Languaging Behaviour as Catalytic Process: Steps Towards a Theory of Living Language[1]

[PART I]

Paul J. Thibault[2],

Faculty of Humanities, University of Agder,
Kristiansand, Norway
Faculty of Education, University of Hong Kong,
Hong Kong SAR, PR China

1. On the Reduction of the Material Dynamics of Languaging Behaviour to 'Text'

In many recent accounts, instances of 'language-in-context' are often thought of as 'text' consisting of verbal patterns and, increasingly, 'multimodal' combinations of, for example, images and verbal text (Kress and Van Leeuwen 2001; Baldry and Thibault 2006; Thibault 1994). The concept of 'text' has been an important one in the language sciences since Harris (1952) inaugurated the systematic study of discourse-level relations in language above the traditional level of the sentence. Since the 1970's, many functional linguists and discourse analysts have made the study of text or discourse a central concern. These approaches have often been inspired by Halliday's definition of 'text' as any instantiation – spoken or written -- of the linguistic system in the form of language-in-context:

> When people speak or write, they produce text. The term 'text' refers to any instance of language, in any medium, that makes sense to someone who knows the language …
>
> (Halliday 2004 [1985]: 3)

'Text', which refers to the instantiation of a language system "in any medium", according to this view, is language that is separated from its material expression, e.g. as the co-orchestration and dialogical coordination of bodily dynamics in talk, or the material inscription of some organized arrangement of visible notational marks on a surface in the case of writing. In this view, talk and (written) text are both instantiated as 'text'. One consequence of this reduction is that the radically heterogeneous character of, respectively,

[1] I wish to acknowledge critically important discussions with Stephen Cowley, Bob Lumsden, Mark King, Neil Mercer, Kai Sorfjord, Sune Steffensen, Susan Stuart, and Feifei Zhou. All of these individuals have contributed in invaluable and inestimable ways to the development of the ideas in this paper. Marcel Kinsbourne also made insightful and helpful comments on an earlier draft. Paul Bouissac deserves special mention for his encouragement, patience, and support. I thank and extend my heartfelt appreciation to these individuals for their willingess to listen and to participate in critical exchanges that have helped me to clarify my ideas.
[2] Contact email: pauljthibault@gmail.com

talk and text is reduced to verbal abstracta. Text becomes the transcription of, above all, the lexicogrammatical choices and patterns in both spoken and written language. The term 'text' itself designates an abstract entity devoid of materiality. Moreover, the very different material dynamics of talk and text are conflated in the term 'text'. This conflation does not account for the material dimension of either as something that is intrinsic to language itself. The discussion below will focus exclusively on talk. However, it is important to clarify the issues discussed above in order to establish a new point of departure for thinking about the material dynamics of languaging behaviour – one which is not based on the more usual idea of 'text'.

The conflation of talk and text in transcription practices – both lay and professional – transforms the experience of the dynamics of talk. It is therefore 'non-neutral' for this reason (Ihde 1979: 21). Ihde draws attention to another facet of the phenomenology of technologically mediated observation and measurement such as the transcription practices under consideration here. Ihde refers to the effect of the "amplification-reduction structure" (Ihde 1979: 21) of these technologies. Whilst the technology of written transcription amplifies some features or aspects of the original event of talk, it reduces others. Ihde (1979: 21) points out that this two-sided effect is an invariant feature of all such uses of a technology to probe some phenomenon. The technological probe is not transparent, although this is so in various ways and to various degrees. In any case, the instrument that is used to probe or record the phenomenon changes our experience of the given phenomenon. In all such cases, the amplification tends to be foregrounded whereas the reduction tends to slip into the background to the extent that it may be forgotten or not even noticed (Ihde 1979: 21). In the case of the written transcription of talk-as-text, verbal abstracta and phenomena on relatively gross time-scales (e.g. discourse moves and turns) are amplified whereas the material, time-locked dynamics of events on smaller, faster timescales and their optic, haptic, articulatory, acoustic and other properties (Bouissac 2006) disappear pretty well altogether.

The resulting reification of verbal transcription-mediated-access to the altogether different phenomenon of talk and its dynamical properties has, of course, been changed significantly in more recent times with the use of more dynamic multimodal forms of transcription and simulation (Galantucci 2005; Steels 2006; Cassell and Tartaro 2007). The main purpose of these brief observations above is to draw attention to the power of writing and the associated doctrine of "scriptism" (Harris 1980: 1-18; see below, this Section) in shaping our understandings of talk in terms of the very different phenomenon of text in both scientific and lay practice. This has fed into a widespread belief in discourse-analytical approaches that talk can be transcribed and analysed on the basis of transcriptions and re-descriptions of verbal abstracta, i.e. what people, including analysts, report they have heard or seen, etc., *as if* these reports were the first-order data to be described and accounted for. In actual fact, these reports are second-order re-descriptions of the data (the material, always time-locked bodily dynamics of agents-in-interaction and their emergent inter-individual patterns) based on the description and transcription of verbal abstracta. Such re-descriptions fail to account for the material dynamics of talk and the small, fast time-scales on which these operate and in terms of which their specific properties and functional capacities need to be explained (see Cowley 2008; Steffensen, Thibault and Cowley 2010; Thibault 2008, 2011).

Overall, the amplification-reduction structure in operation here is implicit in almost all of our scientific, pedagogical, and folk-theoretical accounts of language and their associated practices. This is so to the extent that a folk-theory of abstract verbal forms and the so-called science of these are conflated. Language is analysed and taught for the most part in terms of a culturally promoted literacy-based phenomenology of formal abstracta such as words, sentences, texts, and so on that rarely gets beyond re-description according to the local folk-phenomenology of verbal forms and verbal reports about these forms. On the scientific view, these same formal arrangements are the artefacts on which presumed operations of abstract symbol processing are performed, whether mediated either by internal codes or rules in the brain or by external social conventions. Second-order abstracta also function as scripts that are utilized in educational settings to corral and regiment individual agency and creativity by orienting to standardized norms that fail to do justice to the fact that much of human intelligence, decision-making, and creativity are exercised through the materiality of first-order interaction dynamics between persons, not verbal abstracta (Steffensen, Thibault, and Cowley 2010). This has implications for teaching and learning that remain invisible to pedagogies that are based on standardized second-order constructs.

The text-based model of talk invites us to re-contextualize how bodies engage in real-time meaning-making activity as a form of text-to-be-read (Thibault 1994). Talk can better be better investigated and understood in terms of real-time dialogically coordinated bodily dynamics between persons, the forms of whole-body sense-making that this engenders, and the constraints on these (See analyses along these lines in Sections 3 and 31). As we shall see below, a culturally promoted belief in wordings embedded in associated language practices and interaction routines is one such constraint. Moreover, verbal patterns or wordings are disembodied and reified abstracta that inform our meta-semiotic ways of thinking about and analyzing language. These verbal patterns are described by linguists as corresponding to rule-bound regularities of grammatical or discourse levels of organization. Rules, whether descriptive or prescriptive, are, I contend, second-order cultural constructs and regularities. They are not in themselves the primary data, but second-order re-descriptions of the primary data. The resulting ‘texts’ are typically described and analyzed as verbal and other abstract patterns on different levels of hierarchical organization – e.g. phonological, lexical, grammatical, and discursive – arranged in a sequential order. These approaches – e.g. conversation analysis, discourse analysis - emphasize recurrent patterns. They posit that discourse consists of determinate higher-order units, e.g. discourse-level turns or moves. Discourse is thus seen in terms of program-like regularities of such formal units, rather than dynamical, time-bound properties and functional capacities of the co-orchestrated neural and bodily dynamics of persons-in-interaction.

Whole-body-sense-making-making is, therefore, reduced to predictable arrangements of determinate, recursively organized textual or discourse-level units that derive from a higher-order language code. Interaction reduces to text-in-context (Thibault 1994). Events get detached from time, timing, bodies, activity, environmental affordances and human relationships. Moreover, verbal patterns or wordings are disembodied and reified abstracta that come to inform our meta-semiotic ways of thinking about and analyzing language. These

verbal patterns are described by linguists as rule-bound regularities of grammatical or discourse levels of organization. The modeling of talk using analytical and descriptive abstracta that are based on written notation, including phonetic script, accounts in large measure for the modest progress made by twentieth century linguistics in overcoming its "written language bias" (Linell 2005). The bias identified by Linell is founded on the principle of "scriptism", as defined by Harris (1980: 1-18), i.e. the doctrine that writing is privileged over speaking in literate cultures such that writing constitutes the appropriate meta-linguistic model for talking about and understanding languaging behaviour in all of its manifestations.

In such models, the events in which talk is embedded are de-somatised and de-materialized as 'discourse'. Discourse is seen as sequential and recursive arrangements of abstract verbal forms. The model that informs this approach is that of using text 'in' context (Thibault 1994). Text-analysis *per se* is the basis of theory and description. This model confuses how we think about and use some kinds of written texts with how bodies engage in real-time meaning-making activity. Talk can better be better investigated and understood in terms of the material dynamics of real-time dialogically coordinated neural and bodily activity between persons.

2. The Separation of 'Language' from Dynamics: Trager's Distinction between Language and Paralanguage

The separation of 'language' from biologically grounded neural and bodily dynamics rests upon a number of prior assumptions. Two of these will be discussed here. First, there was the methodological decision, traceable to Saussure, to separate bodily processes from the definition of *la langue* as the object of study of linguistics (Saussure 1993 [1910-1911]: 189; see Section 5 below; see also Thibault 1997; Bouissac 2010). Saussure based his idea of *la langue* on the assumption that "it is the union of an idea with a vocal sign which suffices to constitute the whole of the language system" ["c'est l'union de l'idée avec un signe vocal qui suffit à constituer toute la langue"] (Saussure 1993 [1910-1911]: 189). Saussure's justification for this line of thinking is that phonation in the act of speaking is an individual fact that belongs to "the individual faculty" (1993 [1910-1911]: 189) and therefore that facts concerning the bodily articulation of language are not relevant to the study of *la langue*. In making this argument, Saussure failed to see that human vocal tract capacities belong to a population of agents as part of their evolved biological inheritance. Secondly, the theory of language based on the collective abstraction *la langue* that Saussure developed is a theory of formal abstracta (see Section 5). In constituting the object of linguistic science in this way, Saussure sets up, at least implicitly, an a priori distinction between the verbal and non-verbal aspects of first-order interactive behaviour between persons in talk. This distinction has been maintained in very many folk-theoretical and theoretical accounts of languaging behaviour to the present day. This brings me to the second assumption.

Accordingly, linguistics has proceeded on the further assumption that the verbal part of human behaviour can be isolated from the rest of communicative behaviour and that a science of language – viz. linguistics – based on the verbal part can be set up on the basis of this artificial and entirely arbitrary separation of the 'verbal' and 'nonverbal' components of talk.

For example, one of the earliest theorists of paralanguage, George L. Trager, argued that language was "accompanied by other communication systems, one of motion – kinesics, and one of extra-linguistic noises – vocalizations" (1958: 2). On this basis, communication, according to Trager, "was divided into language, vocalization, and kinesics" (Trager 1958: 2). Language, in Trager's framework, was defined as follows:

> Language will be described here only to the extent of saying that it is the cultural system which employs certain of the noises made by what are called the organs of speech, combines them into recurrent sequences, and arranges these sequences in systematic distributions in relation to each other and in reference to other cultural systems. That is, language has sound, shape, and sense.
>
> (Trager 1958: 3)

Language, in Trager's framework, is embedded in or situated in "the setting of an act of speech" (1958:3):

> When language is used it takes place in the setting of an act of speech. Speech ('talking') results from activities which create a background of voice set (1.2, below). This background involves the idiosyncratic, including the specific physiology of the speakers and the total physical setting; it is in the area of prelinguistics (Trager, 1949, 2-3). Against this background there take place three kinds of events employing the vocal apparatus: language (as described); variegated other noises, not having the structure of language-vocalizations; and modifications of all the language and other noises. These modifications are called voice qualities. The vocalizations and voice qualities together are being called paralanguage (a term suggested by A. A. Hill, who has been interested in the development of these studies). Paralanguage is part of the metalinguistic area of activity.
>
> (Trager 1958: 3-4)

Language is, therefore, made to correspond to the traditional three formal levels of linguistic coding that constitute the basic architecture of 'language', so defined, viz. phonology, syntax, semantics (Trager's 'sound', 'shape', and 'sense'). Definitions of paralanguage vary to include non-linguistic aspects of the speech signal and kinesics such as facial expressions, gestures and body postures. In all cases, the basic assumption is that the verbal and nonverbal aspects of speech behaviour can be separated off from each other as separate entities that are 'combined' in speech. The central contradiction of this view would have it that in speech behaviour abstract verbal forms, on the one hand, and other material, nonverbal audible and visible behavioural events, on the other, are 'combined', to echo Trager's term, though without specifying any causal mechanisms for this curious linkage of incompatibles (i.e. how does one combine formal abstracta with material bodily events?). Language in such a view – the classical formalist one – is identified with wordings, with verbal patterns, which are associated with phonological, syntactic, lexicogrammatical, or other form-based abstracta. In such a view, language is a code-like system of inputs and outputs across these levels that gets

separated from cognitive, affective, and bodily dynamics in real-time (Cowley 2008; Thibault 2008, 2011).

3. Developing Alternatives to Formal Abstracta as the Stuff of Language

How can we move beyond models of language based on codes and abstract forms (e.g. verbal patterns)? The theories of distributed cognition developed by Clark (1997) and Hutchins (1995a, 1995b) provided some early indications. Theories of distributed cognition questioned the view that human thinking is based on the internal manipulation of symbols. In the 1980's, connectionism, neuroscience, dynamic systems thinking, and theories of situated and embodied-embedded cognition also began to challenge the view that human cognition works in the same way that computers process information. Distributed cognition puts the emphasis on cognitive dynamics and on the embodied-embedded nature of cognitive dynamics. Where does language fit in? How can we rethink language in the light of this new thinking? Langacker's cognitive linguistic can serve as a useful starting point.

Cognitive linguists such as Langacker (1987: 57) assume that 'grammar' is an internal psychological representation of a speaker's knowledge of the "cognitive routines that are shaped, maintained, and modified by language use" (Langacker 1987: 57). Moreover, Langacker too, like Halliday, grounds his cognitive investigation of language in second-order abstracta. He argues, for instance, that "the functional and phenomenological characterization of mental experience is consequently more directly relevant to linguistic analysis than descriptions that refer to the firing of specific neurons. For linguistic purposes it is often sufficient merely to establish the existence of a higher-order cognitive entity, regardless of how it might arise from more basic processes … " (1987: 99). This is a species of functionalism that separates cognition from the physical means of its realization.

Accordingly, Langacker evidences no concern with how the conventionalized semantic structures of linguistic cognition are integrated to biomechanical and neural activity. Langacker makes a distinction between higher-order cognition, which he links to a phenomenology of abstract semantic structure. Semantic structure can be described without worrying about the physical stuff that realizes it. Thus, functional and phenomenological characterizations of mental experience can be undertaken without reference to the means of physical implementation of these functional and phenomenological properties. Langacker does not mention bodily processes in relation to cognition. With respect to the brain, he claims that cognitive events of sufficient regularity leave neurochemical traces that facilitate recurrence of cognitive events (1987: 100). On this view, neurochemical traces are the physical stuff that implement and realize cognitive events though this stuff is not necessary for the description of the cognitive event itself. Cognitive events are functional states of the brain. They can be realized by multiple neurochemical traces.

Language, including its 'grammar', according to the distributed language view, is related to interactive loops between brains, bodies, and the external environment rather than internal mental states and their representations or abstract verbal patterns as in current discourse-analytical and form-based approaches. Language is distributed between brain, body, and

environment. The emphasis is on concrete, real, living human individuals who are inter-connected with each other and with cultural artefacts and technologies rather than being mediated by abstract codes or systems. The latter are no more than reified generalities that have no explanatory value. Rather than saying that individuals are mediated by shared abstract systems, the sociality of our interactions is defined and is meaningful in relation to how it is integrated with what has gone before, what is going on now, and what is expected or anticipated to happen next. The distributed view of language emphasises how humans coordinate their actions with those of others in the service of common projects, rather than shared codes or abstract higher-order systems that are said to mediate the interaction. As Harris points out, signs, "provide an interface between different human activities" (2000: 69), rather than mediating between persons and an abstract system or code. The emphasis is on individuating dynamics: how do persons in interaction creatively adapt available resources in order to discover new possibilities of meaning-making, including the inventing of new means, new tools, and new semiotic resources for acting on and transforming the undiscovered potentialities of the materials and technologies to hand? The emphasis is thus on creation and invention rather than the mere use of pre-existing abstract codes, programmes and systems.

The latter view cannot account for human beings as persons: it treats them along mechanistic lines as information processors, mere end-users and instantiators of already defined form-meaning relations, and nodes in computational networks. The real challenge, then, is to investigate and to understand how new assemblages of persons, technologies, and learning systems help people to organize their individual and collective behaviours in novel ways – ways that allow them to organize their actions intelligently and ethically as agents who can evaluate and intervene in events and influence and guide them. On this view, persons are not only shaped and constrained by history; they are in history and are shaping it individually and collectively on, potentially, many different scales of place and time.

The concept of *assemblage* derives from Deleuze and Guattari (2004 [1980]). An assemblage consists of a set of historically individuated component parts and their capacities to affect other parts and to be affected by them. The component parts interact to form a new, emergent whole that is irreducible to the component parts. A key property of such newly emergent wholes is the heterogeneity of the component parts. This means that the parts retain a relative autonomy and identity with respect to the whole in which they function. Unlike structural-functional theories in biology, sociology, anthropology, linguistics, and so on, which assume that the component parts cannot exist independently of their intrinsic relations of interiority (DeLanda 2011: 3) that they have with each other, as assemblage is an emergent whole in which the component parts retain a relative autonomy. As DeLanda (2011: 4) explains, this means that the component parts can be detached or 'unplugged' from one assemblage and attached to another thereby entering into new relations and new interactions. An assemblage is defined by the interactivity between the component parts in terms of relations of exteriority (DeLanda 2011: 3). In an assemblage, the component parts are all historically individuated entities or component processes, not abstract types. Individual persons are individuals in this sense, but so are the events, the objects, the artefacts, etc. with which persons interact to form what I shall call social-affective-cognitive assemblages (Section 28). Assemblages can exist

on many different scales. A conversation between two or more persons is an assemblage. Interpersonal networks, social institutions and historical languages are all assemblages that exist on different spatio-temporal scales. They are all historically individuated entities that have a determinate temporal duration and spatial extension on some scale.

Languaging behaviour is grounded in biomechanics and the natural interactivity and expressivity of our bodies and brains. Discourse-analytical approaches privilege rule-bound sequences of abstract verbal patterns thereby separating such patterns from the individuating dynamics of the social assemblages that are constituted in dialogically coordinated interaction between persons and the matter-energy flows and their modulations that these interactions promote. The interacting components are soft-assembled in the social assemblage in the course of interaction. The component parts are not defined by rigid metric properties, but by their topological connectivity (DeLanda 2002: 56). Specific shapes of, for instance, bodily movements and gestures, including vocal gestures, matter less than their varying and fluctuating dynamical properties, e.g. the varying intensity of the dynamical properties of bodily movements (loudness, orientation, tempo, duration). Bodily dynamics can continuously vary and adapt such that, for instance, a syllable can be lengthened and adjusted in relation to the coupling of a given action to another action. Voice dynamics and other bodily movements can change to adjust to specific coupling relations between different component processes.

As Reed (1996a: 85) points out, there are various levels of bodily postures and nestings of these as well as controlled transformations from one posture to another (movement). A bodily movement is both a change of state from one posture to another at the same time that it is the maintenance of some postural orientation. Vocalizing is no exception. Body movements, seen in this light, have evolved through interaction with the world. Body movements are cognitively salient. They are oriented towards the intentional transformation of the future state of the organism. Vocal tract gestures, facial expressions, manual-brachial gestures, eye gaze, and other bodily resources are all adaptations of the body that sustain and enact intentional, meaning-guided action. The assuming of a particular postural stance, and the mobilization of the neuroanatomical resources required for its enactment, serve to orient the body and thus to indicate to observers the readiness of the agent for some course of action. Many of these postures are, of course, very rapid and subtle; they are micro-temporal or pico-scale bodily events (Section 4) that are rapidly mobilisable owing to the chaotic dynamics of microscopic body dynamics, thereby ensuring a high level of readiness of the system to mobilize and respond quickly to even slight perturbations (Freeman 2000: 87).

