DOCUMENT RESUME

ED 395 516. FL 023 865

AUTHOR	Whincop, Chris
TITLE	Constructivism, Optimality Theory and Language Acquisition. The Shapes We Make in Each Other's Heads.
PUB DATE	96
NOTE	16p.; For complete volume, see FL 023 856.
PUB TYPE	Journal Articles (080) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Information Analyses (070)
JOURNAL CIT	Edinburgh Working Papers in Applied Linguistics; n7 p112-125 1996
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Cognitive Development; *Constructivism (Learning); Foreign Countries; Grammar; Language Processing; Language Universals; Linguistic Input; *Linguistic Theory; Metaphors; *Neurolinguistics; *Neurological Organization; Perceptual Development; *Second Language Learning
IDENTIFIERS	*Optimality Theory

ABSTRACT

This paper identifies a feature of human brain neural nets that may be described as the principle of ease of processing (PEP), and that, it is argued, is the primary force guiding a learner towards a target grammar. It is suggested that the same principle lies at the heart of Optimality Theory, which characterizes the course of language acquisition as a progressive reranking of a hierarchy of universal and violable constraints. It is observed that the hierarchy a learner is in possession of at any particular time is the learner's present characterization of the grammar of the target language and will determine what outputs nets involved in linguistic processing produce for any given inputs to those nets. It is suggested that spatial metaphors may give a clearer insight into the workings of neural nets, and that the process of self-organization of nets is seen in accordance with the PEP as a realignment of the positions of linguistic elements in a multidimensional space that is a characterization of the target language. (Contains 46 references.) (Author/NAV)

Constructivism, Optimality Theory and Language Acquisition the Shapes We Make in Each Other's Heads

Chris Whincop (DAL)

CONSTRUCTIVISM, OPTIMALITY THEORY AND LANGUAGE ACQUISITION - THE SHAPES WE MAKE IN EACH OTHER'S HEADS.

Chris Whincop (DAL)

Abstract

This paper identifies a feature of neural nets which may be described as the principle of ease of processing (PEP) and which, it is argued, is the primary force guiding a learner towards a target grammar. It is suggested tha the same principle lies at the heart of Optimality Theory, which characterises the course of language acquisition as a progressive reranking of a hierarchy of universal and violable constraints. It is observed that the hierarchy a learner is in possession of at any particular time is the learner's present characterisation of the grammar of the target language and will determine what outputs nets involved in linguistic processing produce for any given inputs to those nets. It is suggested that spatial metaphors may give us a clearer insight into the workings of neural nets, and that we can see the process of self-organisation of nets in accordance with the PEP as a realignment of the positions of linguistic elements in a multidimensional space that is a characterisation of the target language.

1. Introduction

What form does language take inside people's heads? How are concepts, ideas and meanings instantiated? How does the learner unconsciously discover linguistic rules? What is it that leads a speaker to be able to judge the relative grammaticality of an utterance?

All these questions can be answered (though not yet fully) if we imagine the brain as a device that self-organises in accordance with what I shall call the Principle of Ease of Processing (PEP), in the sense of minimising energy expenditure in converting an input to an appropriate output. By input/output, here, I do not mean speech we hear and speech we produce. Instead, I am referring to input to a neural net (from other nets) being converted to output which then serves as input to other nets. It will be suggested in this paper that the PEP is a characteristic feature of neural nets in general (biological as well as artificial), and as such is the primary force that drives and guides the course of language acquisition.

2. Language in the environment, and language in the brain

We can make a clear distinction between two types of language - language that exists outside, and language that exists inside, people's heads (roughly, E-language and I-language, respectively - see, for example, Chomsky 1986:19-24; 1995:15-17). The former exists as speech, signs and writing, the latter as patterns of neural activity. The former always results from the latter, and in as much as the language we read or hear (or see, if in the form of sign language) causes neural activity within us, we may view speech and its correlates as intermediaries between the neural activity of speaker/signer/writer and an audience. Indeed, we can say that language as it exists outside our heads is a representation of the language inside our heads from which it results.

Language inside our heads is representational too; representational of the external language that causes it, representational of ideas and concepts in the speaker/signer/writer that we take the external language to represent, and representational of ideas and concepts within ourselves that are in themselves representations of yet other thoughts, ideas, feelings and other mental events together with representations of the external world of which speakers and the external language they produce are a part.[1]

Language is a highly complex representational system that is an integral part of a larger, even more complex representational system that enables us to make sense of our experiential world (see, for example, Bickerton, 1990). Given that it is so complex, how is a child capable of learning the language of her linguistic community with such evident success? This question has long been investigated and the conventional answer is that the child comes equipped with a Language Acquisition Device (LAD), which is taken to be primarily composed of knowledge structures (principles) that are applicable to any and all human languages (by means of setting parameters). These hypothesised knowledge structures are known as Universal Grammar (UG) (see, for example, Chomsky, 1981, 1986, 1988). As Haegeman puts it:

> Given that neither formal teaching nor overt evidence seems to be the source of the native speaker's intuitions, it is proposed that a large part of the native speaker's knowledge of his language, i.e. the internal grammar, is innate.
>
> (1994: 12)

The appeal of the UG hypothesis is strong, particularly if we accept that the learner's ultimate knowledge is underdetermined by positive evidence in the primary linguistic data (PLD) (Wexler, 1991), and that the linguistic data we are exposed to contains a fair degree of <u>noise</u>, in the sense of ungrammatical or otherwise erroneous input (slips of the tongue, false starts, etc.) that the learner should not use as a basis for determining the target grammar.

