for the Study of Human Capabilities.

Blank, D. S., Meeden, L. A., and Marshall, J. B. (1992). Exploring the symbolic/subsymbolic continuum: a case study of RAAM. In Dinsmore, J., editor, *Closing the Gap: Symbolism vs. Connectionism.* L. Erlbaum Assoc.

Browman, C. and Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook*, 3:219–252.

Chalmers, D. J. (1990). Syntactic transformations on distributed representations. *Connection Science*, 2:53–62.

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* MIT Press, Cambridge, Mass.

Elman, J. (1989). Representation and structure in connectionist models. Technical Report 8903, Center for Research in Language, UCSD, La Jolla, CA 92093-0108.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Elman, J. and Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America*, 83:1615–26.

Fodor, J. A. and Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28:3–71.

Gibson, J. J. (1968). *The Senses Considered as Perceptual Systems.* Harcourt Brace, New York.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review*, 87(1):1–51.

Haugeland, J. (1985). *Artificial Intelligence: The Very Idea.* MIT Press.

Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences*, 81:3088–3092.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59:1208–21.

Pollack, J. (1991). Recursive distributed representations. *Artificial intelligence*, page in press.

Port, R. and Crawford, P. (1989). Pragmatic effects on neutralization rules. *Journal of Phonetics*, 16(4):257–282.

Port, R. F. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, 69:262–274.

Port, R. F. (1990). Representation and recognition of temporal patterns. *Connection Science*, 1-2:151–176.

Rosen, R. (1978). *Fundamentals of measurement and representation of natural systems.* North-Holland.

Saltzman, E. and Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1:333–382.

Sejnowski, T. and Rosenberg, C. (1987). Parallel networks that learn to pronounce English text. *Complex Systems*, 1:145–168.

Skarda, C. and Freeman, W. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10:161–195.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11:1–74.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14:355–384.

van Gelder, T. (1991). Connectionism and dynamical explanation. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Hillsdale, NJ. L. Erlbaum Assoc.