I will now consider some aspects of the cognitive significance of bodies-in-interaction in an analysis of a brief episode of infant-mother interaction.

Micro-temporal Phases in an Episode of Infant-Mother Interaction

Typically, a meaning-making trajectory integrates brain, body and world-side factors in short pulses or intervals that are connected to each other along their trajectory. This can be observed in the following episode involving a one-year-old boy, Luke, and his mother, Sheila.

Luke and Sheila are playing a fetching game involving plastic cubes.[3] Sheila encourages and prompts Luke to go and fetch one of the cubes on his own, without the assistance of his mother. In the episode, we see that verbal, postural, kinesic, prosodic, gaze, and facial movements all interact. Each of these controlled movements is a modification in the postural orientation of the body or some part of it. Each such movement is directed towards the realisation of a value in the situation, e.g. the mother wants her son to fetch the cube, to pay attention to her; Luke wants to please his mother, and so on. The main intervals of this brief slice of a trajectory may be summarised with reference to Table 1 as four micro-temporal phases, as follows:

1. The mother takes the green cube in her hand (00.07.28), adjusts her posture inclining backwards 900.08.52) and leans forwards towards Luke, holding the cube in front of him and engaging his interest and attention (00.10.00);

2. The mother shifts her overall posture and body orientation by leaning forwards towards the designated target of the proposed "fetch" with the cube in her hand (00.10.48): this is the onset of tokened or abbreviated "fetch" that the mother is promoting; Luke's gaze and body orientation shift to track his mother (00.11.00 – 00.11.68);

3. She then briefly re-establishes mutual eye gaze with Luke by turning her head to engage Luke's gaze (00.12.60); she then turns her head in the direction of her outstretched hand, holding the cube (00.13.56); maintaining this posture, her gaze then shifts back to Luke (00.14.16);

4. She then turns her gaze towards Luke while maintaining the forward leaning posture (00.14.60). This movement is followed by a rapid sequence of vocalizations, e.g., "go and fetch it", that are rhythmically synchronised with a sequence of rapid head turns towards Luke, on the one hand, and towards the target of the proposed fetch (the cube), on the other (00.14.60 – 00.16.40). The head turns serve (1) to engage Luke; and (2) to direct Luke to the designated target of the proposed "fetch"; Luke begins to move towards the cube (off screen) while his mother directs her gaze to him while holding the posture (00.16.40); Luke starts to crawl towards the cube, smiling, while his mother shifts her posture backwards to give Luke the space to initiate the fetch on his own (00.18.20).

[3] I thank Stephen Cowley for making the video recording of this episode available to me. The analysis is my own.

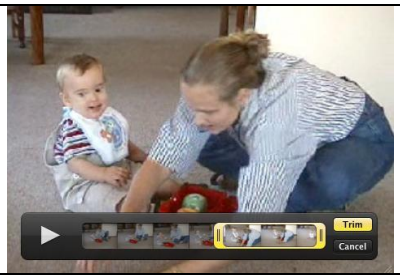
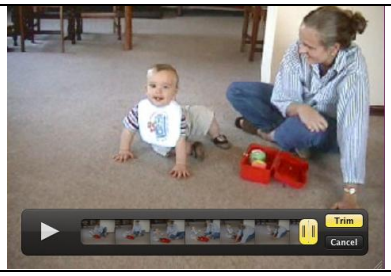
00.07.28	00.08.52	00.10.00
		
Phase 1 →		
00.10.48	00.11.00	00.11.68
		
Phase 2 →		
00.12.60	00.13.56	00.14.16
		
Phase 3 →		
00.14.60	00.16.40	00.18.20
		
Phase 4 →		

Table 1: Four phases in micro-temporal dynamics of tokened rehearsal (fetching game) between mother and infant (original video courtesy Stephen Cowley)

In this brief episode, we see that each of the four micro-temporal phases analysed above forms a pattern of behaviour that is distributed across different bodily postures and their transformations. Moreover, each phase, as well as the episode as a whole, shows the three properties of intentional behaviour identified by Freeman (2000: 116), viz. wholeness, unity, and intent. It makes little sense to split the behaviour into separate channels, codes or modes

of communication that are 'combined' (c.f. Trager 1958; see Thibault 2008 for discussion). Instead, they form a unity from the outset and must be studied as such (Arndt and Janney 1987: 107).

In the present example, we see how the concerted rehearsal that constitutes the fetching game between Luke and Sheila involves the two persons acting on each other, on the relevant objects (the plastic cubes), and on the situation through co-orchestrated and dialogically coordinated real-time bodily dynamics that allow them to generalise and abstract (rehearse) so as to create the meaning structures through which infants like Luke come to perceive and act in the world. Rather than holding an object, a hammer, say, in my hand and exploring it through the haptics and kinesics of tactile exploration, the increasingly compressed and abstracted forms of bodily dynamics enacted in concerted languaging *qua* rehearsal between infants and parents give rise to virtual perception-action cycles that enable infants such as Luke to grasp both in imagination and in action the culturally valued patterns of behaviour that are promoted by caregivers. Engaging in languaging as a form of concerted rehearsal is therefore a means of dynamically shaping and focusing trajectories of value-realizing actions in the human ecology (Hodges 2007, 2009).

Caregivers structure the daily life of infants into meaningful patterns. Meaningful patterns entail participation roles. Caregivers and infants together engage in structured forms of activity involving a differentiation of roles and specializations. Caregivers orchestrate the field of "promoted action" (Reed 1996a: 149) so as to help the child bridge the gap between the child's performance and a competent execution of the skill or action. Melser (2004, 2009) sees language as a mode of concerted rehearsal between individuals. Concerted activities such as languaging entail the acquisition of the requisite perceptual, (inter)actional and cognitive skills for taking part in activities of this kind. Rehearsal implies that the persons who participate in concerted forms of rehearsal such as languaging are readying or preparing themselves for a possible future real performance of the rehearsed action. Rehearsal may also modulate, prompt, guide, direct, orient, correct and organize one's participation in action.

A languaging event consists of a nested set of time-scales, or an array of time-scales that are manifested as different kinds of oscillatory behaviour on different temporal scales. Moreover, in languaging events of this kind different cycles of behaviour on different time-scales are not simply nested the one within the other, but may overlap. There is then the potentiality for radical scalar heterogeneity of time-scales. Biological oscillators may intersect with or overlap with social and cultural oscillators. The speech event unfolds pulse by pulse. Its metric structure is explainable in terms of the relations between oscillators on different time-scales. Rhythmic oscillators have the functional capacity to entrain to each other's temporal-rhythmic behaviour. Thus, two or more persons in conversation have the capacity to form an assemblage – e.g. an occasion of social interaction – by entraining to or synchronising with each other's speech rhythms (Section 17). They can also synchronize to rhythmical oscillations on large scales, both biological (e.g. the night-day cycle or the twenty-four hour rotational cycle of the Earth) and social (e.g. work cycles, classroom cycles, institutional cycles, the collective language dynamics of a population of speakers). Lefebvre (2004 [1992])

uses the term ‘dressage’ to explain how rhythmic patterns are established and integrated over biological, social, and historical timescales. Persons in socially organized assemblages are entrained to and are responsive to these rhythms.

Language is grounded in biomechanics, but is constrained by cultural processes, social conventions and historical traditions. The intrinsic interactivity and expressivity of the biomechanics set the parameters of the capacities of the interacting components of a conversation to couple their neural and bodily dynamics with each other in relations of co-orientation and co-affiliation that also organize the spatial extent of the interaction, its degree of intensity, and so on. Cultural conventions and historical traditions, including lexicogrammatical patterns and social norms, constitute a second parameter. This parameter stabilizes the identity of the assemblage as a whole. For example, ‘grammaticalized’ patterns act as second-order constraints that establish procedures and routines for interaction, the degree or level of formality of the interaction, normatively appropriate ways of differentiating the environment, and interactional obligations, responsibilities, and expectations. Some conversations are highly constrained in these ways, others much less so. Moreover, a conversation does not come into being out of nowhere. The component parts that form a conversation on a particular spatio-temporal scale always come into being when a whole has already come into being and uses its component parts and their capacities to constrain the interactions between them. Thus, persons have intrinsic biomechanical properties and capacities for interactivity and expressivity that can be exercised by these properties. These properties are bottom-up or micro-level contributions whereby individuals actively contribute to and modulate the formation of the overall assemblage. Conversations arise in communities, interpersonal networks, and institutions with their traditions and norms that pre-exist particular individuals. Historical traditions and cultural norms constrain and enable the possibilities for interaction of the component parts of the assemblage. They pre-exist individuals in the sense that they derive from cultural-historical timescales at the same time that they speak through and are immanent in the interactions between the components of the assemblage.

On this basis, the distributed language view has developed the distinction between *first-order languaging* and *second-order language* (Love 1990; Cowley 2007; Thibault 2008, 2011). Verbal patterns (words, wordings, grammaticalized semantic structures, etc.) are second-order cultural patterns that constrain the bodily dynamics of first-order interactivity between persons and between persons and aspects of their environments. A growing body of evidence shows that interactivity, not abstract symbol manipulation, content transmission, or information processing centred on the internal mental processes of the individual, is the key to human learning, cognition and intelligence. Text-based literacies have privileged pedagogies that abstract away from this basic fact because of their focus on a phenomenology of abstract forms and centralized social or cognitive codes. As we shall see below, it is interactivity that enables persons, through their coordinated languaging behaviour, to form the social-affective-cognitive assemblages whereby language operates on and catalyzes socially organized flows of matter and energy.

In the following Section, I define the distinction between *first-order languaging* and *second-order language*.

4. First-order Languaging and Second-order Language

My aim in this section is to offer some preliminary definitions of two terms that will be used to develop an alternative approach to the problems posed by the kind of approach that the early work of Trager exemplifies. First, the term *first-order languaging* refers to the co-constructed and dialogically coordinated body dynamics that occur in real-time interactive behaviour between persons. First-order languaging can therefore be distinguished from *second-order language*, i.e. verbal patterns such as lexicogrammar and discourse, which are stabilized patterns on longer, slower cultural time scales. Agents learn to exploit the dynamics of first-order languaging in ways that get linked to cultural norms and values that connect selves to patterns and traditions emanating from cultural-historical time-scales. The term 'languaging' was first used by Maturana (1970; see also Kravchenko 2003) to emphasise the biologically grounded and processual character of talk as a complex organization in time of biological and social processes oriented to the creation and maintenance of consensual domains due to the structural coupling relations enacted between languaging agents.

First-order Languaging

First-order languaging is embodied inter-individual activity that takes place in real time. It is grounded in body dynamics on fast time-scales measurable in milliseconds to fractions of seconds of fine-grained and co-orchestrated bodily activity between persons (Cowley 2008; Steffensen, Thibault and Cowley 2010; Thibault 2008, In Press). The researchers cited here have developed, in recent years, the notion of *pico-scale bodily dynamics* in order to theorize and to develop methodologies for the study of this scale of languaging behaviour. It is therefore more appropriate to focus on and develop explanations that are grounded in the fast time scales of these pico-scale bodily dynamics that are co-orchestrated and co-synchronized by persons in talk.

It has generally been assumed, for instance, that speech sounds imitate, encode, or express a cognitive-semantic operation or meaning that exists independently of, or is external to, the physical (bodily) activity on some other (more abstract) level of organization or coding. This pervasive view is frequently applied to both the articulatory and prosodic aspects of vocalizations. Instead, the cognitive-semantic operation or process derives **directly** from vocal tract or other intentionally modulated gestural activity. It is the patterns of neuromuscular activation involved in the articulation of a gesture – the neuromuscular program – that directly interact with and operate on the neuro-motor capacities of the addressee. In this way, cognitive-semantic operations are operated in languaging agents. For instance, in English and many other languages, rising tone is an expressive feature of utterances that articulates an anticipatory dynamic (Rosen 1985), i.e. something in the utterance or associated situation is incomplete or unresolved. The rising tone anticipates its completion either later in the same utterance by the same speaker, or by the addressee in his or her response. Rising tone does not imitate, encode, or express the meaning that is glossable

as "incomplete + anticipation of completion" or "unresolved + anticipation of resolution". It is not the arbitrary or conventional coding of such meanings.

Instead, it is an aspect of a dynamically, prosodically shaped vocal tract gestural activity in which the agent assumes an active, dynamical "postural-affective" (Werner and Kaplan 1984 [1963]: 24) state towards some environmental configuration or event that brings about a change in the agent's internal dynamics and organizes a response to it. Viewed in this light, first-order bodily dynamics are intrinsically expressive and cognitive, rather than the arbitrary coding of meanings that are external to them. The latter view is based on the idea that a static sequence of abstract phonemes, e.g. /kæt/, arbitrarily encodes the abstract meaning, "CAT", in the spoken word *cat*. This view arises when utterances are seen as disembedded from the "organismic matrix" (Werner and Kaplan 1984 [1963]: 25) which constitutes their motivation and in terms of which utterances are organized in response to changing environmental events to which the agent needs to adjust. As we shall see in Section 13 below, these observations of Werner and Kaplan tie in with the idea that the internal and external dynamics of the organism and its environment comprise a unified organism-environment system that is the foundation of languaging behaviour as a form of culturally constrained catalytical activity that is grounded in human biology.

The neuromuscular patterns of activation that articulate the rising tone directly enact a cognitive routine; they are not the expression *of* an external meaning that exists on some other level of linguistic coding. These patterns directly enact the cognitive routine in the sense that they bring about a corresponding change in one's interlocutor and therefore in some aspect of the relevant environment. The detection of the pattern by the interlocutor (addressee) may, in turn, result in further changes in the addresser's relationship with the environment as he or she adjusts to the change detected in the addressee. Body dynamics are therefore directly cognitive; they are not the 'expression of' something else in the form of abstract cognitive-semantic processes that pre-exist them. This observation is especially important in the context of first-order languaging because it puts the spotlight on the ways in which agents-in-interaction directly act on and modulate each others' cognitive operations and perceptions in and through their co-constructed micro-temporal bodily dynamics. These bio-physical dynamics are grounded in the natural expressivity of our bodies, i.e. the bio-physical patterns expressed by the body. Moreover, they are organized patterns that compress information and give agents the capacity to interact with other agents in ways that can individuate the self and the other. In second order language, the cultural processes of conventionalization are much more advanced to the extent that this primary biological motivation for the cognitive significance of speech sounds is bleached out though it is not entirely absent.

In actual fact, it is more accurate to say that the rising tone is a behavioural unit, or an aspect of one, in the sense defined by Reed (1996a: 30), i.e. a unit of action on some time-scale that "results in a change in the animal's relationship with its environment" (1996: 30). The 'grammaticalization' of behavioural units, i.e. constellations of phonetic gestures making up 'words' and larger units, is really a preservation or a persistence of pattern that is explainable in terms of the population-level dynamics of higher-order articulatory, acoustic and other

invariants in a community of speakers. These higher-order invariants can function as culturally constrained values owing to the selection pressures of a population of languaging agents. Populations of phonetic gestures habitually relate to populations of environmental affordances on the basis of a history of covariance relations that give rise to statistical effects in a population.

First-order language is time-locked inter-individual activity. It is a bio-cultural process that integrates many different time-scales of organization within and beyond the body and brain of the individual. It is based on co-constructed neural and bodily dynamics that are spread between people in complex processes of circular causation. These dynamics have the potential to connect persons across multiple time and space scales. Moreover, they have the power to coordinate experience around verbal patterns, bodily processes and aspects of particular situations. Furthermore, modulations of body dynamics can bias perception in ways that partition the world into meaningful categorizations. Body dynamics compress and synthesize semantic information that can modify the experience and perspectives of agents-in-interaction. This suggests that hominids that developed this capacity lived in a cultural ecology that allowed them to modify each other's perceptions, cognitions, feelings, actions, identities and values through co-orchestrated body dynamics in micro-time.

As we shall see below, this capacity is linked to languaging behaviour as a form of organism-environment catalysis grounded in experience. The co-synchronization in micro-time of the bodily behaviours of individuals in a particular population transforms perspectives, values and experiences. This capacity depends on the subtle and fine-grained coordination of body dynamics and features of the world that also exploit affect and cultural patterns. The term *languaging* might be a better term for helping us to escape from the limitations and reifying tendencies of the noun 'language'. We need new ways of thinking and new tools of thinking if we are to develop new ideas about 'language'.

Second-order Language

We can refer to verbal patterns – i.e. lexicogrammar -- as *second-order language* (Thibault 2011; see Love 1990 for the original use of this term). Second-order language is what most people have learned to recognize and talk about as 'language' on the basis of a personal and cultural history of metalinguistic practices grounded in literacy and the phenomenology of language form which literacy practices promote. Language is abstracted from bodily behaviour and accordingly treated as a separate and distinctive modality of communication with its own characteristics, sometimes called "design features" (Hockett 1960). When language is theorized in this way, as the example of Trager showed in Section 2, it is separated from the total continuum of communication that occurs in conversation (Harris 1990: 22) and turned into an arbitrary and abstract symbolic code.

Traditionally, language has been viewed as the processing of abstract symbolic forms. Reified synchronic forms are the 'encoders' or repositories *of* meanings, seen as separate from body dynamics. Body dynamics are separated from cognition and first-person experience. Synchronic abstractions replace change and dynamics. This perspective fails to show how

second-order cultural patterns (wordings) get integrated, in real time, to first-order languaging dynamics, not encoded by them. Moreover, first-order languaging dynamics modulate wordings just as the latter constrain the dynamics. First-order language dynamics makes use of audible and visible bodily behaviours in real-time. Our brains tune into and exploit these dynamics in ways that connect experience to normative patterns and values. Body dynamics, not formal abstracta, have the functional capacity to transform perception and action. Co-constructed body dynamics (first-order languaging) have causal and cognitive powers. Interactants are influenced by and respond to these dynamics in micro-time. They exploit pitch, rhythm, cadence, tempo, duration, loudness and other aspects of vocalizations as well as, for example, facial and gestural activities.

5. Typological Thinking vs. Population Thinking: Saussure's Error

Languaging behaviour cannot be reduced to the verbal patterns that linguists transcribe and theorize as 'language' by abstracting verbal patterns from real-time bodily dynamics. On this view, the 'object of study' is a reified abstraction, a hypostatization that has no naturalistic grounding in human biology. This view treats verbal patterns and their purported meanings as abstracta that are separated from the neural, affective, and bodily dynamics of real-time languaging behaviour. Many of the phenomena that are placed under the rubric of paralanguage, kinesics, and so on, are aspects of the very same time-locked dynamics that constitute languaging activity in living, interacting, intentionally charged human agents with bodies and brains. This conception is very different from one that views abstract language forms as the inert and lifeless outcomes of an input/output machine on analogy with the digital computer. Language is grounded in the neurobiology of our brains, our physiology, in perception, in action, and in body morphology. The reduction of language to code-like input/output systems and their regularities denies this grounding. Consequently, language is treated as something that cannot be naturalistically grounded in human biology and first-person experience. It is seen as disembodied and arbitrary cognitive or social processes based on computational models of abstract symbol processing. Cowley (2008) points out that we hear, feel and understand what we say in ways that ground language and interaction in the phenomenology of first-person experience (see also Reed 1996b). On the other hand, if we separate language from biology and first-order experience, we fail to see how languaging is integrated with our feelings and our bodies in productive ways.