3. <u>UG as a genetic endowment</u>

The view of UG that seems to be commonly held amongst many of its advocates is that it is innate knowledge that is built in from the start (the 'strong continuity' view - see Hyams, 1983; Pinker, 1984), and that the process of acquisition is largely one of selection and elimination in response to triggers in the PLD. After all, Chomsky identifies UG as 'the initial state', or theory of the initial state (see, for example, Chomsky 1980; 1995:14). This mainstream view of learning is described as <u>regressive</u> by Quartz and Sejnowski (1995), in that it portrays development as essentially a two-stage phenomenon, with the first stage (preparing the 'initial state') being one of constructing a rich neural pool of prerepresentations as a result of genetic and epigenetic processes, followed by a stage in which the prerepresentations are selectively eliminated (in response to the processing of the PLD) until a subset of the original pool is arrived at which is thought to underlie, though not fully characterise, adult competence.

The conventional portrayal of language acquisition, then, is that neural structures that embody knowledge systems that fit the data are selected for use, and that those that do not fit are, if not actively eliminated, allowed to deteriorate. This would seem to account for the apparent unavailability of UG in older L2 learners that is evidenced by their general inability to reach native-speaker-like competence (as shown by both inferior performance and differences from native speakers (NSs) in grammaticality judgement tasks). We might say, for example, that the L2 learner has difficulty in determining the target grammar because the structures that underlie a characterisation of that grammar are no longer fully available. Despite its attractions, such a view is, I believe, mistaken, as I shall attempt to show in the next section.

4. <u>The rejection of extreme nativism in favour of constructivism</u>

Appealing though it is, the strong continuity view of selecting existent knowledge structures that better fit the PLD and eliminating their inferior rivals, is very probably not well-founded as it stands. Recent neurobiological evidence speaks against it, as Quartz and Sejnowski (1994, 1995) amply demonstrate. Unfortunately, space does not allow more than a brief outline, but the conventional view of a proliferation of neural growth followed by extensive axonal and dendritic arborisation (i.e. growth and branching of the extensions of nerve cells) and subsequent massive neuronal death is, they argue, false. So too is the assumption that the cortex contains a variety of structures that are, somehow, an array of knowledge systems which may be selected in response to the environment (including the linguistic environment) the individual finds itself in. This is evidenced by recent findings that the cortex 'is largely equipotential at early stages' (Quartz and Sejnowski, 1995: 28), the assumption being that equipotentiality does not equate with variety.

What is salvageable from the UG account is that basic structures capable of forming primitive representations grow and develop throughout the course of acquisition, rather than a rich (and varied) set of knowledge structures that characterise core elements of any and all human languages being 'hard-wired' and ... learner from the start. The ways in which the primitive structures develop are determined not only by genetic factors, but by the interaction of these structures with the environment. In attempting to process information derived from the environment, neural nets reorganise (ultimately through progressive arborisation of both axons and dendrites, together with changes in number and position of synaptic connections) to become structures that appropriately deal with the input they get as a result of the environment they are in, in progressively more efficient ways. There will still be a degree of 'pruning' of inappropriate connections, but the emphasis is more to be placed on environmentally determined growth.

Such thinking fits well with a 'maturational view' of UG (such as that espoused by Clahsen*et al.*, 1994), as well as with recent ideas concerning the 'initial state' in SLA (see, for example, Schwartz and Eubank, 1996; Vainikka and Young-Scholten, 1996; Schwartz and Sprouse, 1996; Eubank, 1996). If we look at language acquisition as largely a process of specialisation through growth, rather than a selection and pruning of what is already built in, the initial L2 state will have less and less in common with the initial L1 state, the later L2 acquisition begins. The L2 learner will be obliged to make use of the growth (and thus knowledge structures) already established by the acquisition of the L1, and interference will be inevitable.

To reiterate, the account with which we may replace the regressive/selectionist/eliminativist/extreme nativist UG hypothesis is one in which the neural structures that embody linguistic knowledge develop and evolve in response to the processing of PLD. The principle that governs the development of such structures - indeed, it may be seen as the very engine of learning itself - is one of maximising efficiency in neural nets responsible for linguistic processing. The idea is that any net will develop a structure, through progressive self-organisation, that will process the types of inputs it receives to produce appropriate outputs with minimal expenditure of energy.

5. Energy Sheet Topologies and neural nets

Representations (and constraints and processing propensities) in nets are taken to be distributed (see, for example, Hinton & Anderson 1989/1981; McClelland and Rumelhart, 1985, 1986; Rumelhart and McClelland, 1986; Churchland, 1986; Schwartz, 1988; Smolensky, 1988; Bechtel and Abrahamsen, 1991; Aleksander and Morton, 1993), in that they are not to be identified with the activation of single neurons. There are no 'grandmother cells' (i.e. single cells the activation of which is associated with a single composite concept such as 'grandmother' - see Anderson and Mozer, 1981; Churchland, 1986). How then are we to imagine the operation of distributed representations and distributed 'soft' constraints? The answer is to think of a net as a device for converting inputs to outputs. Inputs and outputs will be distributed too, but we can think of an input of type x entering an energy landscape at point x (or in region x, depending on how specific we want our notion of a particular input to be) with inputs of other sorts entering at other point

Imagine a rubber sheet that reflects the degree of energy required to convert inputs to outputs (technically referred to as a 'Lyapunov sheet' in the literature, see e.g. Kosko, 1992: 77). Inputs of different kinds will fall on the sheet in different places. Outputs will be from the lowest points in areas that 'capture' particular inputs. The idea of capturing leads us to speak of dips or wells in the sheet as basins of attraction. Any input falling within the mouth of such a basin will lead to an output of a type associated with the lowermost point of that basin, unless there is interference from conflicting constraints in other sheets involved in the processing. With a simple two or three dimensional topology, inputs and outputs are pretty much clear cut.