In the recent history of linguistics, this tendency can be traced back to the methodological decision made by Saussure when he set up the then nascent science of general linguistics in the early decades of the twentieth century. Saussure's positing of a collective abstraction, viz. *la langue*, as the object of linguistic sciences, has helped to stymie a proper appreciation of the radically heterogeneous character of language (c.f. Saussure's exclusion of *le langage*). In insisting on a homogenous abstraction as the object of linguistic science, Saussure privileged essentialist and typological modes of thinking traceable back to Aristotle that have remained stubbornly ingrained in linguistics to the present day. Thus, on analogy with the exception-less laws of classical (Newtonian) mechanics, *la langue* was seen to be the repository of everything which the linguist held to be "essential and universal" ["essentiel et universel"]

while at the same time putting aside other factors deemed to be "the particular to the accidental" ["le particulier de l'accidental"] (Saussure 1993 [1910-1911]: 192). On this view, *la langue* is a "collection of abstractions" ["un ensemble d'abstractions"] (Saussure 1993 [1910-1911]: 192) from which everything "essential and universal" about language(s) can be logically deduced as constant formal regularities.

In shifting the focus from specific historical languages (*les langues*) to the theoretical abstraction of *la langue*, Saussure viewed *la langue* in terms of abstract forms and general categories rather than as historical individuals or singularities on some spatio-temporal scale. Consequently, the material dynamics and historical processes which have lead to the morphogenesis of these individuals are set aside. Causal explanation based on material dynamics was replaced by the description of formal regularities, most notably the union of the two mental abstracta – acoustic image and concept – in the formation of the linguistic sign (Saussure 1993 [1910-1911]: 190). Moreover, *la langue* is viewed as being logically necessary for the production of this union (Saussure 1993 [1910-1911]: 190) even though the specific causal mechanisms responsible for this union elude such an account. In adopting this view, Saussure effectively eschewed productive causes in favour of a theory (and its object) founded on formal regularities and abstracta. It was on this basis that Saussure formulated a small number of general principles or axioms, notably the *linearity principle* and the *principle of arbitrariness*, on the basis of which a large number of consequences are deduced. Saussure's rejection of a naturalistically grounded mode of enquiry therefore paved the way for a theory of language based on essences, rather than material dynamics and productive causal relations.

Saussure's theoretical object, *la langue*, is a collective synchronic abstraction. Saussure's essentialist and typological view can be contrasted with the population thinking that was first developed by Darwin. As the brief observations above show, Saussure's way of constructing the theoretical object, *la langue*, makes thinking about variation and change difficult or problematic. Variation is thus seen by Saussure (1993 [1910-1911]: 192) as "the particular and the accidental" and therefore is a departure from the typological criteria established by the abstract and homogenous theoretical object, *la langue*.

Population thinking, on the other hand, views formal abstracta and types as no more than theoretical and heuristic tools rather than hypostatized ontological levels of language. Populations constitute the basis for groupings of individuals. Causal relations between individuals and individual variants are a central part of the explanation of a population and its dynamics. Populations are not abstract totalities, but concrete historical individuals on some spatio-temporal scale. They have extension in space and time. A language (e.g. Catalan, Italian, Putonghua, Urdu) is, then, a concrete historical individual on its spatio-temporal scale just as an individual person is on a very different scale.

On this view, a historical language is an individual on its particular spatio-temporal scale. It is not an abstract system that is instantiated as specific instances that conform to varying degrees to the typological criteria established by the formal regularities and schematic properties that are seen to comprise or define the essence of the language system. Instead, the relationship

between a language and the smaller scale individuals that comprise it is one of parts to whole. Thus, a language qua historical individual on its far greater spatio-temporal scale is made up of the dynamical interactions between the smaller scale individual component processes (e.g. languaging events) that are the component parts that comprise the larger-scale entity. The critical difference for our present purposes between the two views – typological and population thinking -- resides in the fact that a population is not defined in terms of structural arrangements of formal abstracta, but in terms of its material dynamics and the causal processes that change the system over time. A system based on formal abstracta is unable to achieve this kind of causal explanation. Moreover, the population view takes it as entirely unproblematic that a given population may be characterized by varying degrees of heterogeneity or homogeneity without feeling the need to postulate an abstract totality, viz. Saussure's *la langue*, as the object of study.

In population thinking, the relations between individuals on different scales are one of parts to wholes, rather than a type-token or a system-instance relation between different ontological levels, e.g. the system level and the instance level. It is the causal interactions among the individuals – seen as historical singularities – on any particular spatio-temporal scale (not ontological level) that leads to the emergence of the next larger-scale individual. A population on any given scale is an assemblage of co-articulated material processes that has a given spatial extension and that exists over a given time-span. It is created and maintained by its material dynamics and the productive differences these give rise to, rather than the purely abstract and negative conception of difference on which Saussure based his conception of linguistic *valeur*. The fact that a language consists of many sublanguages and sub-varieties of various kinds is itself a contingent consequence of population processes in the socio-cognitive-interactional dynamics of, say, the members of a particular interpersonal network. For present purposes, the important point is that population thinking provides a better theoretical explanation of language at whatever scale in terms of time-locked processes and material dynamics on different scales rather than in terms of synchronic states, formal abstracta, and outputs of formal systems.

'Grammaticalization' is sometimes seen as a matter of compositionality, i.e. the composition of abstract grammatical forms from smaller component parts such as words. In actual fact it is, in the first instance, the preservation and persistence at the population level of very high-order physical invariants in a repertoire of phonetic gestures. These invariants afford access to a world of virtual cultural entities once one has learned the cultural tricks for accessing them. The phenomenon of grammaticalization refers to the automatization of regular, repeated sensori-motor routines in a population in order to reduce the cognitive and sensori-motor costs of linguistic interaction (Bybee 1998; Bybee and Thompson 2000; Argyropoulos 2008). These invariants are a form of higher-order perceptual information that can catalyze in self and others apperceptions of virtual objects and events that need not be physically present in situations or accessible to on-line perception. Once such virtual entities became established in a population as culturally salient and valuable entities that matter to agents, so too did meta-linguistic models and procedures for re-categorizing in socially distributed ways the phonetic gestures as populations of lexicogrammatical forms that are analysable as belonging to

different word classes and their possibilities of recursive recombination. The two developments would have emerged in parallel, viz. in a nutshell, reflexivity and recursivity, the coupling of which induces its own catalyzing effects. On this view, the re-categorization of a repertoire of phonetic gestures as lexicogrammatical categories in a given population of agents is a form of cultural scaffolding. Specifically, it is a cultural scaffolding of the growing diversity of interaction routines and associated situation-conventions that the phonetic gestures afford agents as more and more pressures for standardizing the coordination of agents builds up.

The extraction of prototypes on the basis of the covariance relations between environmental events and the expanding repertoire of recursive phonetic gestures give rise to statistical properties at the population level. These properties arise from the perceptual sampling and exploration of phonetic gestures in a given population. They enable individual agents on particular occasions to determine or interpret the intended meaning of any given gesture through repeated, habitual participation in and observation of the interaction routines and behaviours in which these covariance relations arise. This development would in all probability give rise to selection pressures for agents to generalize from their experience of the situations in which the gestures habitually occur. Through informal and formal learning processes, agents pool these generalizations as part of collective narrative wisdom, again at the population level, such that the accumulated wisdom is potentially accessible by all members of the population. In turn, the pressure to generalize would bring about further selection pressures to segment the continuously varying character of the gestures into component parts that are associated with or can be associated with differential aspects of situations.

Grammaticalization arises in a population as these associations themselves are subjected to more and more layers of cultural constraints as the process of segmentation spreads, owing to its augmented capacity for generalization and the enhanced cognitive and semiotic reach that this capacity affords. The spread of these processes catalyzes the rise of digital semantics and, hence, the expanding pluriverse of virtual cultural entities and events that enable humans to keep track of self and others over an expanding diversity of space and time scales and niches in the extended human ecology. The output of these combined processes catalyzes particular effects. In turn, these effects become the input to new processes that give rise to cycles of self-organizing iterative linguistic structures. These structures exhibit more and more properties of global coherence and identity that persist on the long historical-cultural timescales of a historical language. The units of selection – prototype extraction, generalization, segmentation, association of phonetic gestures with covarying environmental affordances -- are positive feedback processes giving rise to an autocatalytic web of socio-cognitive interaction strategies and behaviours that are perpetuated on cultural timescales as second-order lexicogrammatical and discourse-level and other constraints on these same behaviours.

6. The Scalar Heterogeneity of Language

It is commonplace to say that language consists of different levels of organization. Generally speaking, these levels are seen as inter-related aspects of the formal architecture of a language that can be localized in code-like mechanisms. Such mechanisms are postulated as existing inside the brain of the individual or in externalized artefacts such as 'text' (Section 1). From the perspective of population thinking, language is a radically heterogeneous phenomenon that is causally spread across a diversity of different spatio-temporal scales (Lemke 2000a, 2000b; Thibault 2004a, 2004b; Wheeler 2005), rather than being isolable on any one of them or homogenized as inter-related levels of abstract forms. A focus on any particular scale may be an entirely legitimate line of enquiry for certain purposes, but it cannot be the whole of language. Language and the explanation of language cannot be confined to any one spatio-temporal scale.

According to population thinking, smaller scale entities are the parts whose interactions comprise the next higher-scalar entity. The classic example in biology goes somewhat (and approximately) like this: genes are parts of chromosomes; chromosomes are parts of cells; cells are parts of specialized organs; organs are parts of organisms; organisms are parts of groups; groups are parts of subpopulations; subpopulations parts of species, and so on. On each of these scales, the entities that comprise that scale have the capacity to reproduce. Moreover, there is variation and this variation can affect reproduction and rate of reproduction of the entities that exist on any given scale. For these and other reasons, the different scales proposed in this hierarchy show Darwinian characteristics. This does not mean, incidentally, that all population dynamics are Darwinian ones. Many aspects of social and cultural dynamics entail reproduction though not necessarily through Darwinian processes involving parent-offspring lineage in the biological sense.

As for language, we can postulate that language is spread over a number of scales, populated by their respective entities. The scales across which language is spread can be approximately and by no means exhaustively outlined as follows:

i. Neurons are the component parts of populations of neurons (Edelman 1987, 1989). The central nervous system is a system of oscillators and the modulatory relationships between oscillators. Neurons and neuron circuits are oscillatory and are modulated by influences from other neurons and neuron circuits in distributed networks of brain activity. The patterns of interactions of neurons with other neurons form stable patterns of oscillatory networks of neuronal activation leading to the emergence of populations of neural networks. The content of these networks is a prototype that is created by sampling the objective statistical properties of a population - properties that can be transformed into a subjective felt experience of a linguistic category;
ii. Pico-scale bodily dynamics measurable in milliseconds to fractions of seconds are stable yet varying patterns of bodily behaviour that form populations of dynamical bodily events in the observable languaging behaviour of individual persons and groups of persons;

iii. Populations of phonetic gestures are formed from the entrainment of lower-scalar pico-scale bodily dynamics to higher-scalar patterns of vocal tract articulatory and auditory gestural activity. These stable patterns are based on rich phonetic memory that is grounded in first-person experience of phonetic events (Browman and Goldstein 1992, 1995; Fowler 1986, 2010; Port 2007, 2008), not abstract phonological schema;

iv. Inter-individual interaction routines based on the co-synchronization and entraining of an individual person's repertoire of vocal tract behaviours to group level dynamics and strategies requiring the social coordination of persons gives rise to inter-individual patterns that are not reducible to the component individual persons on the scale below;

v. Language practices of particular social groups associated with particular situation-types and the conventions for coordinating persons in situations;

vi. Lexicogrammatical patterns are the statistical results of cultural processes of institutionalized teaching, standardization, and transmission across generations (Halliday 1991) that emerge through populations of language practices. In this way, a given interpersonal network of interacting agents preserves (and values) higher-order physical invariants of populations of phonetic gestures and associated interaction routines. Phonetic gestures are, accordingly, re-categorized as lexicogrammatical patterns, i.e. normative patterns that stabilize the identity and the socio-cognitive dynamics of an entire interpersonal network or community of interacting agents.

Unlike the relation between a higher-order system and its instantiations, the relationship of smaller scale entities to larger scale ones is one of parts to whole. For example, populations of phonetic gestures are the parts that comprise populations of interaction routines that coordinate persons. Moreover, whereas the system-instance instantiation relation is based on degree of conformity of the instance to the more schematic criteria established by the higher-order system category, the relation of parts to whole in a population is causal. Wholes emerge through the interactions among component parts as a consequence of the specific causal mechanisms at work on a given scale and the often complex interactions over time across scales. For example, interaction routines emerge through the ways in which phonetic gestures entrain to and synchronize with higher-order inter-individual patterns that, in time, characterise particular social groups.

Saussure distinguished the social and individual dimensions of language in terms of *la langue* and *la parole*. These two terms were derived from *le langage*, which Saussure put out the door of linguistic theory on account of its unmanageable heteroclite character. Saussure's reasons were as follows:

> Le langage est un terrain complexe, multiforme, héteroclite dans ses différents aspects. Un consequence, c'est qu'on n'arrive pas à la classer pris dans son tout avec d'autres faits humains. Il est à cheval sur des domains divers (domaine physique, psychique, ou encore: domaine individual, social.) (On ne sait comment lui conférer l'unité.)
>
> (Saussure 1993 [1910-1911]: 276)

On the distributed view, language just is a radically heterogeneous phenomenon that is spread across a diversity of timescales, as outlined above. Saussure's *le langage* can serve as a new point of departure such that real-time dialogically coordinated languaging events between persons are not the outputs or instantiations of a homogeneous and abstract *la langue*, but the emergent outcome of the interactions between component processes and their capacities deriving from a diversity of timescales. A languaging event is an assemblage in which populations of interacting component processes and their capacities on the scale of that event are constrained and enabled by populations of entities and their capacities on other scales that extend from the neural to the cultural-historical, as shown above, and are integrated to that event. The component processes on any given scale are individuals -- populations of individuals, to be more exact, i.e. populations of neurons, of bodily dynamics, of words, of persons, interpersonal networks, and so on. As I said before, populations of component processes on smaller scales (e.g. patterns of neural activation, intentions and body dynamics) are sub-personal component processes of the individual person; they have the capacity to modulate the behaviour of that person just as the coordinated behaviour of the individual person modulates the larger-scale interaction-system between persons that is a conversation. On larger scales, populations of cultural-historical patterns that have arisen in a community in part as the unintended statistical consequences of the very many interactions between persons entrain the dynamics of the lower scales to their own dynamical patterns such that the latter constrain and enable lower scalar component processes.

Grammaticalized patterns therefore are normatively constrained procedures that embody in the dynamics of a population the successful and valued patterns and routines of the past (Section 28). The capacity of grammaticalized patterns for increased abstraction means that they can function as cultural scripts that compress and synthesise cultural-historical information. They are performance and dispositional scripts that have the functional capacity to prompt us as to how to act, orient, attend, perceive, feel, and think, and so on, in specific circumstances. When activated, they get coupled to bodily dynamics, which they shape and entrain in culturally specific ways that lead persons to act in appropriate (and inappropriate) ways. Grammatical scripts can therefore serve to prompt and guide behaviours, though one needs to have access to the culturally compressed information in order to ensure that grammatical patterns are appropriately coupled to bodily behaviours. This is clear in most language pedagogy, which teaches disembodied grammar scripts as verbal forms on the basis of a misleading code view of linguistic 'communication' without attending adequately to their implications for bodily performance in interactional encounters with others and the social-affective-cognitive assemblages that these give rise to.

Another big difference with typological thinking is that in populations of whatever scale the entities on any given scale are historical individuals or singularities that operate at a particular spatio-temporal scale and which interact with entities on other scales. Thus, an historical language, like a biological species, is a historical singularity that exists on a particular spatio-temporal scale just as a single neuron is also an individual existing on a very different spatio-temporal scale. On the other hand, the relationship between a higher-order language system and specific instances is based on the Aristotelian assumption that this relationship is a

hierarchical one. In a relationship of this kind, the entities on different levels of generality are postulated to exist on different ontological levels and therefore to correspond to different ontological categories.

Population thinking emphasises the unique and the singular rather than the typical and the general. Individual neuronal networks are unique just as individual phonetic gestures and individual languages are unique on their respective scales. The collective interactions among the entities on a particular scale, as Mayr (1976) pointed out, can only be described in statistical terms. Lexicogrammatical patterns are, in Mayr's terms, statistical abstractions deriving from the heterogeneity (not homogeneity!) of the interactions among the individuals that comprise the population. However, the collective character of these interactions means that some patterns are selected, taught, and transmitted such that they propagate throughout a population from centres of teaching, power and influence and are reproduced or replicated as normative patterns to which individual behaviours are entrained and to which they are expected to conform for the good of the collective. This does not mean that the statistically unlikely state of homogeneity prevails, unlike Saussure's idealized and homogenized conception of a synchronic *langue*.

Lexicogrammar is a mode of normative empatternment that evolves within the socio-cultural dynamics of collectivities of persons and their institutions. However, the spatio-temporal extension of a particular language qua historical singularity (not idealized system, norm or type) means that heterogeneity and variation are the likely normal state of affairs. In the sense that lexicogrammatical forms are the statistical consequences of collective and socially distributed interactions between persons, they are virtual entities that individual persons learn to perceive and orient to as actualized in the own and others' linguistic behaviour as specific wordings. This achievement takes place through participation in meta-linguistic practices of learning to attend to aspects of the dynamics of vocal behaviour qua higher-order articulatory, acoustic and other invariants that are entrained to and shaped by cultural attractors. In the first instance, these practices begin, most fundamentally, in the parent-infant dyads that are created in the first year of the infant's life (Halliday 1975). We learn to talk about talk in ways that shape and (self-)regulate talk around norms that the socio-cultural milieu requires us, either informally (e.g. a sense of solidarity or belonging to a group and having an identity in that group) or formally (e.g. pedagogical enforcement of language standards in schooling), to conform to for the collective good. Moreover, conformity brings its own emotional rewards at the same time that nonconformity is punished, e.g., group solidarity brings a sense of belonging and therefore a sense of group identity; non-conformity can lead to forms of social stigmatization of non-standard varieties and their speakers.

The vertical coupling of scales is central here. For example, an individual person's brain consists of populations of neural networks and the interactions between these that occur on a smaller spatiotemporal scale with respect to the individual. Neural dynamics are embodied in a larger-scale entity, viz. the central nervous system and body of the individual organism. Second-order cultural patterns such as lexicogrammatical differentiators are control parameters that index emergent, self-perpetuating virtual relations in a population of

interacting agents. Second-order patterns of this kind persist on the long, slow historical-cultural timescales of a population of languaging agents. Our ability to adopt the 'language stance' (Cowley 2011) on the dynamics of languaging behaviour and to perceive lexicogrammatical patterns (wordings) in real-time languaging events means that these patterns are also seen and heard as intentional semantic contents on the short timescales of real-time interaction. Intentional semantic contents are part of the endlessly evolving hierarchies of control structures (Juarrero 1999; Van Orden and Holden 2004; Thibault 2004b. They emerge and are perpetuated in time by means of circular causality as systems of symbolic associations (Deacon 1996) on cultural timescales.

Living systems originate in the circular, positive feedback processes of chemical autocatalysis. In autocatalysis, the output of a chemical process becomes in its turn its input and catalyzes the same process. Autocatalysis thus perpetuates the reaction in continuous cycles of chemical reaction. The chemical reaction itself appears, on its time-scale, as a coherent, self-organized, iterative structure. Because lexicogrammar patterns are perpetuated in time as webs of symbolic associations by autocatalytic processes, they have the functional capacity to constrain processes on smaller, shorter timescales of real-time interaction between individual persons. This helps to explain lexicogrammatical patterns as both normative constraints emanating from cultural timescales, on the one hand, and intentional semantic contents that are attributed to individual persons in interaction as aspects of what we perceive in the dynamics of first-order languaging behaviour, on the other. The developmental and individuating trajectories of persons vertically couple more and more scales or organization and process.

The vertical coupling of different timescales on account of biofeedback processes makes this possible. Vertical coupling means, for example, that neural events on their time-scale are not seen as random fluctuations, as seen, say, from the viewpoint of the cultural time scale. Instead, they are coupled to events on longer, slower cultural time-scales (amongst others). Bio-feedback processes in complex bio-cultural systems such as in languaging behaviour are vertically coupled on many different timescales, e.g. those stipulated above, taking in the neural to the cultural. As Van Orden and Holden (2002: 94) point out, this leads to the emergence of "fractal patterns of long range correlation", e.g. between the neural and cultural time-scales, or between intentional semantic contents of utterances of individual persons and the lexicogrammatical constraints emanating from the cultural scale.

Linguistic utterances are operators on implicit underlying networks of representations rather than encodings in propositional form of representations of states of affairs (Harris 1991; Bickhard 2004; Thibault 2005a, 2005b, 2011). In individual brains, these representations are stable patterns of neural activation leading to neuronal networks the content of which is created by the individual agent's sampling of the objective statistical properties of a population of, e.g., observed covariances between parts of utterances (e.g. words) and aspects of situations. In a particular situation, both standing constraints and contingencies will conspire to activate and transform this stored prototype into an intersubjective experience when two brains come to resonate with each other owing to the co-synchronization of each

person's neural and bodily dynamics in their dialogically coordinated languaging behaviour. Because representations are also conventionally and normatively associated with particular situation-types and their conventions, which utterances operate on and transform, the intention that modulates the utterance and flows through its trajectory, can be inferred or interpreted due to the fact that utterances take place in relation to situations which are normatively constrained. If the utterance, "Can I help you?", uttered in a shop by a salesperson, has acquired the semantic-pragmatic status of a Sales Bid, it is because it takes place in a situation and in relation to its implicit underlying representations. This means that agents are able to correctly interpret the customary intention that guides the speaker's uttering of this vocalization as a bid to initiate a sale with a potential customer. The intention becomes customary and therefore easy to interpret owing to the intensity of the significance that persons attribute to the situation and the frequency and density of the agent-situation interactions in relation to which the intention is interpreted in a given population of interacting agents.

7. Abstract Symbolic Forms and/or Material Bodily Dynamics

Languaging behaviour is forever under construction. Every utterance that occurs is a historical singularity. Many current approaches to language processing assume the opposite: they assume that the same conditions yield approximately the same output, or that the same choices are instantiated from the same abstract system, consisting of a priori schema and component processes. Input/output processing models of language are based on the idea (in various versions) that language consists of a set of inter-related levels of abstract form – e.g., phonology/graphology, morphosyntax/lexicogrammar, semantics -- mediated by coding relations. Such approaches assume that language has its basis in smaller component atoms that combine through rules to create language output.

On this view, high-level abstracta such as phonology and morphosyntax mediate and filter out irrelevant small-scale detail on any given level of the coding cycle so that only the abstract formal categories of the system are considered relevant to its internal architecture and functionality. Small-scale bodily dynamics and fine-grained details of the situation do not enter into the description of linguistic inputs and outputs assumed by the views described above. They are not therefore seen as relevant to or as affecting languaging behaviour. How do such approaches explain context-dependent languaging behaviour? One way is to assume that stable linguistic components combine in situationally specific ways according to stable social or generic conventions and expectations that skew the outputs of the system. The outputs of the system in this approach are described as 'text' or 'discourse' in many recent accounts. Text and discourse are seen as the manifestation of formal regularities and patterns above the level of the sentence, which previously was the main concern of classical linguistics. There are two critical problems with this approach.

First, the preoccupation of linguistics with formal abstracta is preserved though it is now shifted or extended to higher levels of linguistic organization 'above' or 'beyond' the level of the sentence. Secondly, the components of linguistic behaviour are never found to exist in context-independent ways that allow us to define any context-independent criterion for the

establishment of stable forms that are reliably the same from one situation to the next. The attempt to understand languaging behaviour on the basis of stable, abstract forms and the regularities of their combinations has resulted in languaging activity being removed from and studied independently of its biological implementation as inter-individual bodily dynamics in a population. Albeit with varying emphases, this is no less so of many current cognitive and functionalist approaches to language and discourse. These theories are based on a 'levels' account of language, seen as comprising a number of inter-related levels of abstract form and the coding relations between the different levels.

For example, Halliday (2004 [1985]: 24-26) proposes that the content stratum of language is internally stratified into two layers: an 'interfacing' stratum (semantics) and an 'organizing' stratum (lexicogrammar). Thus, the semantic stratum interfaces "with what goes on outside language: with the happenings and conditions of the world, and with the social processes we engage in." (2004 [1985]: 24). Halliday argues that this involves two steps, as follows: First, "experience and interpersonal relationships" in the ecosocial environment are transformed into meaning. This is the semantic stratum. The semantic stratum interfaces with the world outside language and transforms it into semantic content. The semantic categories of language construe the ecosocial environment in which language is embedded. Meaning, in a further step, is transformed into wordings (lexicogrammar) on the next level down. On this view, lexicogrammar is the stratum on which meanings are organized as wordings. Cognizers and perceivers are not explicitly mentioned in this account; the emphasis is on the internal stratification of the content plane of language into semantics (content substance) and lexicogrammar (content form). In any case, the account is in conformity with formal abstract symbol processing models of 'meaning'.

Such models assume the existence of a world or environment 'out there' that is transformed into meanings by the categories that are internal to the language system. Halliday does not discuss the biological basis of this assumption. Rather, there is an implicit division of labour between the work of the linguist and the work of the neuroscientist. The task of the former in Halliday's account is to specify the functional architecture of the linguistic system and the role it plays in the overall process of symbolically processing or coding the world as meaning according to the semantic categories of language. In spite of the absence of any explicit discussion of the cognizers and perceivers who participate in these processes, the implicit assumption shared with the symbolic processing view is that information in the form of "experience and interpersonal relationships" is transformed into meanings. This entails some kind of representational process that underlies these processes of transformation. In sidestepping the role of cognizers/perceivers in order to focus on the architecture of the language system, we are left with a split between the "ecosocial environment" and the "linguistic system." Moreover, the absence in this account of cognizers/perceivers qua active agents provides no way of seeing that languaging behaviour is inseparable from and is an extension into the environment of the bodily processes of agents. In the alternative account to be explored and developed below, languaging is a mode of behaviour that catalyzes or mediates the organism-environment relation as a unified system. In this account, lexicogrammar acts as constraints on behavioural events whereby agents actively catalyze

their environments. Rather than the symbolic processing or construal of experience into meaning, as in Halliday's account, unified organism-environment catalysis just is experience.

On the other hand, careful analysis of real-time languaging behaviour shows that the components of linguistic behaviour exhibit a wide degree of natural and intrinsic variability. It is therefore difficult to sustain the view that the dynamical properties of real-time language events are reducible to instantiations of higher-order systemic or schematic properties. The view of language as consisting of levels of abstract form that express or realize semantic functions of various kinds is a logical reconstruction based on formal type-token or system-instance instantiation relations internal to a synchronic language system. Instead, I shall argue that language is not adequately definable as a well-defined synchronic 'object of study', but consists of populations of dynamical diachronic processes that are spread across diverse spatio-temporal scales ranging from the neural to the cultural. In languaging activity, these scales are integrated with each other in varying ways and to varying degrees in real-time languaging behaviour between persons.

8. Functions of Language: Etiological and Normative Approaches

According to functional theories of language, functions are expressed or realized by formal properties of linguistic utterances. Certain formal properties are seen as having evolved to perform some semantic or other function. Semantic functions are thus realized by properties of lexicogrammatical form on the content stratum. These functions are abstracted from the material dynamics of utterance activity. In the dynamics of first-order languaging behaviour, it is more problematic to separate function from the material dynamics in this way. Can first-order dynamics be subsumed by and reduced to a set of abstract semantic functions? The complex material dynamics of first-order languaging cannot be reduced to or explained by in terms of a set of abstract functions instantiated in abstract form. In other words, there is a problem of instantiating, in a levels account of language, abstract semantic-pragmatic functions with their own formal properties and modes of realization within the material dynamics of vocal tract gestural events.

A vocal tract gesture is a physical act of an individual that is guided and modulated by both intentionality and meaning. It is an exploratory act, or a 'choice' from a seeking system, in the sense that in acting on some aspect of its world by means of the gesture, the individual seeks responses that will provide further information about the world. Consider an agent who is intent on getting a cup of coffee in Starbucks -- he or she produces the utterance, *a tall cappuccino*, on being called to the counter by the salesperson. The choice of a particular action trajectory reduces the uncertainty of 'what next?' Both the speaker's desire for a cup of coffee and the salesperson's readiness to provide the desired service could be said to elicit or call forth the utterance, which then reduces the uncertainty of what will happen next. The utterance qua physical gesture thus serves as a constraint on the entropy production of the event (Salthe 2011: 423). It serves the function of getting the speaker what he or she desires – the tall cappuccino. The utterance does not *have* this function; this function is not in any way encoded in the formal organization of the utterance. Bickhard (2008) has discussed at length the difference between etiological approaches to function (i.e. something *has* a function) and

his idea of emergent normative function (i.e. something *serves* a function). The utterance does not have the function of requesting a cup of coffee; its component parts and their relationships cannot be explanatorily reduced to the function of requesting coffee.

On this view, the parts of the whole utterance and their inter-relations have functions in the following sense: The parts that constitute the whole have dependency relations with each other that presuppose a system of relations in which the parts have the functions that they do. Both the parts and the whole have the functions they do because of the functional relations that derive from the structures and forms of organization of the system in which they occur. On the other hand, normative function in Bickhard's sense can be modelled in terms of how something *serves* the function of contributing to the self-maintenance of a far-from-equilibrium system. In its own modest way, the utterance *a tall cappuccino* serves the organized functioning of the self-maintenance of the agent qua far-from-equilibrium system. By the same token, the agent's recursive self-maintenance requires some means of differentiating environments and the interaction potentials of these environments. Differentiations must be connected to possibilities for interaction. The utterance *a tall cappuccino* has this differentiating function. But first the speaker must assess the multiple possibilities of interaction with the relevant environment and make a selection. The desire for coffee, the opportunity this presents to chat with friends, the social milieu of Starbucks, etc. are all culturally saturated factors that potentially inform the selection of a particular interaction. Thus, the speaker determines that these factors provide indications as to the appropriateness of the environment for the selection of an interaction of the requesting coffee kind. The indications functionally presuppose that the conditions for the success of the interaction are valid in this environment.

As Bickhard (2008) shows, the interaction can only succeed if the environment is one that is appropriate to the interaction. The utterance qua interaction with the environment is a differentiator. It seeks to differentiate some aspect of the situation that is selected as the current locus of cognitive processing. Differentiations can be judged to be true, false, correct, incorrect, appropriate, inappropriate, etc. If the environment supports the differentiation in the anticipated way, then the presuppositions of the indications received are verified and the interaction will succeed. On the other hand, if the environment falsifies the presupposition, then the interactive differentiation in the utterance is falsified for the agent precisely because the agent can detect that it is a differentiation that is not supported by the environment in question. Vocal tract gestural activity is a sub-personal component of the agent that has this function.

Human agents are living systems with a specific form of embodiment. Our embodiment confers the required degree of stability on the agent such that vocal tract and other forms of gesture can be articulated with the requisite degree of reliability and stability from occasion to occasion. The differentiations that can be made by these gestures must be stable and guaranteed by the embodiment of the organism. A stable embodiment ensures that the agent can engage in the relevant environmental differentiations and interactions. The selection of a given utterance functionally presupposes that the current environment is an environment that

is appropriate for the agent-environment interaction indicated by the utterance. The example thus presupposes that the environment is one in which the uttering of the utterance will lead to a transaction resulting in a tall cappuccino. If, on the other hand, I said the same thing to the sales person in the Post Office, the current environment may not be one that is appropriate for the indicated interaction. The selection of the utterance presupposes there is an environment in which coffee can be obtained. If the presupposition is wrong, then the interaction is dysfunctional with respect to the given environment (Bickhard 2008).

It is now possible, I think, to clarify what it means to say that an utterance is a form of seeking activity. A particular environment typically has many possibilities for interaction. Starbucks is no exception. Of the multiple possibilities, one has to be selected and then acted upon. As I said before, the particular action trajectory that is selected serves as a constraint upon the entropy production of the event, guiding and honing it towards its conclusion. The multiple possibilities of the environment constitute a search space -- a global informational entropy – that must be reduced. It is precisely the choice of a particular action that narrows the space to a more determinate trajectory through that space. This implies a Shannon-type reduction of uncertainty. The action is pulled into its future by the speaker's desire for coffee in the service of contributing to the self-maintenance of the agent's far-from-equilibrium condition. Difference plays a productive role in the driving of this process. Differences, e.g., the desire of the person wanting a tall cappuccino and the readiness of the salesperson in Starbucks, mesh productively with each other so as to yield new connections in the assemblage as a whole. The productive differences between the two individuals – the desire of the customer and the readiness of the salesperson – constitute an intensive difference (Deleuze 1994) when the two individuals are meshed in a new assemblage constituted by the interaction between them. This assemblage is, as Deleuze would say, in the first instance a continuum of intensity that is progressively differentiated as capacities are exercised and tendencies released (DeLanda 2011: 91).

Some normative functional capacities of utterances are:

i. Operate on and activate implicit presupposed systems of representations through apperception;
ii. Transform the social status of participants;
iii. Coordinate actions, perceptions, understandings, etc. between persons;
iv. Locate and track individuals (human and nonhuman) over space and time;
v. Modulate the bodily feeling states of persons;
vi. Enact social-affective-cognitive assemblages of persons and other classes of individuals by coupling languaging activity to the material world;
vii. Catalyze flows of socially organized matter and energy in social-affective-cognitive assemblages;
viii. Direct attention of persons and create joint attention frames between persons;
ix. Create virtual experience of cultural entities and events that are constrained by an attunement to linguistic patterns.

Persons qua living systems are recursively self-maintenant systems. Languaging behaviour is one of the sub-personal behaviours that persons have so as to maintain themselves over time through their transactions with their environment. Davia (2006: 260) argues that if we see that

in living systems there is no distinction between function and metabolism, then the notion of function becomes redundant. If we take this step, we can focus on languaging behaviour as an aspect of the biological process – a culturally constrained aspect – whereby persons and social groupings of persons metabolize their self-maintenant relationships with their environment. The material dynamics of languaging behaviour, not abstract form, is in itself a key means whereby humans catalyze their experience of their worlds. On this view, the properties of the material dynamics constitute metabolic processes that serve to maintain the person's relationship with the world.

9. The Intrinsically Expressive and Cognitive Character of First-order Languaging Dynamics

How does orderly, patterned inter-individual languaging behaviour that is shaped by culture arise? How do globally coherent representations of environmental phenomena arise? How can we explain language as intentionally and affectively modulated behaviour between persons? What is the place of living, feeling, moving human agents in such an explanation? If language is reduced to input/output processing models, we will never develop satisfactory answers to these questions. The view from dynamical systems theory helps to provide some new perspectives on these questions. Languaging behaviour is self-organizing in the context of the larger system of human agents embedded in their environments (Thibault 2004b: 48-49). Rather than positing internalized codes or rules, abstract social conventions, and input/output processing models as the basis for explanation, languaging is a self-maintaining process that exploits gradients of uncertainty both in the environment and in the internal milieu of individual organisms (Hollis, Kloos and Van Orden 2009: 212). In languaging behaviour, a number of potential states are possible at any given moment until one particular outcome is selected by circumstances. The set of potential outcomes is determined by the constraints that are in operation. Constraints may exist in the agent or in the environment, but they do not determine actual behaviour. Accordingly, actions are graded as more or less likely. Hollis, Kloos, and Van Orden (2009) make the following relevant observations:

> On this basis, we equate the states of gradient potential sets with affordances and effectivities in cycles of action and perception (Davia, 2005; Gibson, 1977; Swenson & Turvey, 1991). *Affordances* are descriptions of the environment directly relevant for action, with reference to an organism and its effectivities; *effectivities* are descriptions of the organism directly relevant for action, with reference to an environment and what it affords (Turvey & Shaw, 1995). To this we would only add that *gradient potential states* are also descriptions of the history of an organism, directly relevant for action, with reference to its immediate future.
>
> (Hollis, Kloos, and Van Orden 2009: 216)

Utterances are patterns of behaviour that have the potential to change the organism's relationship to its environment. Such patterns persist in space and time and therefore have the ability to exert consistent selection pressures on a given population of interacting individuals (Reed 1996a: 30-31). On this view, meaning is the result of the individual's efforts to come

into a relationship with the affordances of the environment when the individual interacts with those affordances. A key concept here is *interactivity*. Interactivity is in the first instance the active sensori-motor exploration of the affordances of one's world. As Gibson (1983 [1966], 1986 [1979]) showed, perception is an active process of seeking stimulus information in the environment. This information has the capacity to change the organism's relationship to its environment. Perceptual exploration is a mode of searching for information that constrains the information that is detected. In interacting with the environment, the organism enacts changes in the organism-environment system that yields differentiations that may be useful to the organism and its further interactions with the environment. Differentiations are internal interaction outcomes that guide further activity and provide a basis on which context-dependent apperception con proceed (Bickhard 1998: 198). Perceptual exploration is a specific kind of interaction with the world that yields stimulus information that is picked up by anatomically specialized perceptual systems (Gibson 1983 [1966]).

Utterances too are interactive in the sense that they operate on some aspect of social reality and transform it (Bickhard 1998: 205). However, many interactions are not confined to specialized sub-systems such as the classical perceptual systems. Perceptual systems are just a sub-class of a wider range of possibilities for exploring the environment in order to differentiate it. Exploratory activity is a means of modulating the environment in order to differentiate it. Languaging activity is an extension of this basic fact, not an exception to it. For example, vocal tract action is a means of acting on and transforming the social world in ways not necessarily tied to the immediate environment. Vocal tract gestural activity is a way of modulating the environment of other persons in order to obtain information about them as well as to allow for the tracking of apperceptions of social realities across diverse time and place scales. Lexicogrammatical patterns are second-order cultural constraints on languaging activity that provide conventionalized solutions to the problem of coordinating the diverse understandings that agents have both of the situation they find themselves in and of each other's understandings of the situation.

The search space of all possible ways to interact with some phenomenon is a virtual continuum of unactualised potentialities until one is selected by a particular search path. This virtual continuum of unactualised potentialities is nevertheless weighted such that some pathways are more likely than others. The continuum of virtual possibilities is value-weighted so that the continuum realizes what Hollis et al (2009: 216-217) call a "gradient of uncertainty": potential states are graded according to their likelihood of being realized.

> Gradients of uncertainty are like energy gradients in physical systems in several ways (however, see Keijzer, 2003; Turvey & Shaw, 1995). Each gradient is a potential set for action that includes the likelihood of respective actions without fully prescribing the particular action that will occur.
>
> (Hollis et al 2009: 216-217)

Exploratory activity changes the content of uncertainty gradients (Hollis et al 2009: 217). A particular body movement or shift in orientation, no matter how subtle or fine-grained,

changes the uncertainty gradient. Each time an act of perceptual exploration occurs, new potentialities for action and perception are created that change the potential set. Hollis et al point out:

> Actions perpetually update potential sets, which ensure a locally unstable system - a system close to a critical state. *Critical states* are states in which oppositional "forces" (constraints that favor one or another available outcome) are precisely balanced against each other. Critical states are thus a kind of boundary between qualitatively different behaviors.
>
> (Hollis et al 2009: 217)

These constraints are the embodiment of a population's history of exploring the search space and the action trajectories it adopts to explore that space. Successful action trajectories are selected and consolidated on the basis of habit. Habit is the principle process whereby a system acquires definite boundaries and contours that are maintained through time. The habitual repetition of an action thus integrates present and past and thereby gives rise to the anticipation of potential futures that can reliably be expected to occur when the given action is performed. Habitual patterns arise through a history of success in dealing with potential future interactions. In time, these habitual patterns become more valued such that they are selected as conventional routines on account of their success in dealing with the search for the information that the organism requires for the self-maintenance of the far-from-equilibrium conditions essential for its integrity and survival. These nonholonomic constraints – the emergent conventional patterns assembled through a history of organism-environment interactions – are control parameters that constrain and direct the flow of dynamical pico-scale bodily events in languaging behaviour.