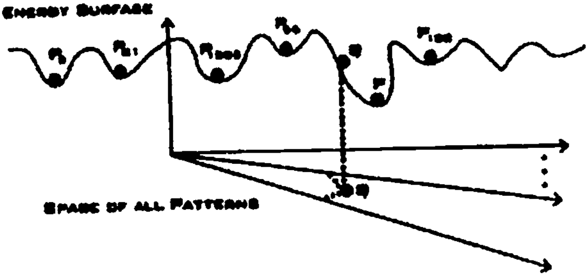

Figure 1a. Here, the input pattern Q will move down the slope of the basin of attraction as the system settles and result in the output P (from Kosko, 1994: 212).

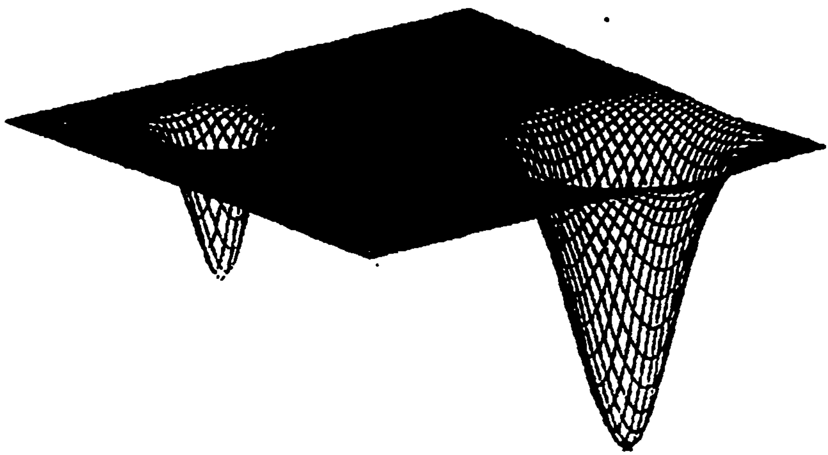

Figure 1b. Basins of attraction in an energy sheet. Imagine an input as being like a ball placed on, or falling onto, the sheet. Any input falling into the area of the basin will result in an output corresponding to the lowermost point of that basin. (From Wasserman, 1993:18)

Topologies determine processing propensities and reflect sets of 'soft' constraints. Simply (albeit inaccurately) put, if we imagine a topology in three dimensions, the peaks and slopes represent areas of constraints against processing, where more energy is required to convert input to output, while the valleys and wells represent propensities, areas where the conversion of input requires less energy.

This is taken to reflect the fact that learning in brains is thought to reside in protein synthesis that strengthens appropriate synaptic connections and thus eases the evocation of patterns of activation that are representations of what has been learnt or, alternatively, lead to replications of the behaviour that has been learnt (Rose, 1993). In either case, we can say that the pattern of connectivity creates a topology in which constraints and propensities are inherent, and that this is how knowledge of any kind is instantiated.

How true this view of learning is is difficult to say, but it is the view that is presently favoured by most researchers, and a great deal of evidence appears to support it (see Rose, 1993 for an overview). It is often referred to as the Marr-Albus theory (Marr, 1969; Albus *et al.*, 1989), which states that conversion from short term memory (STM) to long term memory (LTM) takes place as a result of long term potentiation (LTP) or long term depression (LTD), or a combination of the two.

LTP results from connections in a net being strengthened, making the transmission of signals from neuron to neuron easier, while LTD is the reverse. We can associate LTP (known to take place in the hippocampus) with creating valleys or wells in the topology, or deepening existing ones, and LTD (known to take place in the cerebellum) as creating peaks, shelves or plateaux, or raising the height of existing ones. LTP thus illustrates the role of the PEP in driving learning in that the strengthening of connections responsible for appropriate processing makes subsequent processing of a similar type easier. LTD does not, however, violate the PEP. It is simply an alternative method for sculpting energy sheets.

The task of the language learner is to develop energy sheet topologies that characterise knowledge of the target language by making progressive changes in neural architecture. The changes made will be in accordance with the PEP in that the learner will maximise the extent to which linguistic constraints are satisfied in the processing of the PLD.

6. Language Acquisition as a search problem in Hypothesis Space

The course of language acquisition may be characterised as a search problem through a multi-dimensional space (a hypothesis or acquisition space), each point in the space representing a permissible human language grammar (Gibson and Wexler, 1994; Turkel, 1995). One of the attractions of UG and Principles and Parameters Theory is that the hypothesis space to be searched will be limited. One possibility is to characterise the learner as an error-driven hill-climber, searching for a global optimum in accordance with Greediness (see Gibson and Wexler, 1994) - moving only to positions that reflect an improved analysis of the data at hand - and the Single Value Constraint (SVC) of Clark (1992) - moving to a grammar which differs minimally from the one presently entertained, i.e. to one that is a near neighbour in the search space. The problem for such a model is that the learner (even in relatively simple, idealised. spaces) is likely to encounter traps -areas of the hypothesis space which are local optima, and from which, given Greediness and the SVC, she cannot escape in that no near neighbour grammar is better than that characterised by the present position (see, for example, Gibson and Wexler, 1994; Pulleyblank and Turkel, in press).[2]

Various solutions may be suggested for dealing with this problem, one being the use of Genetic Algorithms (Clark, 1992; Pulleyblank and Turkel, in press)[3]. Whatever formulation is chosen, however, the learner needs some criterion by which to evaluate the superiority of a proposed position in respect to the one presently occupied. In fact, it is not a conscious choice on the part of the human learner, but a choice made by nets responsible for linguistic processing in accordance with their nature. The choice is based on the fact that nets self-organise by following the PEP.