10. The Auditory System is calibrated for the Perception of Source Events via Higher-order Acoustic Values

Agents perform the activities of articulating and/or perceptually exploring environmental events, e.g. vocal tract gestures. In doing so, they deploy bodily skills that allow them to integrate certain activities – e.g. optical or auditory scanning – with the given event. The (rapid) movement of our eyes along a given graphic arrangement of notational items on, say, a page gives rise to a changing array of optical stimulation in time. We perceive and acquire sensitivity to the patterns we detect on the basis of our active engagement with environment events. Perception-action invariants therefore reside in the relationships between an organism and its environment. For example, we learn that the visual-spatial stimulus information that we pick up is constrained as a function of the movement of the eyes, the head, and the torso through space, in the process of optically scanning the given optical array. The information that we pick up covaries with the movement of our eyes, our head, and the orientation of our bodies as we scan the sequence. Similarly, perception of phonetic gestures qua environmental events is dependent upon our active orientation to and exploration of the acoustic array of stimulus information that is caused by someone's vocal tract activity during the act of speaking. The time-locked perceptual exploration of the acoustic properties of speech (and

other acoustic) events is also motivated by the behaviour of the source that caused the acoustic event (Warren and Verbrugge 1981: 224).

Warren and Verbrugge (1981: 225) refer to experimental results that "support the general claim that sound in isolation permits accurate identification of classes of sound-producing events when the temporal structure of the sound is specific to the mechanical activity of the source … " (Warren and Verbrugge 1981: 225). They further point out: "If higher-order information is found to be specific to events, while values of lower-order variables per se are not, then it may be more fruitful to view the auditory system as being designed for the perception of source events (via higher-order acoustic functions), rather than for the detection of quasi-stable sound elements." (Warren and Verbrugge 1981: 225). In other words, the identification of the source event and its behaviour is the primary task of the auditory system. Higher-order acoustic functions are values that we perceive in source events such as a person's vocal tract activity. The higher-order perceptual information specific to that event can be used by the listener during the course of the event to assess, for instance, the speaker's feeling-state as well as the feeling which the speaker's vocal (and other) activity induces in the listener (Sections 11, 24). Or the higher-order information may specify and provide off-line access to experience of virtual cultural entities (Sections 23-25).

Bodily dynamics are intrinsically expressive and cognitive. Persons respond to the real-time dynamics of languaging behaviour as both expressive and cognitively salient in their worlds. Unlike code and symbol processing views, the theorist can focus on how agents hear and feel voices and respond to dynamics. Individuals align with and orient to affect, feeling, and aspects of events through expressive dynamics such as pitch, cadence, rhythm, tempo, duration, loudness, etc. Moreover, semantic priming shows how individuals dialogically coordinate and align their vocalizations in ways that give rise to the creation of shared semantic categories and values. Coordinated first-order languaging behaviour biases and shapes perceptions. It is by means of the emergence of social strategies in the form of inter-individual interaction routines that persons are able to partition their worlds on the basis of digital semantic categories (Ross 2007). A fundamental question, therefore, is: How do the bio-physical properties of languaging behaviour enable persons to partition their worlds using digital semantics in order to modify what persons perceive, orient to, and value?

In the following Section, I shall examine some possible answers to this question.

11. Body Dynamics, Felt Experience, and Values: The Feeling of What Happens

Time-locked and continuously changing bodily dynamics in interaction enable individuals to co-synchronize and to coordinate their actions and perspectives with those of others. Dynamics, not hypostatized symbolic forms, are what catalyze linguistic cognition. Cognition arises out of coordinated languaging behaviour as diverse factors are semantically synthesized in linguistic catalysis and integrated to core consciousness (Damasio 2010: 205-207). Change and dynamics are the key factors here. Language dynamics can induce in the self an emotional reaction that "alters the master interoceptive maps, a modification of the proto-self

ensues thus altering the primordial feelings." (Damasio 2010: 205). Damasio continues, as follows:

> Likewise, the sensory portal components of the protoself change when an object engages a perceptual system. As a consequence, the regions involved in making images of the body are inevitably changed at protoself sites—brain stem, insular cortex, and somatosensory cortices. These varied events generate microsequences of images that are introduced into the mind process, by which I mean that they are introduced into the image workspace of the early sensory cortices and of select regions of the brain stem, those in which feeling states are generated and modified. The microsequences of images succeed each other like beats in a pulse, irregularly but dependably, for as long as events continue to happen and the wakefulness level is maintained above threshold.
>
> (Damasio 2010: 205-206)

Continuously varying language dynamics alter the "sensory portal components of the protoself" (Damasio 2010: 205) such that the micro-sequences of images that are fed into the mind process have a first-person feel (see also Cowley 2006, 2007, 2008: 331, 333), which Damasio has designated as "the feeling of what happens" (Damasio 1999). Damasio (2010: 2007) stresses the likely central role played by timing as continuously varying dynamics instigate changes in the protoself, as follows:

> Timing is likely to play a role here too, when the causative object begins to be processed and changes in the protoself begin to occur. These steps take place in close temporal proximity, in the form of a narrative sequence imposed by real-time occurrences. The first level of connection between modified protoself and object would emerge naturally out of the time sequence with which the respective images are generated and incorporated into the cortege of the mind. In brief, the protoself needs to be open for business—awake enough to produce the primordial feeling of existence born out of its dialogue with the body. Then the processing of the object has to modify the varied aspects of the protoself, and these events have to be connected to each other.
>
> (Damasio 2010: 207)

In first-order interaction, one agent's dynamics affect or inhibit how the other perceives, acts and orients. The other modifies his or her responses on the basis of experience-based sensitivity to aspects of events as they unfold in a narrative-like sequence that is imposed by the time-locked character of the unfolding dynamics. This establishes a first-level connection between the protoself and first-order languaging dynamics. Felt experience is narrative-like. Schögler and Trevarthen (2007: 291) have shown how the changing dynamics of inter-individual movement patterns, consisting of coupled shifts in posture, orientation and vocalizations, engage participants and observers with temporally and interpersonally coherent and intentionally modulated mimetic displays that elicit 'narrative' interpretation (see Thibault 2011 for further analysis and discussion). As agents modify the environment of each other's actions through their inter-individual dynamics, they hear, see and feel what their

languaging behaviour means as its dynamics continuously modify aspects of the protoself in core consciousness.

Prosody, for example, thus continually perturbs core consciousness such that the resulting modifications of the proto-self's primordial feeling states "now become differentiated feelings of knowing relative to the engaging objects." (Damasio 2010: 209). Cognition dawns. In becoming "differentiated feelings of knowing" with respect to the dynamics and their relations to aspects of the world (the situation, the environment) with which agents co-engage in their interaction, the "objects of knowing" are made salient and are assigned values at the same time that they are placed in the felt self-perspective of the agent in core consciousness. Language dynamics therefore bias action and perception in self and other in ways that are value-weighted. This gives rise to semantic priming, based on how agents coordinate aspects of the dynamics with aspects of situations. In time, this is how we perceive wordings in dynamical languaging behaviour when we learn to attend to syllabic patterns and coordinate them with objects and events – both actual and virtual -- with which they covary in a given population of languaging agents.

12. The Sensori-motor Basis of Sign-making

The pre-linguistic vocalizations and related behaviours of infants are, of course, frequently and primarily motivated by the need to fulfil basic metabolic and physiological needs and requirements of the organism. For instance, babies will cry when they are distressed, feel hungry, or are in pain (Stark et al 1974; Demos 1988: 33). Many aspects of infant vocalizations are not exclusively motivated by metabolic processes, but by changes in the cognitive and communicational dynamic of the dyad linking caregiver and infant as a distributed cognitive-interactional system. The dyad itself is thus operated on and transformed in ways that may also lead to the satisfaction of the infant's metabolic requirements. However, the operations and resulting transformations occur at the higher level of the dyad in ways that are not fundamentally metabolic, but cognitive. Infant vocalizations are, with respect to metabolic processes, second-order processes. These may certainly play a role in the regulation of the infant's metabolism (getting food, warmth, etc.).

By the same token, the sensori-motor patterns generated by the infant's vocal and related bodily activity play a key role in extending the infant's agency beyond its body and of coordinating the infant with selected aspects of its world. Vocal activity is therefore about more than the self-regulation of first-order metabolic processes of the organism. It is also a second-order form of sensori-motor coordination with the world beyond the organism. The prosodic patterns of the infant's vocal activity are behavioural patterns that enable the infant to tune into and to interact with forms of organization in its environment, notably the behaviour (vocal and otherwise) of caregivers (Cowley et al 2004). Early infant semiosis is dependent upon the embodiment of the agent. The infant's embodiment affords a repertoire of sensori-motor behaviours such as vocalizing, exchange of gaze, smiling, facial expressions, etc. that are constitutive of the infant's interactions with his or her world.

The dialogically coordinated character of the infant's engagements with others from the earliest moments of life and their constitutive basis in the infant's sensori-motor coordinations with the environment (the dyad) suggest that first-order languaging has its developmental basis in earlier forms of exploratory activity in pre-linguistic caregiver-infant dyads (Halliday 1975; Cowley et al 2004; Thibault 2005c). It is a form of exploratory activity aimed at seeking and obtaining information about the dyad by acting on it and getting responses from the other. Utterance-activity embodies an anticipatory dynamics in the form of stances of attention and expectation formed by intentional dynamics in the limbic system (Freeman 2000). The entire sensory cortex is therefore primed to be in an active state of anticipation. Sensori-motor behaviour is exploratory because it seeks and obtains responses from the environment of the agent – responses that confirm or deny these hypotheses formed in the intentional dynamics of the limbic system. These hypotheses are tested by "state transitions giving amplitude modulation (AM) patterns that converge into the limbic system, in mammals into the entohinal cortex." (Freeman 2000: 97). The formation of a new hypothesis primes the agent for a range of possible responses, each with its means of being effected as sensori-motor action in the world.

This is, I argue, the foundation of all dialogically coordinated processes of cognition and communication. The infant's sensori-motor coordinations with its world place tight constraints on the kinds of explanations we can come up with. Given (1) that human organisms are complex, dynamic, self-organizing systems in far-from-thermodynamic-equilibrium conditions and (2) that they actively maintain themselves in this statistically improbable state by continuously producing and synthesizing the processes that maintain them along their time-locked trajectory, we can say that the infant's sign-making is not a progression towards an evermore sophisticated means for encoding and decoding the world. Instead, in exchanging matter, energy, and information with their environments through their active engagements with it, infants learn to participate in and to semantically synthesize sign-making behaviours that come under more and more cultural constraints in the course of their development as persons (Cowley et al 2004). They do so by actively exploring their worlds and obtaining information about it in the form of the responses to their own and others' activity. Signs do not encode inner and outer states of affairs qua representations of an independent world; instead, human agents (infants, adults), manufacture signs for themselves and for others through the catalyzing actions of their languaging behaviour. They integrate and semantically compress information about their worlds from multiple sources and sensory modalities (Cowley 2008). This takes the form of sensori-motor coordinations and couplings that have the potential to bring about adaptive behaviours or orientations to behaviour that change the agent's relationship to its world (Reed 1996a).

In the following Section, I shall discuss more fully the catalytic nature of linguistic processes in relation to recent development in the biological sciences.

13. The Catalytic Nature of Languaging Behaviour

Davia (2006) has proposed a new theory that grounds biology and neuroscience in a general theory of the process of catalysis. In the following Sections, I shall further explore the

implications of this general theory for our understanding of language. Davia's hypothesis is that at every scale, "living processes are processes of catalysis, and that all biological processes mediate transitions in their environments, employing the same mechanisms as enzymes." (2006: 255). Enzymes are the prototypical catalysts: autocatalytic reactions occur when an enzyme catalyzes a chemical reaction without itself being changed (Davia 2006: 261). Following the theory of autopoesis developed by biologists Maturana and Varela (1980, 1987), Davia has questioned the view that cognition represents information about an independent world 'out there' that is external to the organism. The theory of autopoesis proposes instead that organisms create their worlds through their interactions with it.

Davia's theory of catalysis goes further and argues that the organism and its environment are unified by the mediating or catalyzing role played by solitons or travelling ways in relation to its environment. Carpenter and Davia (2006: 1081) draw attention to extensive neurobiological research that shows "the wave-like or resonance-like nature of perceptual experience itself." (Carpenter and Davia 2006: 1081). Carpenter and Davia further remark:

> These observations may reflect a single, unifying principle, namely, the wave-like processes are the way by which living organisms mediate (catalyze) their environment, and they relate directly to the organism's experience (Davia 2006).
>
> (Carpenter and Davia 2006: 1081-1082).

Organisms catalyse or mediate their environment by means of wave-like processes, including the languaging behaviour of human agents. The environment does not exist independently of the organism, but is a result of the organism's catalyzing activity. Organism and environment comprise a unified organism-environment system on account of this catalytical activity.

According to Carpenter and Davia ((2006), the nervous system is an excitable medium on account of the metabolism of glucose and other nutrients:

> We argue that the brain is best understood as an excitable medium as a consequence of the metabolism of glucose and other essential nutrients. This energy gradient is dissipated by neural activity that is constrained by the organism's structure, including that arising from its history and ongoing activities. Thus, the non-linear neural waves can be seen as a self-maintaining and self-sustaining dynamic, a solution to the boundary conditions implicit in the structure of the organism and the relation between the organism and its 'environment.'
>
> (Carpenter and Davia 2006: 1082)

Carpenter and Davia (2006: 1083) generalize catalysis from the prototypical instance of enzyme catalysis. A soliton-like wave is "a localized, nonlinear solution to the boundary conditions that constitute its environment." (Carpenter and Davia 2006: 1083). On this view, catalysis on all scales (enzyme, cell, organ, and organism) involves the overcoming of structural constraints to dissipate energy (Carpenter and Davia 2006: 1083). Boundary conditions constitute a problem space of what Deleuze would call "virtual multiplicities"

(Deleuze 1987: 23-25; 1994: 212) in the sense that "the virtual possesses the reality of a task to be performed or a problem to be solved" (Deleuze 1994: 212). In this sense, catalysis is a solution to a problem space of virtual possibilities. In individuating a specific organism-environment relation grounded in experience, a solution to the problem posed by the current balance of constraints is individuated.

Carpenter and Davia argue that a living organism is a unitary process of catalysis at all fractal levels of, for example, the human brain and body, seen as media of catalysis. A soliton-like wave is a disturbance of a medium that propagates through that medium, which is able to replenish its energy. A medium of this kind is what Carpenter and Davia define as an "excitable medium." On all levels, soliton-like waves facilitate these transitions. On this view, catalysis is generalized and applicable to many scales of organismic organization from enzymes, cells, organs, and organisms (Carpenter and Davia 2006: 1083): "A living organism can be understood as a unitary process of catalysis, mediating its environment." (Carpenter and Davia 2006: 1083). Extrapolating, languaging activity consists of soliton-like waves of a unitary process of linguistic catalysis comprised of waves of neural and sensori-motor activity. In the first instance, phonetic gestures are time-locked behaviours that exhibit patterned, wave-like properties. Phonetic gestures disturb the medium of air in structured ways and are propagated through the air to impact on the excitable media of an observer's brain and body. Furthermore, this applies across all scales of the organism-environment system as massively parallel and inter-connected processes from cell, to organ, to organism, to organism-environment system. The brain does not solely determine or control this process, which depends on structural constraints arising from a history of the organism's interactions with its environment, including interactions with other persons, perception-action invariances, normative constraints and values emanating from cultural timescales, contingencies of the situation, and affective and other subjective and inter-subjective gradients.

The perceptual pick up of first-order languaging activity and its higher-order invariances constitutes what Carpenter and Davia (2006: 1083) call a "unified dynamic." This dynamic, Carpenter and Davia continue, "is a solution to the boundary conditions arising from our eye movements, our head movements, and so forth, interacting with impinging energy (see also Noë 2004). Our experience manifests the perception-action invariance as a unified event that is more thermodynamically stable than non-unified patterns." (Carpenter and Davia 2006: 1083). A linguistic event is only implicit in a discrete set of statistical invariances in a phonetic gesture and consequent acoustic array in the environment. The unity of such an event is implicit and depends upon the organism's ability to confer structure on it through its own catalyzing activity as the perceiver orients to, attends to, and perceptually explores the event. The organism's active exploration of the sensori-motor contingencies of the event therefore is a catalytic process that is dependent on the way in which the organism actively experiences the event as a complex, multidimensional neural standing wave that is associated with the perception of the event. The ability to discern the implicit linguistic structure of such an event depends on a history of the organism's culturally constrained experience of such events.

According to Davia, the nervous system is an excitable medium, i.e. a landscape replete with energy that can be consumed and replenished (see above). The brain is just such a medium in which glucose is consumed and replenished. A soliton or travelling wave is a time-locked invariant structure that emerges and exists as a coherent, ordered entity on some scale within an excitable medium. There is, then, no such thing as an environment that is independent of the organism. Carpenter and Davia point out that motor activity is "what an observer perceives of the organism's catalytical process." (2006: 1083). Carpenter and Davia provide the following example:

> Thus, a millipede's solitonic motion as it crosses a sand dune (to morph Herb Simon's classic example of the ant) is an observer's perspective of the millipede catalyzing an aspect of its environment. These findings, if generalized, may point toward a non-reductionist reconciliation of the observation of waves in physiology and the observation of waves in behavior; that hypothesis, however, requires further development.
>
> (Carpenter and Davia 2006: 1083)

Languaging events clearly consist of behavioural waves that can be observed from diverse points of observation in a given dialogical array (see Hodges 2007, 2009 for this term). Seen from this perspective, it is the speaker's catalyzing an aspect of his/her environment. Gibson's (1986 [1979]) ecological theory of perception identified invariances that arise from the organism's interactions with its environment. Invariances are not 'out there' in the world, but are the relations of the organism's perception-action. O'Regan and Noë (2001) developed their theory of sensori-motor contingencies in order to explain how the organism makes these invariance relations explicit through its experience. Invariances are detected through the organism's active exploration of and interaction with the events and objects that constitute its environment. First-order languaging activity is bodily activity that dynamically changes the environment. It is a mode of performatory activity that makes information available to other agents in the environment (Reed 1996a: 155) and exploits the potentiality to affect other individuals (and be affected by them). It is also and simultaneously a mode of exploratory activity that seeks information from the environment by engaging with others and eliciting responses from them – responses that provide information about the agent and the agent's world to others in the dialogical array.

For example, how the listener/observer hears and feels the dynamical properties of the speaker's utterance-activity dynamically changes the listener/observer's bodily feeling states, her body dynamics, her focus of attention, and locus of cognitive processing in ways that are detected by and affect the first speaker. These changes provide information to the speaker concerning the listener, the listener's affective states, her intentional orientation, and so on. Agents' perceptions depend upon the action they perform. Perceptual experience is created or enacted as a result of the organism's active engagement. The invariants that are detected crucially depend on and acquire content for us as a result of what Noë calls "sensorimotor knowledge" (2004: 9). Noë elaborates on this notion as follows:

> Genuine perceptual experience depends not only on the character and quality of stimulation, but on our exercise of sensorimotor knowledge. The disruption of this knowledge does not leave us with experiences we are unable to put to use. It leaves us without experience. For mere sensory stimulation to constitute perceptual experience – that is, for it to have genuine world-presenting content –the perceiver must possess and make use of *sensorimotor knowledge*.
>
> (Noë 2004: 10)

We access and scan articulatory events through movements of, for example, our eyes, our ears, and our head as we attend to and perceptually explore such events. Moreover, much of this movement is self-actuated in response to the changing dynamics of the event. It is through self-movement that we actively probe and gauge the event. In this way, we learn what the relevant patterns of sensorimotor dependency are (Noë 2004: 13). The dynamical properties of such events do not necessarily yield a single or unitary visual, auditory, tactile, or other experience. In languaging too, their properties arise through our active exploration, which provides information concerning things like syllabic patterns, pitch, tempo, rhythm, degree of loudness, tone, and so on. In this way, we obtain information concerning the organization of the event on the basis of the ways in which perceptual stimulus information covaries with our exploratory activity (auditory scanning, positioning of the ears, head turning and adjusting to track the event, and so on).

14. Linguistic Catalysis and Semantic Synthesis: A Biologically Grounded Alternative to Language as Constructed Code

In the tradition of ecological psychology founded by Gibson (1977, 1983 [1966], 1986 [1979]), Verbrugge (1977, 1985) had already argued that linguistic utterances are catalysts that have the functional capacity to either activate or inhibit flows of affect, cognition, action, and so on, in both self and others (see Thibault 2011). Enzymes are the prototypical biological catalysts (Section 13). The catalytic process greatly increases the rate at which molecular reactants form a thermodynamically more stable product. The products of this process are thermodynamically more stable than the reactants.

> In enzyme catalysis, the reaction ultimately occurs because the product(s) is/are more thermodynamically stable than the individual reactants. The catalytic process facilitates the transition from the reactant(s) to the product(s) by overcoming the structural constraints of the reactants' structure and dynamics. Research suggests that catalysis takes advantage of the invariance (symmetries) of the biological structure (the protein-substrate complex) to deliver energy where it is needed to change the molecular structure. The process appears to be 'vibrationally-assisted,' a wave-based facilitation that involves a type of localized, non-linear wave, called a *soliton*.
>
> (Carpenter and Davia 2006: 1082)

Something analogous occurs on the very different organism-environment scale in utterance construal. Here I draw on Cowley (2008), who points out that changes in how we report the dynamics of talk are not reports of objectively heard patterns, but reports on how we feel the dynamics in core consciousness (Cowley 2008: 329; Sections 11, 14, 24, 30). Dynamical languaging events are, in part, coordinated waves of vocal tract and other bodily activity that are propagated into the environment when they perturb and restructure environmental media such as air and light. They are the perceived dimensions of the agent's languaging behaviour as the agent catalyzes an aspect of his or her environment. This dynamical wave-like activity triggers events that tend to settle on a stable pattern in a process of *semantic synthesis*, not encoding/decoding. This process is akin to Barbieri's (2003) organic process model of protein synthesis (Cowley 2008: 330-331). However, we do not only hear lexicogrammatical patterns; we also hear voice dynamics (e.g. prosody). Construal of a linguistic utterance as a stable percept integrates diverse factors during the catalyzing process. These factors include: lexicogrammatical constraints (e.g. co-occurrence restrictions) and lexicogrammatical patterns heard as 'wordings'; voice dynamics; recent events held in working memory; prior experience; and speech timing. To quote Cowley:

> … utterances set off events that stabilize on a pattern. In settling on a sense, participants exploit dual recognition. While hearing wordings, orientation (or reorientation), they also use prosody. As Rączaszek et al (1999) emphasise, symbols constrain dynamical events (c.f. Pattee, 2000). Using dual recognition, dynamics are integrated with a heard pattern. Organic coding thus offers a model of how we use connotations. Neural 'openness' uses recent experience and timing to integrate recognition that manufactures cognitive-communicative fusion. While akin to Cowley's (2006) felt response, in the work of Rączaszek et al (1999), the parallels with organic coding apply to neural functioning. The percepts that are identified become manufactured syntheses that arise as artefacts (utterances) prompt brains to fill out incomplete information.
>
> (Cowley 2008: 331)

The semantic synthesis of a self-organizing semiotic artefact (a linguistic utterance) depends on dynamics and timing. Following Pattee (2000), the fusion of symbols (e.g. wordings) with dynamics means that symbols constrain the dynamics. Symbols, according to Pattee, are rate-independent as distinct from rate-dependent phenomena such as the material dynamics. When symbols are integrated to dynamics, the resulting linguistic catalysis works to either amplify or to inhibit flows of cognition, affect, feeling, coordination, and action in self and others (Verbrugge 1985). Phonetic gestures qua behavioural waves and their physical properties do not stand in a correspondence-type relation to the semantic synthesis mentioned above. Instead, our experience of the semantic synthesis of diverse factors corresponds to specific patterns of neural activation (Maturana and Varela 1987: 22), not with the behavioural wave and its specific, measurable properties per se. The resulting meaning is unique to the individual just as the specific patterns of neural activation are. The physical form of the wave does not have meaning in itself; it elicits the construction of meaning in observers (Freeman 2000: 14-15). Maturana and Varela (1987: 22) show that the phenomenal experience of seeing

the colour green "can be triggered by a number of different light perturbations" such that "we can correlate our naming of colors with states of neuronal activity but not with wavelengths." (Maturana and Varela 1987: 22). The "states of neuronal activity" – neural standing waves – associated with the experience of seeing a particular colour are triggered (not caused) by "a number of different light perturbations." (Maturana and Varela 1987: 22). The same observation is valid for perceptual experience generally.

For example, the experience of hearing a particular wording in a given phonetic gesture can be triggered by different properties of the sound wave in different persons. A particular wording does not correspond to fixed physical properties of the sound wave, as analysed in different ways by phonetic notation, spectrographic analysis, PRAAT-type acoustic analysis, and so on. The soliton-like wave of languaging behaviour catalyzes a self-organizing, time-locked process of semantic synthesis of diverse, previously loosely structured factors. These factors include lexicogrammar qua virtual patterns from cultural timescales, felt dynamics, first-person experience, expectations, contents of working memory, and physical events (e.g. vocal tract gestural events). These factors are synthesized as semantic artefacts when brains are prompted to make explicit otherwise implicit and incomplete information (Cowley 2008: 331; Carpenter and Davia 2006: 1084). If the invariances are not 'out there', objectively speaking, in the world, but are in the relations of the agent's perceptions and actions – in the agent's sensori-motor contingencies – then it is, as Carpenter and Davia point out, the agent's experience that makes them explicit through its exploratory activity (Carpenter and Davia 2006: 1083; see also Maturana and Varela 1987: 22). The particular patterns of neural activation that are triggered by a particular phonetic gesture and therefore the particular semantic synthesis that is experienced by that person depend upon the intrinsic properties and structure of that person – the person's internal complexity – at any given time as well as that person's history of experiencing such events. Socially coordinated languaging behaviour is a way of concerting the unique, experience-based meanings of individual persons through attunement to gestural structures that afford processes of successive approximation of each other's meanings so that person's can assimilate to the extent that is necessary their own neural structures to the meanings of others (Freeman 2000: 15).

Brain and body are the media of catalysis in linguistic cognition. As the discussion above shows, linguistic utterances are not ready-made, unified events, e.g. form-meaning pairings or codings. The processes of semantic synthesis described by Cowley arise by virtue of the agent's experience, by neural standing waves. We do not in the first instance perceive a continuous, unified dynamic in the form of an already constituted linguistic event. Our perception of a dynamical languaging event, e.g. vocal tract gestural activity, is an exploratory activity that takes place in time. Soliton-like waves of neural and bodily activity unify the boundary conditions arising from the organism's interactions with its world. The localized, robust qualities of phonetic gestures qua soliton-like behavioural waves afford their active, reciprocal perceptual exploration by observers on the time-scales of dialogically coordinated utterance-activity between persons. Moreover, their properties are non-arbitrarily related to and embedded in the invariances and boundary conditions of the environments in which they occur and in part constitute and respond to (Section 4). Unlike the idea that

language encodes or represents phenomena in an external world in terms of the categories internal to the language system, linguistic catalysis is one instantiation of a more general phenomenon whereby biological processes mediate or catalyze experience by channelling energy via structure (see above). Enzyme catalysis does it when:

> The protein chains of the enzyme may support soliton waves that alter the conformation of the enzyme-substrate complex, affecting the width of the energy barrier. The conformational change lessens the distance between specific parts of the enzyme and thereby lessens the distance between the molecular reagents that are bound to it. This shortening increases the possibility of 'quantum tunnelling' and increases the reaction rate.
>
> (Carpenter and Davia 2006: 1084)

A catalytic reaction occurs because the products of the reaction are more thermodynamically stable than the individual reactants. The catalytic process helps to bring about and indeed accelerates the transition by overcoming the structural constraints of the reactants' structure and dynamics (Carpenter and Davia 2006: 1082). Similarly, complex, globally ramifying patterns of oscillatory behaviours across many scales of neural and bodily dynamics in languaging behaviour constitute macroscopic coherent states of matter and energy that unify many fractal scales of the organism's organization as whole-body sense-making (Sections 30-32). This suggests that oscillatory patterns of neuronal activation are embodied aspects of the organism's catalytic processes rather than representations of an external world. Linguistic catalysis is facilitated by the wave-based properties of phonetic gestures. These properties have the capacity to exploit the invariances and symmetries of biological structure "to deliver energy where it is needed" (Carpenter and Davia 2006: 1082) to change that structure.

Freeman (2000) has described the transition to a new global state of coherent oscillatory patterns that ramify across the entire cortex in these terms. Meaning, Freeman argues, is the transition from one thermodynamically stable state to another; it is triggered by the processes of assimilation when the self adapts its brain and bodily dynamics to the waves of languaging behaviour of self and others during the process of perceptual exploration referred to above (Freeman 2000: 121). Freeman explains meaning in these terms as follows:

> Meaning emerges in sequences of global AM patterns of oscillatory neural activity coordinating the neurophil of an entire cerebral hemisphere. The amplitude of the oscillations is high in local patches and low in others, as viewed through brain imaging, in which, as with all patterns, both the highs and the lows are necessary. Each pattern is a construction of the brain, with onset and termination by global state transitions. The contributions to the pattern are both local and large scale. Local details are provided by synapses that have been modified by previous learning and now shape local bursts as they emerge in patches of the forebrain, including the primary sensory cortices, the limbic system, and the brainstem nuclei. The interactions of these patches with each other and with the brainstem create a global state, which organizes and constrains local

> activity in a process of circular causality. At any given time, this state –which is constantly in flux as individuals grow and learn—is the meaning in a person.
>
> (Freeman 2000: 143)

Global states of complex oscillatory patterns as described by Freeman unify or semantically synthesise the discontinuities between energy and the diverse factors mentioned above in relation to the processes of semantic synthesis. As previously noted, a phonetic event, corresponding to a complex, multidimensional soliton-like behavioural wave, does not constitute in itself a continuous, unified dynamic. The perception of the wave affords its active exploration by the observer. This means that time-locked neural waves associated with the sensorimotor contingencies involved in our exploration of that event constitute a unified dynamic, e.g. a semantic synthesis. This dynamic is an emergent, time-locked solution to the boundary conditions arising from our bodily exploration of such events – e.g. eye movements, ear movements, head movements, etc. – interacting with the energy of the soliton-like wave that impinges on the excitable media of the brain and the body as whole-body sense-making at all scales of organismic organization (Section 32). The physical properties of the behavioural wave amount to the complex intersection of virtual futures that is immanent in the complex of tendencies and forces that accompanies the event and which anticipates a process of resolution in the form of, e.g., actualized experience. The catalytic process actualizes in experience a solution to this virtual problem-space created by the multiple physical properties of the event. The phonetic event itself occurs in a complex, symbiotic relationship to aspects of its environment - the situation, the persons participating in it, their past experience (of such events), the feeling states of each participant, cultural constraints, and so on. The phonetic event is itself an aspect of a dialogically coordinated relationship between the persons involved. The ability to hear and make semantically salient certain kinds of patterns, e.g. "wordings," in such events, or to perceive and feel a "friendly" voice as distinct from a "neutral" or "unfriendly" one are actualized patterns catalysed by observers along specific vectors of interest, attention, relevance, motivation, feeling states, and so on.

The phonetic event qua soliton-like behavioural wave physically expresses a virtual multiplicity of patterns in the auditory array of perceptual information that is specified about the environmental event (the vocal tract gesture) that caused a disturbance in the array. The adopting of a 'language stance' (Cowley 2011) by an observer in the dialogical array actualizes a specific vector – let us call it the 'wording' or lexicogrammar vector' – as an actualized singularity by virtue of the catalyzing process of the observer. Semantic synthesis and the kinds of thinking, acting, feeling, and so on that it affords agents in linguistic catalysis are based upon these processes of the actualization of singular patterns from the virtual multiplicity that the behavioural wave constitutes. It is not the encoding of a pre-existing experience as a semantic content into linguistic form. Meaning and value do not reside in any of the components that are synthesized, but emerge in and through the processes of semantic synthesis that is linguistic catalysis. First-order languaging dynamics and the affordances of their physical arrangements (patterns) therefore afford virtual multiplicities that we selectively

actualize as particular, individuated singularities according to the balance of constraints – cultural, biological, situational, subjective, and so on – at play in particular situations.

As Cowley's discussion shows, the parallel with protein synthesis is useful because we can see how the higher-order percept – the perception-action invariants – is manifested by our experience as a unified event (a semantic synthesis) that is, as Carpenter and Davia 2006: 1083) put it, "thermodynamically more stable than non unified patterns." These non-unified patterns include the various factors mentioned by Cowley, e.g. wordings qua virtual patterns emanating from cultural timescales, the bodily feeling of what we hear, expectations, norms, physical events, and contents of working memory. The self-organizing semantic synthesis of linguistic artefacts entails the parallel operation of many different processes, rather than a linear or sequential ordering of encoding/decoding terminating in the reception of an encoded message. Instead, semantic synthesis involves parallel, relatively unstructured and loosely coupled processes that offer enhanced possibilities for innovation by altering, say, the coupling or timing relations between the diverse factors involved in the synthesis that I mentioned above with reference to the work of Cowley. Participation in cultural routines gives rise to globally coherent neural activity, which Carpenter and Davia characterize as a soliton-like wave "that maintains its organization, mediating the impinging transitions and giving rise to the organism's experience." (Carpenter and Davia 2006: 1083). It is the catalytic, mediating work of this globally coherent activity that gives rise to a unified semantic synthesis of, for example, the various factors mentioned above.

Unlike enzyme catalysis, linguistic catalysis is a bio-cultural process involving self and other(s) with complex internal organization. In linguistic catalysis, the addressee of the catalysis is, in many respects, causally responsible for the cognitions, feelings, actions, and so on, which may be catalyzed. Persons have internal dispositions, inclinations, and capacities which enable them to adopt and transform in the service of their own projects the observed behaviours, routines, ideas, and so on, of others. If enough people adopt the observed behaviour in this way, the catalytic process has a chance of spreading more widely though the population.

Soliton-like waves do not reduce to activity of the nervous system; they are emergent phenomena whose existence depends on the current balance of constraints. Constraints on languaging activity derive from biology, culture, a history of interactions grounded in first-person experience, norms, current circumstances, the situation, the current states of the agent, and so on. Constraints hone and fine-tune the possibilities of cognition and action before it occurs by narrowing down or restricting the many degrees of freedom so that orderly activity can occur (Hollis, Kloos, and Van Orden 2009: 214). Constraints do not directly cause behaviour. Instead, they alter the probability of particular behaviours occurring by narrowing down the potential set to a smaller one consisting of fewer possibilities. On the other hand, the development of new skills and capacities means that prior constraints are relaxed and the degrees of freedom are increased such that the cognitive and contextual reach of the organism is extended. The potential for the soliton-like waves that characterize languaging behaviour is therefore dependent upon the agent-environment system, rather than being located in either

the agent or the environment per se. Capacities, dispositions and constraints may be located in either of these, but none of these as such causes behaviour. Capacities and constraints narrow the probabilities of behaviours occurring, but actual behaviours are triggered by the contingencies of particular situations. Contingencies make particular behaviours happen. Capacities, dispositions, and constraints specify the probability landscape in which these can occur, but do not determine them.

The notion of linguistic catalysis suggests that language is not something we 'use' by instantiating tokens from a system of types or a code that we deploy to decipher an independent world. In animated first-order languaging activity, participants do not have a sense that they are separate from their languaging or that the objects and events that are coordinated with their languaging are separated from the self. Instead, they have a felt sense of being caught up in an event and its flow. The idea of language as an input/output machine consisting of different levels of coding localizes language to some specific performatory and receptive functions of the body or reifies it as 'text'. However, the catalytical view suggests that life processes at all levels are unified by self-similar fractal patterns on multiple scales of the entire body-brain system: languaging is whole-body sense-making in this sense. On this view, linguistic catalysis triggers large-scale complex oscillatory patterns in the nervous system that constitute large-scale unified states that potentially ramify across all scalar levels of the organism's organization and affect it (Section 32).

Subtle modulations in first-order languaging dynamics can alter the gradient of uncertainty in interaction (Hollis, Kloos, and Van Orden 2009: 216; see Section 9 above). So too can changes in intensity of affect, the bodily orientation of interactants, and aspects of situations, amongst other factors. First-order languaging dynamics and the reciprocities they create promote habits and routines in caregiver-infant dyads that lead to a process which Stern (2004) called "moving on." The processes of moving on lead to the creation of a repertoire of now-moments. As dialogically engaged persons such as infant and caregiver move along together, they enact and create a series of now-moments that have their origins in the narrative processes generated in core consciousness by the proto-self (Section 11). These brief narrative units lasting no more than a few seconds give rise to intersubjective moments of affect-charged "meeting." According to Hart, meetings of this kind "increase the ability of the nervous system to intensify and coregulate with someone else's activity, and they occur only when the infant's and the caregiver's nervous systems have been able to engage in mutual adjustment and self-regulation, the condition that Stern refers to as moving along." (Hart 2011 [2006]: 27). The now-moments reciprocally bind the caregiver's and the infant's neural and bodily dynamics in states of intersubjective entanglement (Section 32), resulting in mutual recognition and shared experience. Moments of meeting intensify the flow of affect through the entangled dynamics of the two agents. Meeting "promotes the nervous system's capacity for self-regulation and attention control." (Hart 2011 [2006]: 27). Moments of meeting, as defined by Stern and Hart, are, to quote Hart, "a dyadic expansion of the consciousness." (Hart 2011 [2006]: 27).

The intersubjective processes of moving on and the resulting now-moments described by Stern and Hart are the precursors of conversational dialogue. Dialogue can thus be seen as successive loci or rhythmic pulses of intersubjectively coordinated and deictically grounded experience in the form of short pulses or temporal intervals that are connected to each other along their trajectory. Lieberman (1967) showed the link between speaking and expiration in this connection. Expiratory airflow from the lungs is one of the parameters in terms of which dialogue is organised. The so-called "breath-group" is one of the factors that chunk speaking into temporal intervals that correspond on higher levels to different syntactic units – clause, phrase, and word – though in ways that tends to be language-specific rather than universal. Pulses of experience are catalysed in and through cycles of dialogically organized catalytic activities that loop between participating agents in dialogue and their environment(s). Dialogue is a succession of now-moments experienced in this intersubjective space-time that are deictically grounded as metricised rhythmic pulses or intervals of intersubjectively coordinated and deictically grounded experience that interactants co-orient to. Seen in this light, second-order lexicogrammatical resources of clause grammar and the resources for combining clauses into larger complexes (Halliday 2004 [1985]: chap. 7) are forms of cultural scaffolding that enable agents to solve problems of coordination as they attempt to move along together in dialogue.

15. Phonetic Events as Travelling Waves

Brain and body are the media of linguistic catalysis. Phonetic events resulting from vocal tract activity are not in themselves unified phenomenon. They are unified by the agent's experience. The neural waves associated with the perception-action invariances of an event of this kind constitute a unified dynamic. This dynamic is a solution to the boundary conditions arising from our ear and eye movements, head movements, and so on, interacting with the energy from the phonetic event as it impacts the organism. It is experience that registers the perception-action invariance as a unified event that is more thermodynamically stable than are non-unified patterns (Section 14). It is the skilful exploration of audible and visible patterns of languaging behaviour that give rise to coherent neural activity in the form of a soliton-like wave that maintains its organization as it mediates the transitions in the energy impinging on it and give rise to the agent's experience. How can we connect these observations to phonetic events? In the remainder of this Section, I make some preliminary observations directed to the answering of this question with reference to the gestural phonology of Browman and Goldstein (1992, 1995) and Fowler (1980, 2010).

Browman and Goldstein (1992, 1995) showed that phonological entities are temporally overlapping or co-articulated gestural activities of the vocal tract that they have called "phonetic gestures." Phonetic gestures are bio-physical actions, not abstract mental categories (Fowler 2010). Phonetic gestures are biomechanical activities of the vocal tract that create and release constrictions in the vocal tract in different ways and to different degrees. For example, the consonant /p/ is articulated by two gestures: (1) lip closure or constriction of the lips; and (2) a devoicing gesture constricting the vocal folds in the larynx. A phonetic gesture is a distinct unit of action that takes place in the speaker's vocal tract. For example, in the lip-

closure gesture in /p/ the two lips typically come together. This involves the movement of three independent articulators that all play their part in effecting this particular configuration of the lips, i.e. lip closure. These movements are: (1) the upper lip is displaced downward with respect to the lower teeth; (2) the lower lip is displaced upward with respect to the lower teeth; and (3) the raising of the entire mandible. The lip closure gesture is thus the result of a synergy of interacting factors that give rise to a functional unit called a coordinative structure (Turvey 1977).