Optimal structures are those that maximise harmony, harmony being the extent to which constraints are minimally violated and maximally satisfied when converting input to output. If, as a result of a proposed change, the harmony of the system in processing the PLD is increased, the change will be adopted. This feature of the system is the PEP.

Thus nets involved in linguistic processing will happen upon structures that better fit the PLD in that they are more suited to an analysis of the data than those the nets presently employ. In so doing they will develop linguistic knowledge. The greater the efficiency in converting inputs to outputs, and the greater the appropriacy of the conversion, the more accurate is the set of neural structures as an embodiment of linguistic knowledge.

Such an account is very much in line with the thinking behind Optimality Theory (OT), a relatively new and fast growing research programme within Linguistics, as well as being one which fits well with the neurobiological facts as we know them.

7. The Optimality Theoretic view of language acquisition

OT views linguistic knowledge as being instantiated in hierarchies of a finite set of violable universal constraints (see, for example, Prince and Smolensky, 1993, Tesar, 1995, Legendre *et al.*, 1995). In any hierarchy, some constraints will dominate, or outrank, others, and thus will be respected over lower ranking constraints with which they conflict. We may call such structures Constraint Domination Hierarchies

(CDHs). Differences between languages are thus explained by variation in the CDHs that characterise them. In consequence, OT views language acquisition as a process of progressive reranking of constraints in a hypothesised CDH in determining a CDH that satisfactorily accounts for the PLD of the target language that the learner is exposed to, together with the acquisition of the lexicon (Tesar and Smolensky, 1993).

7.1 The size of OT spaces

Because OT hypothesis spaces are so large and complex, learning techniques based on brute force serial searches or chance may be ruled out - the learner would in all likelihood be exceedingly old by the time an appropriate CDH was happened upon. As Pulleyblank and Turkel point out:

> In constructing a theory of parametric variation, changing the number of binary parameters from N to $N+1$ doubles the number of possible grammars. Adding a constraint to a system of N constraints results in $N+1$ times as many grammars. To get an idea of the magnitude of the space, consider a learner which enumerates the possible grammars, and is able to test one grammar per second. On average, the enumerative learner will have to test 1/2 of the grammars before finding the target. For a system of 5 constraints, the learner would take about 1 minute to find the target grammar. The average learning time goes up to about 231 days for a system of 11 constraints. For a system of 20 constraints, this learner would take about 38.5 billion years.
>
> (Pulleyblank and Turkel, in press: section 4)

In view of such considerations, the learner's search of the hypothesis space cannot be random or enumerative. Instead, the learning process must be principled, and the most likely principle is that which governs learning in artificial neural networks; the strengthening of appropriate connections and the weakening or elimination of inappropriate ones, appropriate connections simply being those that more easily produce appropriate outputs from the inputs received.

To an extent, such an answer does, admittedly, beg the question of how new connections are formed in order to be tested to see whether they are appropriate. I suspect that the learner builds, by means of the action of interneurons[4], a number of virtual patterns of connectivity which represent CDHs that are near neighbours of the present position and tests these out (i.e. sees to what extent they are superior in analysing the data by judging which ones process input based on the data with a greater degree of harmony). The learner might then move to any position that is found to be superior to the one presently occupied by physically instantiating either connections that lead to a real pattern of connectivity that embodies the virtual pattern tested, or by strengthening the connections that facilitate the pattern of firing of the interneurons that are responsible for creating the superior virtual pattern of connectivity that is built on the real pattern that exists. We could also imagine a blending of the two scenarios with a progressive development of the real pattern under the influence of the maintained success of the virtual in processing input based on the PLD.

7.2 CDH reranking compared with parameter setting

It is also likely that constraints are not reranked independently but as sets of tied changes; rerankings of sets of related, interdependent, constraints. Constraints, then, may be seen as coming in families. The parallel with the idea of parameters (particular 'settings' of sets of interdependent constraints) and principles (constraints) is striking. The fact that the position of some constraints in a CDH must be dependent on the position of certain others is obvious if we consider the fact that many constraints are conflicting - there are pairs of constraints in which not both may be satisfied - the satisfaction of one rules out the satisfaction of the other. A simple example is the constraint pair Repeat and *Repeat (do not repeat) (see Yip, to appear, for a discussion of various types of repetition and its avoidance in Javanese). One either repeats, say, a word in a speech stream and thus satisfies Repeat-w (repeat word) and violates *Repeat-w, or one doesn't, in which case *Repeat-w is satisfied and Repeat-w is violated. Repeat /*Repeat can never occupy the same stratum in a CDH; at any moment in time one must outrank the other. This is not to say that one must always outrank the other. Sometimes we may wish to repeat a word for the sake of emphasis ("It's a long, long way.""It's very, very complex."), at other times we may not. At least some elements of our CDH must, therefore, be variable, depending on the context and what it is that we wish to express. How might such a change in CDH structure be achieved? One possibility is that we continually create a virtual pattern of connectivity in our linguistic nets (as suggested above), and that this pattern of connectivity embodies the CDH of the moment. This virtual pattern may be varied in some respects, at any given moment, through the intervention of interneurons acting in a way similar to so-called Sigma-Pi units as outlined in Rumelhart and McClelland (1986: 73/74). Admittedly, things are unlikely to be quite so simple, and it is more probable that changes are effected in a number of different ways by many groups of neurons acting in parallel and in competition with the influences of yet other groups.

A way of imagining how such sets of changes in a CDH may take place is by considering linguistic elements as representations that exist in multi-dimensional space, and the changes being realignments of clusters of elements within that space, as we shall see in the next section.