During the time-bound performance of the lip closure gesture, the three movements mentioned here form a coordinative structure in the sense that the diverse articulatory movements actively synchronize with and modulate each other. The three movements are independent variables that can function in other structures, but in the lip closure gesture they form a cooperative coalition or synergy of the variables that lasts for the duration of the performance of the gesture (Fowler 1980). Lip closure is thus a unit of action that can be described microscopically as involving a very large number of interacting co-articulated muscular movements at the same time that it can be described macroscopically as a unit of action involving a much smaller number of variables required for the performance of a given task (e.g. lip closure) (Browman and Goldstein 1992, 1995).

A defining characteristic of the articulatory phonology of Browman and Goldstein (1989, 1992, 1995) is the view that the goals of speech production are defined by local constrictions effected by specific organs of speech in contrast to the traditional view in linguistics that the goals of speech actions are defined in terms of the properties of the sounds that result from the actions of vocal tract gestures. Browman and Goldstein (1989) argue that organs constitute an intrinsic partitioning of possible vocal tract gestures into categorically distinct types. For example, lips and tongue body are "intrinsically different structures" rather than "points on some continuum" (Browman and Goldstein 2002: 9). Phonologically, the actions of distinct organs are basic. For example, most if not all languages effect contrasts based on vocal tract organs, e.g. Labial (lips), Coronal (tongue tip/blade) and Dorsal (tongue body) stops, and Nasality (velum). Browman and Goldstein (1989) also cite developmental evidence based on the experimental work of Meltzoff and Moore (1977; see also Kuhl and Meltzoff 1982, 1984) on infant facial mimicry of adult facial gestures. When experimenters performed facial gestures involving specific facial organs (e.g. lips, tongue, eyes), infants produced movements that approximated the adult movement in organ-specific ways. For instance, the infant may imitate a tongue protrusion task, in which the experimenter protrudes the tongue sideways, with a straight out movement of the tongue. The early ability of infants to individuate and match organs on the basis of optic and kinaesthetic information is of course underpinned by mirror neurons. Studdert-Kennedy and Goldstein (2003) examined a corpus of phonetic transcriptions of infant speech - their own and that of Vihman (1996) – to show that children show a lot of consistency in the vocal tract organ that they use to articulate the initial consonant of a word. To quote Browman and Goldstein (2002: 10):

> For example, a given word might always begin with a lip gesture, but vary as to constriction degree (stop, fricative, or approximant) or in voicing. Thus, children appear

> to acquiring (sic) a relation between actions of distinct organs and lexical units very early in the process of developing language. This relation is apparently not mediated by a phonological unit specified for additional features. At this stage, it is the organ identity itself that infants incorporate in early lexical items.
>
> (Browman and Goldstein 2002: 10)

On this view, a gesture is a constriction action of one vocal tract organ, as we saw above in relation to the lip closure gesture. In that case, three articulators – upper lip, lower lip, jaw -- form a functional synergy to effect lip closure. Of course, a number of vocal tract organs, not just the lips, are involved in the formation of constrictions during speech. Vocal tract organs include lips, tongue tip, tongue body, velic aperture, and glottal aperture. Vocal tract variables pertain to each organ. Table 2 summarises and adapts discussion in Browman and Goldstein (2002: 12).

Vocal Tract Organ	**Vocal Tract Variable**	**Articulators Involved**
lips	lip protrusion	upper and lower lips, jaw
	lip aperture	upper and lower lips, jaw
tongue tip	tongue tip constriction location (location of constriction along vocal tract)	tongue tip, tonguebody, jaw
	tongue tip constriction degree (size of variable serving as rest position for constriction action)	tongue tip, tongue body, jaw
tongue body	tongue body constriction location (location of constriction along vocal tract)	tongue body, jaw
	tongue body constriction degree (size of variable serving as rest position for constriction action)	tongue body, jaw
velum	velic aperture	velum
glottis	glottal aperture	glottis

Table 2: Vocal tract variables and articulators for the five vocal tract organs (adapted from Browman and Goldstein 2002: 12)

Lip closure and devoicing are two discrete gestural units that have the potential to enter into productive relations with each other, e.g. in /p/, or with other combinatoric phonological

units. Thus, /p/ is an organized assemblage formed from the constriction relations between the two discrete gesture units. Talk is decomposable into a small language-specific repertoire of such action units that have the additional properties of discreteness and context-invariance. These gesture units are not the same as the segmental units and features of conventional phonological representation, though they are of a similar degree of graininess (Browman and Goldstein 1995). They are in any case the basic units of phonological structure. This is so in two senses. First, they exhibit properties of the combinatoric units characteristic of traditional phonology. Secondly, they are measurable bio-physical processes qua units of action. In the first sense, they are seen as time-invariant, discrete, and low-dimensional – all key properties of combinatoric systems. In the second sense, they are seen as time-varying, continuous, and high-dimensional – all key properties of real-time bio-physical processes.

In the traditional view, the two sets of units were viewed as incommensurate. Properties pertaining to the first sense were not found in the physical speech signal. The solution was to postulate that formal phonological units belonged to an abstract mental code that existed in the brain/mind of the individual. On this view, the physical properties of real-time co-articulation have no systematic relation to the abstract formal units of this internal mental code. It was the task of the speaker/listener to infer the abstract units of the code from the messy details of real-time co-articulated speech events (see Fowler 1980, 2010). Moreover, Browman and Goldstein (1995) showed that the two senses – the formal, combinatoric and the bio-physical – can mutually constrain each other in various ways. The solution, according to Browman and Goldstein (1992, 1995), lies in shifting away from abstract formal units as the components of combinatoric systems to distinct action units, i.e. phonetic gestures. These action units have the properties of discreteness, context-invariance, and are small in number. They are the most basic units of phonological structure. They therefore fulfil the requirements of both formal combinatoric and bio-physical processes, simultaneously.

The creation and release of constrictions in vocal tract gestural activity articulates a unified and dynamical wave-like pattern of energy and organization in the form of the combinations of phonetic gestures that articulate words and larger units. The constriction gestures that form and then decay and which we associate with specific phonological segments, traditionally defined, turn out to be concurrent, rather than sequential. Browman and Goldstein (1992, 1995) have called this overlap "co-articulation." As we saw above, the consonant /p/ involves the co-articulation of two overlapping gestures, i.e. lip closure and devoicing. These two gestures are concurrent, rather than sequential. In utterance-activity, temporally overlapping phonetic gestures are performed by distinct vocal tract organs in the production of phonological units such as consonants and vowels, as shown by Browman and Goldstein. In the case of /p/, this means that the lips can yield a closure action at the same time that Voicing occurs. Consonants and vowels are, then, complex, multidimensional events that express physical information about the behaviours of more than one organ simultaneously.

In Gibson's terms (1986 [1979]), this is information about a source-event (an organ-specific gesture or constellation of gestures) in the environment rather than information about the acoustic signal (Fowler 1980, 2010). The organ-specific character of phonetic gestures means

that the gestures produced by these organs do not blend during speech production (Browman and Goldstein 2002: 18). The discreteness of the source organs involved in the articulation of consonants and vowels multiplies the differentiations that are effected by a particular phonological unit. The increasing capacity for complex, multidimensional differentiation by means of phonetic gestures in a given population of interacting agents augments the available pathways for energy dissipation through linguistic catalysis. The fine-grained sensorimotor differentiations made by phonetic gestures mean that languaging agents are able to constantly fine-tune and adjust their catalysis to respond to a vast range of different environmental events (Thibault 2004a: 184-187). This accounts for the vast range of environmental differentiations that a person is able to articulate and experience in and though linguistic catalysis. On this view, phonetic gestures are physical (bodily) activities that are linguistically significant qua physical (not mental) events. They are the actions that cause the acoustic speech signal to be propagated in its environment from a source-event (a speaker's vocal tract activity) relative to some dialogical array and its observer positions. In keeping with a central tenet of Gibson's ecological theory of perception, Fowler (2010) points out that "the relation between acoustic structure and corresponding perceptual object is considerably more direct than it can be if language forms are mental categories." (2010: 291). Fowler further argues that language forms are adapted to two critically important parity constraints.

First, they are physical events that are perceived, rather than mental ones that are inferred. Secondly, language forms are preserved throughout a communicative event such that speakers and listeners do not lose them and therefore do not have to recover them in the course of that event (Fowler 2010: 292). They are held as contents in working memory on timescales that persist throughout a communicative event. Different phonetic gestures physically and causally structure the air differently such that the patterning in the air over time provides information about the gestures themselves and their source qua environmental event. On this view, Fowler argues that language forms are public events that are "adapted to the achievement of parity." (Fowler 2010: 292). Seen in this light, phonetic gestures appear to be good candidates for travelling waves or solitons. Soliton-like waves are localized and robust in the sense that they do not easily dissipate (Carpenter and Davia 2006). Solitons need not be solitary waves, as in the case observed by J. Russell Scott in the mid-nineteenth century of the solitary wave of water that maintained its structure for two miles when a boat suddenly stopped in a canal. Solitons can be complex, multidimensional patterns.

Phonetic gestures are time-locked environmental events that persist and are preserved during a languaging event as coherent, ordered events exhibiting higher-order perceptual invariants within the environmental space and energy of an excitable medium, e.g., the air in the external environment, the auditory cortex in auditory perception, and the brain. The formation and persistence of phonetic gestures qua soliton-like waves is related both to the invariances and symmetries of its environment and to context-varying information. Simple invariant acoustic and other properties have proved elusive owing to the fact that the acoustics are not produced by a given, localized segment, but by the time-locked trajectory of the vocal tract gesture as a whole during its formation. Phonetic gestures exhibit another general characteristic of soliton-like waves insofar as they mediate transitions by channelling energy

into structure. They do so in time. The simultaneous production by the vocal tract of vowel and consonant restrictions and the transitions between these in a complex, time-locked temporal trajectory of vocal tract activity is just such a channelling of energy into structure. If vocal tract gestures truly were sequential rather than temporally overlapping or co-articulated, they would not be capable of expressing the large amount of high-dimensional information – the rich physical detail – that is in fact characteristic of vocal tract gestural activity.

Catalysis in general involves the overcoming of structural constraints to dissipate energy, as in the paradigmatic case of enzyme catalysis. Body and brain are excitable media with travelling waves of excitation and inhibition. Carpenter and Davia (2006) extend the domain of catalysis to apply to the macroscopic level of biological systems embedded in their environments. According to Carpenter and Davia, all biological processes mediate transitions by channelling energy via structure. As pointed out before, these researchers argue that this occurs at all scales from enzymes, cells, organs, and organisms. At all of these scales, a living entity just is a unitary catalytic process of mediating its environment. Phonetic gestures are one instance of the complex, multidimensional behavioural waves whereby an organism unifies the boundary conditions resulting from the organism's interactions with its environment. Brain and body are the media of linguistic catalysis.

Co-articulated gestures, as shown in the work of Browman and Goldstein (1989, 1992, 1995), overlap in real-time, are continuous, and show considerable context-sensitive variation from speaker to speaker and occasion to occasion just as prosodies do. Decades of research and experimentation in phonetics (e.g. Shankweiler et al 1977; Browman and Goldstein 1992, 1995; Port 2007, 2008) have shown that topological-continuous variation in the dynamics of articulation is, in actual fact, the norm as distinct from the long-standing assumption that the articulatory and auditory space of speech events is segmented into discrete, invariant formal segments, e.g. phonemes. This observation applies to parameters such as voice-onset time, vowel quality, consonant place of articulation, the duration of vowels and consonants, and the temporally overlapping, context-sensitive variation that is characteristic of co-articulation (Browman and Goldstein 1992, 1995; Fowler 2010; Port 2007, 2008). As pointed out in Section 13, observers of such events actively explore and probe these and other parameters in order to discover the relevant patterns of sensorimotor dependency (Noë 2004: 13). The parameters mentioned above as well as vocal tract organ-specific constrictions are some of the invariants that are detected in phonetic gestures. They are aspects of very fine-grained sensorimotor discriminations that effect transitions in co-articulated phonetic events. As noted in Section 13, these invariants acquire content through the application of the perceiver's "sensorimotor knowledge" Noë (2004: 9) as perceivers mediate these transitions during linguistic catalysis.

A languaging agent who is sensitized to the patterns of invariance in the given phonetic gesture will manifest a travelling wave across his or her auditory cortex as he or she attunes to the acoustic signal and its dynamical, time-locked properties. The wave is a complex pattern of AM activity (Freeman 2000): neurons firing at particular amplitudes constitute the wave structure. The wave is context-sensitive such that the environmental events with which the

wave covaries in experience give rise to statistical learning in a population of interacting agents. This means that populations of travelling wave patterns consisting of phonetic gestures or gesture-types are associated with particular classes of environmental events in apperception (Verbrugge 1985, Thibault 2011). Phonetic gestures therefore exhibit another feature of travelling waves: they propagate a fixed structure and its associated energy within a dialogical array on account of the unity of structure and energy that is characteristic of travelling waves. They are, as Fowler shows, typically preserved rather than lost during a languaging event. They therefore do not dissipate easily. Also, phonetic gestures are stored as rich and detailed (high-dimensional) auditory and articulatory memory for linguistic patterns (Port 2008: 4). The capacity of phonetic gestures to be preserved over a given languaging event lends support to the idea that it is the rich, physical detail of such events which agents remember and recall on future occasions on the basis of their direct experience of such events, rather than sparse phonological schema (Port 2007, 2008).

Languaging behaviour consists of cycles of dialogically coordinated catalyzing activity that loop through agents and their environments. Human agents make explicit organized linguistic structure in the implicit structure of phonetic gestures by their catalyzing activity (Section 13). This is an active and skilful process whereby agents adapt their bodies to the skilful sensorimotor exploration of the phonetic event, in the process making body and brain in some respects conform to and become similar to the event (Freeman 2000: 120). As Freeman also points out, the forms that one discovers in such events when one shapes body and brain to the event in this way do not pass from event to brain. Instead, the application of sensorimotor knowledge and the discovery of forms – high-order physical invariants – in such events catalyzes in the imagination flows of cognition, affect, action, and understanding such that we are able "to create the internal structures with which we can act and understand." (Freeman 2000: 120).

In the following Sections 16 to 19, I shall consider in more details how this happens. These Sections (and others) will appear in Part 2 of this article. Part 2 will be published separately in the following issue of PJOS III.3 (2011).

16. Intentions as Control Parameters and Grammatical Order Parameters

Intentions flow through and modulate the dynamics of articulatory trajectories. Thus, the intention to articulate the lexeme /pat/ rather than /bat/ is, in part, an intention to articulate a lexeme with an initial voiceless consonant /p/ in contrast to the initial voiced consonant /b/. Gafos (2006: 67) points out: "The intention to communicate a lexeme with a final voiced consonant, in particular, is defined as a part of a dynamics that attracts the order parameter toward the intended voicing. In turn, intentions are constrained by the grammar dynamics, namely, by how forms 'should be produced' in specific contexts." Gafos argues that "intentions are communicative goals." (2006: 67). I would see this as only partially true. Intentional dynamics, following Juarrero (1999), flow through the entire action trajectory from its onset to its conclusion and modulate it. They are not simply target goals though I am in agreement with Gafos, who also says that, "Intentional dynamics adds an attractor at the required value of voicing {-x0, x0}, where x0 '=' [-Voiced], -x0 '=' [+Voiced]." Intentions

are non-grammatical control parameters that can vary in intensity. This control parameter interfaces with and interacts with a grammatical order parameter (Gafos 2006: 68). Gafos observes: "Order parameters describe the macroscopic form of phonology and grammar principles refer to such parameters … " (2006: 68).

Lexicogrammatical patterns and formats are socially and culturally distributed re-descriptions of fragments and combinations of fragments of many utterances experienced by speakers as rich, high-dimensional phonetic gestures in the contexts in which they covary with aspects of human experience. Port (2010: 316) shows that fragments of these concrete memories "are assigned to categories in many ways by speakers of the language related to what we call semantics (e.g., singular/plural, etc.), phonology (e.g., +/-voice, /b, d, g/), and syntax (e.g., grammatical categories)." (Port 2010: 316). Grammar order parameters arise in a population of speakers owing to the increasing conventionalization and standardization of aspects of rich, high-dimensional phonetic gestures as phonological, lexicogrammatical, and semantic categories. Such categories represent formalizations of different aspects of first-order languaging dynamics as formal categories in a second-order rationalization or re-description of language, seen as consisting of different levels of abstract form in a language code. Aspects of dynamics are, accordingly, frozen or hardened as normative constraints on the dynamics through selection pressures arising from informal and formal teaching and learning, the enforcement of cultural norms, the modeling of speech in terms of writing, and so on.

Thus, phonetic gestures get re-categorized as socially distributed and increasingly norm replicating phonological forms that stabilize pronunciation routines. In turn, these phonological routines get associated with semantic categories in ways that are stabilized at the population level as regular, habitual patterns of association between phonological forms and semantic categories. The cultural standardization of these patterns of association thus gives rise to 'grammar' as the re-categorization of these routines and their habitual semantic associations such that the two are mutually constraining and linked to particular situation-types and their conventions. By the same token, the emergence of 'grammar' goes hand-in-hand with the augmented cognitive and meta-linguistic capacity to reflect on and to selectively analyse (aspects of) lower-scalar dynamical processes in terms of culturally sedimented formal categories that provide the basis for folk-theoretical ways of scaffolding and coordinating interactions between persons around appropriate norms. The formal categories thus provide a distorting lens through which lower-scalar dynamics are perceived, analysed and evaluated. 'Grammar' therefore gives rise to an increasingly self-maintaining and quasi-autonomous cultural dynamic of meta-organizations composed, in turn, of real-time lower-scalar dynamical processes on which 'grammar' induces us to reflect in culturally regimented and constrained ways. Lexicogrammatical patterns are population-level (e.g., cultural) constraints on the bio-physical dynamics of phonetic gestures. The gestures compress in their dynamical properties accumulated historical-cultural information and higher-order invariant properties (Section 10) that have the functional capacity to evoke culturally specific forms of virtual experience and virtual semantic entities. These 'digital' semantic differentiators bias action, feeling, cognition and perception in value-weighted ways that give rise to higher-order behavioural control (Sections 4, 11, 18, 25).

Seen in this light, 'grammar' specifies culturally constrained and standardized order parameters that also interface with and are modulated by the intentions of languaging agents. Intentions flow through real-time neural and bodily dynamics of first-order languaging behaviour and are not reducible to the order parameter set by culturally regimented grammatical form. By the same token, the intentional dynamics of agents are attracted to and flow into the attractor spaces that are set by lexicogrammatical order parameters. Intentions are control parameters because they flow from an intentional source – the neural dynamics of an agent – through body dynamics and out into the environment along the temporally unfolding action trajectory, e.g. a vocal tract gesture. In this sense, they are an aspect of the mechanisms whereby agents control and modulate their own and others' behaviours relative to that environment. At the same time, 'grammar' is an order parameter in the sense that it is a mechanism-independent quasi-causal constraint or attractor qua future cause that shapes, guides and sculpts the development of the action trajectory as it is pulled into its future. This occurs in ways that are constrained by cultural norms that serve to orient, scaffold, and coordinate the intentions of agents when they interact with each other. Intentions, then, are not so much goal-directed, but anticipatory.