8. A-theory and realignment of elements in Language/Linguistic Space

Culicover and Nowak (1995) set out arguments for a conception of language acquisition based on 'adaption' that they consequently term 'A-theory'. According to their view, knowledge of language is built up through the establishment of representations and links between representations in what they term Linguistic Space, in response to frequency of forms in the PLD.

The links between lexical categories establish permissible trajectories through Linguistic Space. The more closely aligned and parallel the trajectories between categories (the categories themselves being clusters, ultimately, of representations of lexical items), the more confidence the learner will have in the 'rule' expressed by the 'envelope' of trajectories. Thus the learner will judge a novel sentence as being grammatical if it is composed of lexical items that can be assigned to the clusters of linguistic categories that lie along those permissible trajectories.

For such a system to work, linguistic items with similar properties are taken to be represented in the same region of representational space. Culicover & Nowak call this the Local Optimization Principle:

> A representational space tends to self-organize in such a way that elements with similar properties are relatively close to one another.
>
> (*op.cit.* : 22)

An important point here is that of 'self-organization'. If properties of a particular set, A, of representations in an area of Linguistic Space presently also occupied by representations B are realised to make A more similar to the set of representations X than was realised hitherto, the space will reorganise so as to bring A and X together. This will entail either locating A, B and X together in the same region of space, or moving A to X, or moving X to A and relocating B. The choice made will depend on how optimal (in terms of maximising harmony/ease of processing) the subsequent alignment of trajectories is, and is clearly analogous to parameter setting. If the alignment of trajectories formed by moving A to X proves a better basis for analysis of experienced data than grouping A, B and X together, or moving X to A and relocating B, this will be the preferred move. It will be preferable in that, in accordance with the PEP, it will provide a more efficient configuration in which fewer constraints are violated in processing the data.

9. OT and multi-dimensional energy sheet topologies

We can begin to see how multi-dimensional energy sheet topologies relate to an OT view of linguistic processing if we imagine that the (initially weak) patterns of activation (the candidate set) created by any input will resonate and ultimately settle on a strongly activated pattern that requires least energy to be the output of the net, i.e. a pattern that has the greatest *harmony* with the topology, the one that satisfies the constraints that operate in the relevant part of the topology to the greatest extent or violates them minimally (Smolensky, 1986; Prince and Smolensky, 1993). The optimal candidate for output is therefore the pattern of activation that maximally satisfies or minimally violates the constraints in the energy sheet topology.

It is important to remember that nets map inputs to outputs according to the energy topologies they create. According to OT, a 'grammar is a specification of a function which assigns to each input a unique structural description or *output*' (Tesar, 1995: 1). Smolensky (1995: 1) states that the "*Universal* Components of Grammar" (my italics) are:

a. Input Set

b. *Con*: Constraint Set

c. *Gen*: Candidate Set for each input

d. *H-eval*: Formal procedure for evaluating Harmony/optimality

(*ibid.*)

Smolensky goes on to state:

> OT is a theory of how one level/component of a structural description is projected from another: optimal satisfaction of ranked and violable constraints.

(*ibid.*)

As we saw in section 5, the idea of energy sheets reflecting the conversion of inputs to outputs is a good way of imagining the effect of distributed soft constraints and other forms of representation within a net. If we imagine a more complex situation in which many energy sheets intermingle in the same multi-dimensional space, Linguistic Space, we come closer to an image that captures the OT view of inputs creating candidate sets of patterns of activation which, by a process of competition, ultimately lead to an output being selected which is deemed optimal in terms of constraint satisfaction and therefore grammatical given the CDH or set of topologies that applied in the selection process.

Optimal output candidates will be ones that exhibit the greatest harmony. Different parts of the system will come up with optimal candidates, based on input to those parts of the system and the topologies in the resolution of Linguistic Space that is reflective of those parts. The optimal candidates will then become the input for other parts of the system. We will attempt to portray this schematically in the next section.

10. What do neural networks do?

Neural nets map vectors[5] in regions of multi-dimensional space (called, alternatively, vector space or state space - see Churchland, 1986, though she predominantly uses the term phase space). Input vectors are mapped to output vectors. In cell assemblies pertaining to linguistic processing, many types of vector mapping are carried on in parallel. The chain is unlikely to be so simple, but the following elements seem likely:

auditory stimulus ⇓

creation of candidate set for output in auditory space ⇓

creation of candidate set in phonetic space (phonetic feature mapping from auditory space) ⇓

creation of candidate sets in morphemic and thus lexical spaces ⇓

creation of candidate sets in semantic, syntactic, logical and conceptual spaces

This will then lead to feedback through the chain to make the whole system settle on a solution (thus making all inputs together make sense - comparing the 'as it appears to be' with what it 'should' be). A more familiar way of expressing this process might be as follows:

hits eardrum, signals travel to Wernicke's area ⇓

possible morphemes in the phonological/morphological spaces are primed ⇓

possible lexical items are primed ⇓

suitability of primed lexical items is checked in semantic, syntactic and logical areas ⇓

morphemes are decided upon, given feedback from semantic via lexical areas ⇓

best choice lexical items are identified

Production in the form of speech, would presumably involve elements of the following:

generation of possible output candidates in conceptual, semantic and logical spaces ⇓

generation of candidate sets in syntactic and lexical spaces ⇓

generation of candidate sets in morphological and phonological spaces ⇓

feedback (this probably goes on from the beginning of the process) ⇓

strong priming of (reduced number of) syntactic, lexical and thus phonological candidates in their respective spaces ⇓

priming (activation of candidate set) in Broca's area for production of sound stream ⇓

final feedback loops and settling of system, resulting in selection (strong activation) of optimal outputs ⇓

production (speech)

This is, of necessity, highly simplified, but gives us a basis for discussion and examination, and is a useful starting point for progressive refinement.

The diagram is a schematic attempt to portray a dynamic system in which all the constraints exist and in which they always play a part, but the state of the system at any particular moment in time will influence the degree to which some constraints are respected or violated. This being the case, there will be subtle differences in the CDH which are <u>context</u> dependent. This is not necessarily to be linked only to particular lexical items, it is likely that emotional, semantic and pragmatic factors will also play a part, as we shall outline in the next section.

11. <u>One or many CDHs?</u>

The idea of a speaker entertaining a number of CDHs is at first sight problematic in that it would require alternate, virtual, patterns of connectivity, but it seems likely that that is precisely how we are able to speak more than one language. To what extent we use the same nets for processing two or more languages is still an open question, but it is impossible to seriously contemplate the systems for two languages being entirely separate. At the very least, we will use the same systems of conceptual and logical space. We will use other spaces too, to the extent that this is practical. It is possible to imagine a separate lexical space for an L2, but it is doubtful, even in the case of lexical items, that total separation is possible. It is inevitable that we will strive to find similarities between lexical items in our L1 and those in the L2. Every time a similarity is discovered, there will be a point of intermingling of L1 and L2 space.

Even if we speak two languages quite well, there will be very little mixing of the two in production if we live in a setting where we normally only speak one of the languages, so it seems that if the two languages, determined by their respective CDHs, do somehow exist in the same space we can automatically restrict selection to only one language within a space. How might this be done? One possibility is that interneurons set up virtual patterns of connectivity within a net - in this sense the L1 and L2 would be like virtual machines run on the real machine of the net.

Sometimes our selection mechanism relaxes - if, for example, we are trying to communicate with a native speaker of the L2 who also speaks our L1. In such circumstances it is not uncommon, and nor is it unreasonable, for a degree of mixing to take place. Obviously, this does not mean that we mix indiscriminately - such a policy would quickly lead to a breakdown in communication. It is not a haphazard affair, but something of an ongoing negotiation with both languages being used in an attempt to maximise communication (and perhaps also express empathy).

We tend to stick to one language in a monolingual setting for the simple reason that it would be unreasonable for us to express ourselves in a language that our interlocutors didn't understand. Even in an L1 setting, however, we rarely speak precisely the same language with everyone we try to communicate with. If I return to the region I grew up in, Yorkshire, and meet local people in a pub, I notice not only that my accent (rhythm, pronunciation, intonation, speed) changes, but that I also use words and expressions that I would not normally use. Does this mean that we entertain a whole panoply of subtly different CDHs? This seems rather doubtful in that it would require that we have an alternate pattern of connectivity for every CDH. What seems more likely is that the CDH is the (perhaps virtual) topology of the moment and the topology is variable in as much as the pattern of connectivity can be influenced by context. In this sense, our CDH is constantly under review and being changed by what we have just heard, how we have reacted to it, what we have decided we want to express and how, etc.

If this is true, we do not entertain countless CDHs, i.e. countless CDHs are not physically instantiated in our heads, but we possess a system which is able to variably evoke countless variations in energy sheet topologies, CDHs, though only one CDH exists at any particular moment.

Rather than viewing the learner as having a single grammar/CDH, it is perhaps preferable to imagine her being in possession of a cluster of potential CDHs, remembering that we are constantly changing our CDHs in subtle ways in response to all input processed. We may view such a cluster as a sort of candidate set from which to choose, noting that much of any CDH chosen will have a great deal in common with the others. In this sense we might imagine the cluster being a set of variable elements of the favoured CDH, any of which can be evoked at any particular moment of time.

12. Conclusion

At first sight, it seems obvious that knowledge must emanate from either the organism (that which is built in, i.e. innate), or the environment. Given the poverty of the stimulus argument, the appeal of the UG hypothesis is strong, and it would be folly to reject it without determining a mechanism for explaining aspects of a NS's knowledge that cannot be accounted for by positive evidence in the PLD. At the same time, we should not presume that such knowledge, if not derivable from the input, is necessarily built in. The third possibility, is that the knowledge arises from the interplay between what is built in and the environment. This might be termed a feature of the dynamic system that the linguistic system is.

Such reasoning is unlikely to appeal to those who would like to keep the world conceptually simple - it is satisfying to draw clear lines between domains, and to limit domains in number - but we should not ignore good evidence, and we ignore the insights derived from machine learning and the neurosciences at our peril. The fact is that the real world is not so simple and sometimes refuses to fit into clear conceptual categories that are easy for us to grasp. The positive corollary is that the real world, in its complexity, is found to be more interesting, and thus even more worthy of investigation.

References

Albus J., P. Kanerva and E. Loebner. 1989. 'Cerebellar Models of Associative Memory'. In Three Papers from IEEE Compcon Spring '89. (ed.) M. Raugh. RIACS Technical Report 89.11. Research Institute for Advanced Computer Science, NASA Ames Research Center.

Aleksander I. and H. Morton. 1993. Neurons and Symbols: the stuff that mind is made of. London: Chapman & Hall.

Anderson J. and M. Mozer. 1981. 'Categorization and selective neurons'. In G. Hinton and J. Anderson (eds.) 1989/1981. 251-74 (in updated edition, 213-36 in original).

Bechtel W. and A. Abrahamsen. 1991. Connectionism and the Mind: An Introduction to Parallel Processing in Networks. Oxford: Blackwell.

Bickerton D. 1990. Language and Species. Chicago: The University of Chicago Press.

Chomsky N. 1980. Rules and Representations. New York: Columbia University Press.

Chomsky N. 1981. 'On the representation of form and function.' The Linguistic Review. 1/1: 3-40.

Chomsky N. 1986. Knowledge of Language: Its Nature, Origin, and Use. Dordrecht: Foris.

Chomsky N. 1988. Language and Problems of Knowledge. Cambridge, MA: MIT Press.

Chomsky N. 1995. The Minimalist Program. Cambridge, MA: MIT Press.

Churchland P. S. 1986. Neurophilosophy: Toward a Unified Science of the Mind-Brain, Cambridge, MA: MIT Press / Bradford Books.

Clahsen H., S. Eisenbeiss and A. Vainikka. 1994. 'The seeds of structure: a syntactic analysis of the acquisition of Case marking'. In T. Hoekstra and B. Schwartz (eds.). Language acquisition studies in generative grammar: papers in honour of Kenneth Wexler from the 1991 GLOW workshops. Philadelphia, PA: John Benjamins, 85-118.

Clark R. 1990. 'Papers on Learnability and Natural Selection'. Technical Reports in Formal and Computational Linguistics. 1. Geneva: University of Geneva.

Clark R. 1992. 'The selection of syntactic knowledge'. Language Acquisition 2: 85-149.

Culicover P. and A. Nowak. 1995 'Paradoxes and Puzzles of Triggering'. ms. Ohio State University, draft version of 9th May, 1995. Available from http://www.cog.ohio-state.edu/homepage/culicover.wip.html as paradox.ps.

Eubank L. 1996. 'Negation in early German-English Interlanguage: more Valueless Features in the L2 initial state'. Second Language Research 12, 1: 73-106.

Gibson E. and K. Wexler 1994. 'Triggers'. Linguistic Inquiry 25: 407-454.

Haegeman L. 1994. Introduction to Government and Binding Theory. (2nd ed.) Oxford: Blackwell.

Hinton G. and J. Anderson. 1989/1981. Parallel Models of Associative Memory. (Updated edition). Hillsdale: Erlbaum.

Hyams N. 1983. 'The acquisition of parameterised grammars'. PhD dissertation, City University of New York.

Kosko B. 1992. Neural Networks and Fuzzy Systems: a dynamical systems approach to machine intelligence. London: Prentice-Hall International.

Kosko B. 1994. Fuzzy Thinking. London: Harper Collins.

Legendre G., C. Wilson, P. Smolensky and W. Raymond. Forthcoming. 'Optimality and wh-extraction'. To appear in Papers in Optimality Theory, Beckman, Urbaczyk, & Walsh (eds.), University of Massachusetts Occasional Papers.

Marr D. 1969. 'A theory of cerebellar cortex'. Journal of Physiology. 202: 437-470.

McClelland J. and D. Rumelhart. 1985. 'Distributed memory and the representation of general and specific information'. Journal of Experimental Psychology: General. 114/ 2: 159-88.

McClelland J., D. Rumelhart and the PDP Research Group. 1986. Parallel distributed processing. Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models. Cambridge, MA: MIT Press/Bradford Books.

Pinker S. 1984. Language learnability and language development. Cambridge, MA: Harvard University Press.

Prince A. and P. Smolensky. 1993/forthcoming. 'Optimality Theory: Constraint Interaction in Generative Grammar'. RuCCs Technical Report #2, Rutgers University Center for Cognitive Science, Piscataway, New Jersey [also to appear in MIT Press].

Pulleyblank D. and W. Turkel. In press. 'The logical problem of language acquisition in Optimality Theory'. In Is the Best Good Enough? Optimality in Syntactic Theory. Cambridge, MA: MIT Press & MIT Working Papers in Linguistics.

Quartz S. and T. Sejnowski. 1994. 'Beyond modularity: neural evidence for constructivist principles in development'. Behavioral & Brain Sciences 17: 725-726.

Quartz S. and T. Sejnowski. 1995. 'The neural basis of cognitive development: a constructivist manifesto'. Unpublished article submitted to Behavioral & Brain Sciences, available from http://salk.edu/cgi-bin/pub-search#unpublished-articles.

Rumelhart D., J. McClelland, and the PDP Research Group. 1986. Parallel distributed processing. Explorations in the microstructure of cognition: Vol. 1. Foundations. Cambridge, MA: MIT Press/Bradford Books.

Schwartz B. and L. Eubank. 1996. 'Introduction: What is the 'L2 initial state'?'. Second Language Research 12, 1: 1-5.

Schwartz B. and R. Sprouse. 1996. 'L2 cognitive states and the Full Transfer/Full Access model'. Second Language Research 12, 1: 40-72.

Schwartz J. 1988. 'The new connectionism: developing relationships between neuroscience and artificial intelligence'. In S. Graubard (ed.) The Artificial Intelligence Debate, 123-141, Cambridge, MA: MIT Press.

Smolensky P. 1988. 'On the proper treatment of connectionism'. Behavioral and Brain Sciences. 11: 1-74.

Smolensky P. 1986. 'Information processing in dynamical systems: Foundations of Harmony Theory'. In D. Rumelhart, J. McClelland, and the PDP Research Group (eds.) Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations. 194-281. Cambridge MA: MIT Press/Bradford Books.

Smolensky P. 1995. 'On the internal structure of the constraint component *Con* of UG'. Available at the Rutgers Optimality Archive as construc.ps. Version of 15th September, 1995.

Tesar B. 1995. 'Computational Optimality Theory', unpublished Ph.D. thesis, University of Colorado. Available from the Rutgers Optimality Archive as tesar.ps.

Tesar B. and P. Smolensky. 1993. 'The Learnability of Optimality Theory: An Algorithm and Some Basic Complexity Results', ms., University of Colorado, Boulder.

Turkel W. 1994. 'The Acquisition of Optimality Theoretic Systems'. ms., University of British Colombia.

Turkel W. 1995. 'The NK Model of Parameter Space'. ms., University of British Columbia.

Vainikka A. and M. Young-Scholten. 1996. 'Gradual development of L2 phrase structure'. Second Language Research 12, 1: 1-5.

Wasserman P. 1993. Advanced Methods in Neural Computing. New York: Van Nostrand Reinhold.

Wexler K. 1991. 'On the argument from the poverty of the stimulus.' In A: Kasher (ed.)The Chomskyan Turn. Cambridge, MA: Blackwell.

Yip M. Forthcoming. 'Repetition and its avoidance: the case of Javanese'. To appear in K. Suzuki and D. Elzinga, (eds.) Proceedings of the Arizona Phonology Conference 4: Features in OT. Available from the Rutgers Optimality Archive as java.wp6.

---

[1] To clarify this, let us use a few symbols to represent such representations and what they represent. Let's say that there is a speaker and listener and in an exchange the speaker says 'dog'. Let's call this sound stream an E-rep in that it is an external representation (a sign) of what the speaker wishes to convey to the listener. It is a pattern of sound waves in an external medium, usually air. The listener creates a representation of this sound stream in response to the pattern of movements of his eardrum when the E-rep hits it. This is the word 'dog', as it sounds to the listener. Let's call this I-rep1, with 1 indicating the outermost level of representation on the part of the listener. I understand that some readers may already balk at the idea of this experience (created by the sound stream and the listener's sense of hearing) being a representation. Such readers feel that the experience is a sound, and not 'just' a representation of a sound. However, it is evident that the experience (of sound) we have as a listener when we hear a sound, is not the same thing as the (physical) sound which causes the experience. Even if we are reluctant to say that the experience is a representation of the external world 'object' that the sound stream is, we at least have to recognise that sound as a physical entity outside our heads and sound as an experience inside our heads are different entities in a causal chain and exist at different coordinates in spacetime.

[0]So far, we have an E-rep (which we may call "dog"E) and an I-rep1, ("dog"I). Next, the listener identifies the phonetic form (which might be the same for different lexical items) 'dog'. This is the next level of representation, and we may call it I-rep2. Again, some people may balk at this, wondering how the phonetic form of a word can be said to represent anything. Let's say then, that it is a bridge between the raw, in itself meaningless, sense experience of a particular sound and the range of linguistic entities that that sound can represent, which we may call (I-rep3)$^{1-n}$. The linguistic entities will be composed of further sets of representations which we may identify as bundles of concepts pertaining to sense experience and categorisation. In the case of the lexical item dog (as a noun referring to the canine), these will include such things as 'canine', 'animate', 'faithful', 'potentially dangerous' together with a range of visual, auditory, olfactory and tactile qualities based on the individual's experience of what dogs look, sound, smell and feel like. as well as linguistic features such as 'able to take agent or patient roles'. Such concepts lie at a level of representation that underpins I-rep3 and we may call it I-rep4.

Thus we may think of entities that exist at E-rep and I-rep1,2,3 as elements in a causal chain of association, with any element in the chain being such that it is capable of evoking the next. This is what representations do. We may describe this schematically:

Production		**Comprehension**
Speaker	sound stream	Listener
→		
I-rep4 → I-rep3 → I-rep2 → I-rep1 →	E-rep →	I-rep1 → I-rep2 → I-rep3 → I-rep4

[2] Such traps are evidently avoided in normal L1 acquisition, but may play a part in accounting for fossilisation in L2 acquisition.

[3] Genetic Algorithms are a heuristic search method involving the 'breeding' of a population of candidate networks, with individual networks competing with each other on the basis of 'fitness' (their relative success in converting inputs to appropriate outputs) for the chance to 'survive' and 'breed' and thus create the next generation of candidate networks.

[4] Interneurons, as their name suggests, stand between other neurons and affect the connection between them. What I really have in mind is the sort of activity carried out by what Rumelhart and McClelland call 'conjuncts' or 'gated pairs' (see Rumelhart and McClelland, 1986: 73/74) in which there is a branching connection between two neurons, A and B which joins before reaching a third neuron C, that also receives an input from a further neuron, D. We can simulate the activity of such a branching connection if we multiply the outputs of A and B. We can then calculate the net input to C by adding this product to any other inputs C receives from other units. For example, if C's input from the other unit, D, is always +1 (excitatory), if A's output is 1, and B's 0, C's net input will be 1, and if A's input is 1 and B's is also 1, C's net input will be 2. Without the branching connection, i.e. if A and B were each directly connected to C, the net input to C would have been a simple sum of the inputs from A, B and D, giving 2, (1+0+1), and 3, (1+1+1), instead of 1,((1*0)+1), and 2, ((1*1)+1). A link such as that between A and B is called a gate, and the pair of units linked by the gate are referred to as conjuncts (Rumelhart, Hinton and McClelland, 1986: 73) the units as a whole are called Sigma-Pi units.

[5] Vectors, in this sense, are elements of vector space. Mathematically, a vector is a quantity that has magnitude and direction. For example, I can map certain qualities of someone I know by placing a point at a certain distance along a line that represents that quality. Given three qualities, such as kindness, appearance and intelligence, we can place a single point inside a cube that represents the degree to which the person is kind, good-looking and intelligent, the axes (dimensions) representing each of the qualities in turn. If we take this point and map it to another vector space, say one describing attractiveness (if we assume that this depends on how kind, good-looking and intelligent a person is), we have performed the function of a neural net. In this case, the input vectors are kindness, appearance and intelligence, and the output vector is, say, attractiveness.