Gafos points out that "intentional strength is a scalar variable" (2006: 70) that varies continuously along an interval or cline. The intention to express this or that contrast, e.g. the contrast between [Voiceless] and [Voiced] in the /pat/ vs. [bat] pair is motivated by contextual factors. However, intentions are not causes of actions. They are aspects of the processes themselves (the actions) that may change a system from one state to another. They are immanent in the material dynamics of action trajectories such as phonetic gestures. Intentions refer to the actual world at the same time that they interface with the virtual world of 'grammar' by virtue of the multiple attractors that compete for the dynamics of the unfolding action-trajectory, defined as a distribution of singularities. A particular phonological value, e.g. Voiced or Voiceless, is a singular point or a singularity -- a stable point in phase space towards which vocal tract trajectories tend to converge. A trajectory is defined as a stable behaviour that changes as the behaviour changes state, i.e. it passes through a series of points in phase space as it converges towards its attractor.

Attractors are singular points in this sense. A state space can have several attractors. Its singular points are thus surrounded by a basin of attraction, a region that affects the behaviour of trajectories. If behaviour starts in a particular basin of attraction it will end up at the attractor. Attractors thus define regions of stability; they stabilize an attractor around a set a values that restrict its degrees of freedom. Intentions, on the other hand, are control parameters that function to take a stable system (an order parameter) to one of its critical thresholds. They are scalar variables that vary in intensity such that a critical threshold is reached and a bifurcation occurs, i.e. one stable distribution of attractors (an order parameter) is changed into a different, contrasting one, as shown above in the discussion of the lexemes /pat/ and /bat/. Intensive variables such as intentionality are control parameters that can transform one distribution of attractors into another one once a critical threshold of intensity is crossed. Whereas order parameters define regions of stability, control parameters such as intentionality define the potentiality for phase transitions to occur that dynamically transform

one stable distribution of attractors into another, different one.

Control parameters and order parameters together define an immanent space of possibilities. Intensive properties such as intentions qua control parameters are not therefore extrinsic to this space of possibilities, but are immanent in them. A phase space so defined corresponds, in the Deleuzian ontology, to a *multiplicity* (Deleuze 2004: 230-235). Multiplicities define the spontaneous capacity of a material process to generate pattern without external intervention (DeLanda 2002: 26). A multiplicity is defined by the distribution of singularities in a material system that define its tendencies, and the critical transitions that take the system from one phase state to another. Intentions can thus be seen as a topological-continuous space that progressively differentiates and specifies itself into the partitionings of vocal tract variables into stable, discontinuous regions that afford the reciprocal coupling and mutual attunement of speakers to each other's vocal tract dynamics.

Gafos sums up his discussion as follows:

> There is a parameterization in terms of an order parameter and a control parameter, in (11a, b) respectively. Order parameters describe the macroscopic form of phonology and grammar principles refer to such parameters (see Gafos 2002 on gestural coordination relations). In our example, the control parameter is intentional strength. As shown in (11c), there is also an 'interface', the hypothesized model relating these two parameters, dx/dt = G(x) + intent * (xREQ– x), where G(x) = – k + x – x^3. Crucially, however, this 'interface' does not translate symbols to continuous signals. Rather, it states a dynamic linkage, in the form of a testable relation, between a grammatical (order) parameter and an extra-grammatical (control) parameter. The linkage is dynamic because the two parameters it relates are interdependent and changing quantities, …
>
> (Gafos 2006: 69)

Intentional strength is a scalar phenomenon. Once the intentional strength interfaces with the grammar, the latter moves away from equilibrium. If the intentional strength is made intense enough, the grammar system can undergo a phase transition to a new stable state. Intentional strength is defined by continuous intensive properties that can differentiate into discontinuous regions corresponding to particulate phono-grammatical categories. These processes of differentiation occur because of the intrinsic partitioning of possible vocal tract gestures into distinct categorial regions that afford attunement between the members of a population of speakers and therefore serve to differentiate lexicogrammatical units. Constriction parameters of the vocal tract are assigned quantitative values (Stevens 1989).

17. Constriction Parameters, Mutual Attunement and the Dialogical Coupling of Languaging Agents as Bio-dynamical Social Engines

Stevens showed that the relation between constriction parameters and their acoustic properties exhibit nonlinear properties. A restricted range of these values will afford attunement between speakers, whilst others will not. The topological continua of vocal tract variables can

therefore be partitioned into discrete regions that afford attunement. Stevens shows that the relation between constriction parameter values and acoustic properties divides into stable and unstable regions. In the stable regions, small changes in the constriction parameter result in small changes to the acoustics. In the unstable regions, small changes to constriction parameters result in large changes to acoustic properties. According to Goldstein and Fowler (2003: 35-39), the former afford mutual attunement between speakers because their properties afford imitation by the members of a population of speakers whereas the latter do not. The interaction between intentions and vocal tract variables can be many and varied. Intentions are immanent in the dynamics of these material processes.

Strictly speaking, imitation in the early months of the infant's life refers to the infant's capacity to relate to and to mirror the external behaviour of others (Hart 2011 [2006]: 23-24). However, mutual attunement is more than imitation so defined. Imitation alone could not get the dialogical coupling of persons through their reciprocal languaging off the ground, so to speak. Mutual attunement is based on the capacity to affect each other's **internal** dynamics and to modulate them. Mutual attunement is affective attunement and, as Hart points out, it "is essential for our ability to feel other people and to feel that we are felt, which facilitates the development of attachment capacity and enables us to relate to significant others throughout our life span." (Hart 2011 [2006]: 23). The capacity to affect each other's internal dynamics is therefore an essential foundation for the capacity of listeners to covertly simulate through the activation of neural structures that mirror the vocal tract gestures of the speaker, as proposed in the motor theory of speech perception (Liberman and Mattingly 1985; see also Kinsbourne 2005).

Sander (1977) showed that the infant's ability to self-regulate develops in the earliest stages of infancy. Caregiver and infant engage in interaction routines that promote affect-based forms of co-regulation (Trevarthen 1998; Bråten and Trevarthen 2007; Bråten 2007). The recticular activation system in the brain stem creates a temporal structure for brain activity. Neurotransmitters in this system control sensorimotor exploration, attention, and motivation. Motivation arises from the interaction between the recticular activation system, the diencephalon, and limbic system (Hart 2011 [2006]: 48). Bråten (2007), Trevarthen (1998), Stern (1977), Cowley et al (2004) and others have further shown that synchronized interaction between infant and caregiver gives rise to joint motivation and attention. The mutual regulation of arousal between infant and caregiver also results in physiobiological and psychobiological attunement between them (Stern 1984). As Hart (2011 [2006]: 49) points out, caregiver and infant modulate each other's energy states through their mutual attunement and the interpersonal routines in which this is embedded.

New patterns of neural activity form in the brain as a result of the interactions between persons and their environments. Regular, repeated experiences and intense affect establish new patterns of neural activity that reactivate more readily (Hart 2011 [2006]: 49). When mutual attunement occurs, there is 'deep coordination' (Hart 2011 [2006]: 49) of the autonomic and limbic centres in the brain, which are important for the fine-tuning of the emotional regions of the brain amongst other things such as "heart rate, breathing, digestion"

(Hart 2011 [2006]: 49). Synchronized attunement, which is based on the limbic regions, is referred to as "limbic resonance" by Hart (2011: [2006]: 49). Limbic resonance, Hart (2011 [2006]: 49-50) explains, "requires that an internal state be expressed externally." (Hart 2011 [2006]: 49). The many fine-grained motor distinctions of the face, the hands, and the vocal tract are able to articulate many very fine-grained motor discriminations and combinations of these, which partly come under the control of the cranial nerves in the brain stem. The cranial nerves connect to circuitry involving the diencephalon, the limbic system, the insula, and the amygdala. They also "later connect with an area deep inside the frontal lobes: the orbitofrontal cortex, which enables us to feel and understand facial expressions." (Hart 2011 [2006}: 50). In this way, the human brain and that of other primates can link the perception of gaze and facial expression to emotion, motivation, and meaning.

Simondon (2010: 385) points out that motivation are not based solely on individuals, but on the potential energy that the individual person has in relation to the social group. Individuals, Simondon argues, are partially autonomous modulators of social realities. Individuals can activate ["déclencher"] flows of energy qua intentions and motivations through action trajectories that operate on and adjust the relational dynamics of selves in their social milieu. For this reason, individual persons may be described as *bio-dynamical social engines*. The material dynamics of vocal tract activity is what enables the flow of energy through the system, i.e. along its time-locked trajectory. Forms qua topological constraints on the dynamics modulate the flow of energy. Stable regions of vocal tract activity yield organ-specific gestures that have the potential to specify differential contrasts in phonological units such as /p/ in contrast to /b/ (Section 15). Contrasts can be either typological-categorial or topological-continuous. Perceived contrasts in vocal tract and other bodily behaviours enable languaging agents to mediate transitions in their environments as they catalyze their worlds (Section 13). Contrasts -- topological and typological -- mediate transitions in the flow, distribution and modulation of the energy that flows through the action trajectory and into the environment in which it is articulated. Matter and form are thus unified in linguistic catalysis as structured energy fields that make information available to agents willing and able to respond to it.

As shown in the work of Fowler (1986, 2010), speech sounds provide observers with information about the more fundamental environmental event – the vocal tract gestures of the speaker –, which the listener attunes to so that the coupling of speakers and listeners can take place. The capacity for covert simulation in one's own neural structures of the other's vocal tract gestures clearly is a case of the speaker affecting the listener's internal dynamics through the coupling of their internal (neural) and external (bodily) dynamics. The entrainment of a population of languaging agents to similar patterns of vocal tract dynamics thus provides a solution to the problem of assimilating one's own internal dynamics to the behavioural and neural dynamics of others. It is in this way that persons can assimilate their own behavioural and neural dynamics to the experiences, feeling, intentions, points of view, and the meanings of others in socially coordinated ways that have no need for abstract codes. Languaging behaviour is an external resource whose regular, habitual, norm replicating patterns in the form of vocal tract gestures in a population give rise to population-level entrainment effects.

Kinsbourne (2005) further points out that this means that entrainment is not only in terms of bodily interactional synchrony, but also involves the entrainment of the neural structures required for the covert simulation of the other's vocal tract gestures in the forebrain of the listener. This means that the neural structures involved are entrained to the intrinsic potential for complex, fine-grained differentiations that vocal tract gestural activity affords.

Stevens (1989) showed that the stable regions of the articulatory-acoustic map are the basis for phonological units. According to Browman and Goldstein, "the partitioning (sic) such continua into discrete regions emerges from the constraint of mutual attunement under conditions where the relation between constrictions and acoustics exhibits such non-linear maps." (2003: 14). The attunement constraint means that, in time, agents will converge on and eventually settle on the same constriction values. The attunement constraint is therefore responsible for the partitioning of the constriction degree continuum into discrete regions that are interactionally salient for agents due to the population-level entrainment effects mentioned above.

As we noted in Section 15, the phonological units of a language are formed from combinations of discrete organ-specific phonetic gestures. The combination of the organ-specific gestures of lip closure and devoicing is a value-realizing action that has the value /p/. The lip closure gesture in /p/ is also present in the initial consonant of words like *mat, pick,* and *bat* in combination with other organ-specific gestures that combine with lip closure to form consonants such as /m/ and /b/. Thus, the combination of the gesture primitives [lip closure] and [devoicing] produces /p/; in contrast, the combination of [lip closure] and [voicing] yields /b/. The phonological units /p/ and /b/ can occur in initial position of the lexemes *pat* and *bat*.

The vocal tract actions of different individuals in the same community must be attuned to each other – around six months infants begin to align their own vocal tract activities to the range of constriction parameters that afford attunement among speakers. The constriction parameters that afford attunement in a population of speakers draw attention to the fundamentally ecological, value-realizing character of languaging (Hodges 2007, 2009). A primary value that is realized by languaging is the building of dialogical connections with others. The mutual attunement afforded by the given range of constriction parameters is therefore a fundamental value that is realized when speakers achieve mutual attunement to each other's vocal tract actions. Persons physically modify their vocal tract gestures in order to attune to others. Moreover, the phenomenon of 'gestural drift' (Sancier and Fowler 1997) shows that speakers can modify their vocal tract actions throughout life in response to the changing demands of the different speech communities in which they participate. Attunement is a primary value because it underpins all other forms of value-realizing that can be enacted in languaging (Hodges 2007, 2009).

The constriction parameters that afford attunement constitute a most basic way in which languaging contributes to the recursive self-maintenance of persons through their languaging behaviour. Speakers contribute to their own self-maintenance by attuning to the constriction parameters of the given population and tuning out of those that are not pertinent in that

population. Attunement can be considered to be an appropriate interaction outcome because it serves the function of contributing to the maintenance of the far-from-equilibrium stability of the person as a member of that speech community. The detection of the salient constriction parameters in the vocal tract actions of others implicitly predicates that the relevant environment is functionally appropriate to a range of interaction possibilities and that these possibilities can hold or not hold in that environment (Allen and Bickhard 2011: 109).

A given constriction parameter is an indicator of a set of interaction possibilities. If one of these possibilities is articulated, i.e. by a particular constriction parameter, then the interaction will succeed, e.g. mutual attunement will be achieved, if the environment is appropriate to the indicated possibility. The observables of the environment are the constriction parameters that are detectable in the vocal tract actions of others. These parameters implicitly predicate that the current environment is one that is appropriate to the possibility of attunement with other speakers. This entails the learning of a linkage between the appropriate parameters and the implicit predicate to be learned that is internal to the person. That is, the agent generates an internal model of its agent-environment transactions that operates on a slower timescale with respect to the agent's *in vivo* interactions with its environment.

Here, the environment consists of the other speakers with whom one can be mutually attuned through the detection in other's vocal tract actions of the appropriate parameters just as others can detect the appropriate parameters in one's own vocal tract actions. What governs the behaviour of languaging agents, on this view, are the specific features of two environmental qualities: (1) the co-presence of other speakers; and (2) attunement to the same constriction parameters. Rather than notions like shared meanings, we see that our discussion is focused on receptors (auditory, visual, haptic, etc. systems) and actuators (vocal tract activity, facial expressions, hand-arm gestures). Receptors and actuators are tightly coupled with each other and with what von Uexkull (1928) called the organism's *Innenwelt* in the creation of the action circuit whereby the *Innenwelt* is determined through organism-environment interactions. The *Innenwelt* that is so created just is the world for that organism. Moreover, the interactions that give rise to the *Innenwelt* are value-realizing. Value-realizing behaviours ensure that the encountered constriction parameters are not just a motley collection of discrete entities, but they are integrated to a meaningful whole that realizes a value for the organism.

Constriction parameters are elementary differentiators that have the capacity to indicate possible interactions with the relevant environment. In this sense, they constitute anticipations of possible interaction outcomes. The constriction parameters that are salient in a given population serve this function because they differentiate the environment in relation to the agent's value-realizing behaviours. In the first instance, this means that they have the functional capacity to achieve mutual attunement between speakers. Importantly, this shows more clearly why the constriction parameters have value for agents in their worlds – not merely because they participate in an abstract play of differential oppositions between phonemes, but because they are positive and productive differentiators (Deleuze 1994 [1968]: 254-256) whereby agents affect others and are affected by them. Their value resides in the fact that languaging agents are highly attuned to the behaviour of others. Moreover, specific

constriction parameters can directly affect interaction outcomes by amplifying or inhibiting the responses of others. This has important implications for a catalytic theory of language.

Hodges (2009: 638) reports on recent work in ecological psychology using non-linear dynamical analyses that can help to better understand these issues in relation to fractal patterns:

> Recent work in ecological psychology (e.g., Van Orden, Moreno, and Holden 2003; Van Orden, Holden, and Turvey 2005; Van Orden, Kello, and Holden, in press) using non-linear dynamical analyses, have revealed pervasive, long-range patterns in linguistic performances (e.g., word pronunciation, lexical decisions, semantic categorization). These data are startling and have important implications for our understanding of intentional activities. If, for example, someone repeatedly says a word thousands of times, and measures such as the latency and duration are taken, aperiodic waves of variation in amplitude occur across time. Instead of random variation around a "true score" of latency or duration, the waves of variation at different scales of times (e.g., blocks of 10 trials, 100 trials, 1000 trials, 10,000 trials), are similar at every scale. The similarity is so precise that if the amplitudes are graphed relative to frequencies on a log/log scale, the relationship is linear. Such a pattern is *fractal* (sometimes referred to as 1/f scaling).
>
> (Hodges 2009: 638)

Hodges further points out that no one scale or unit of measurement can capture or describe the fractal phenomenon. These observations readily apply to the constriction parameters discussed above. Hodges identifies three ways of looking at such patterns: the patterns are global, collective, and paradoxical (Hodges 2009: 638). Fractal patterns are global because the larger the sample and the more precise the measurement device, the more variability is revealed. Fractal patterns are collective because they cannot be reduced to any single causal principle or module. Instead, they are the emergent results of the interactions between many different components and many different processes and are irreducible to any single principle or component process. The fractal pattern is immanent in the interactions between the component processes over the space-time of the interaction between the components. It ceases to exist once the components cease to interact. Fractal patterns are paradoxical because a single movement, e.g. an organ-specific constriction parameter in vocal tract action, correlates with an increasing diversity of patterns across scales. Hodges observes: "Actions that occur on the scale of 0.2 sec show dependencies on other activities that are on the order of 11,000 sec (Van Orden, Holden, and Turvey 2005). Interdependence implies that each part reflects something of the whole in its behavior." (Hodges 2009: 639). For example, the correlations between organ-specific constriction parameters, combinations of organ-specific gestures, and morphemes, etc. suggests the interdependence, context-sensitivity and global inter-dependence of the different scales discussed by Hodges (2009: 639).

18. Mutual Attunement, Phonological Structure and Value

Mutual attunement of vocal tract gestures arises and is maintained in a population of languaging agents not only because of the reciprocal coupling of the organ-specific gestures of agents via the medium of air; it also arises in response to the need to solve problems of coordination between agents in emergent social situations. To participate in social situations means that one is, implicitly, a social person (Bickhard 2006). To promote oneself as a social person is to make usually implicit and inherently normative claims as to one's capacity to participate in such situations. Moreover, it means that one's own actions, including utterance-activity, occur in and are promoted within a heteroglossic field of others' desires, motives, intentions, and plans. This discovery goes hand in hand with the discovery that the desires, motives, etc. of others may contrast with one's own. One learns that the social world is populated with other persons' points of view (Reed (1996a: 167). Reed points out: "There is thus strong pressure on children to learn to select utterances that produce desirable effects." (1996: 167).

Learning about others' points of view is, Reed argues, "a fundamental cognitive milestone" (1996a: 167) in the life of the child. This developing awareness of the points of view of others entails an inherent reflexivity in the coordination problems that arise between agents in social situations and in their solutions. The social coordination of diverse points of view in the dialogical array of observer positions and perspectives in some situation calls for solutions that can provide some sort of resolution of, say, your understanding of my understanding of your understanding … in the given situation. The heteroglossic diversity of points of view that underpins this inherent reflexivity are then "key elements in structuring the child's linguistic environment in such a way as to provide information concerning the significance of varying speech structures and that set the child on the road to making discoveries about the generative patterns inherent in the language around him or her." (Reed 1996a: 167).

The mutual attunement of vocal tract gestures is, as Goldstein and Fowler (2003: 35-39) show, a process of attending to and tracking the information about the other's vocal tract gestures that is made available through the restructuring of the medium of air that occurs in the act of speaking. In this sense, phonetic gestures are public, observable events, not private, mental ones; they take place between persons in social situations. They are events which persons actively and intentionally select and organize for the purpose of presenting information to others. In so doing, they make others aware of relevant environmental information even when the other person is not physically present in the situation and has no direct, online access to the situation.

Attunement begins in the earliest stages of infant-caregiver interaction so that infants align their vocal tract behaviour to that of the others in their environment (Halliday 1975; Vihman 1996; Cowley et al 2004; Kuhl 2007). Reciprocal attunement of vocal tract gestures takes place through the acoustic medium which functions to couple the vocal tract behaviours of agents to one another though visual and haptic information can also play a role in these coupling dynamics (Meltzoff and Moore 1977; Kuhl and Meltzoff 1982, 1984). Browman and Goldstein (2002) make the following pertinent observations:

> Because the relation between constriction parameters and their resulting acoustic properties exhibits some nonlinear properties (e.g., Stevens, 1989), certain ranges of values of the constriction parameters will afford attunement among talkers (i.e., those values can be effectively imitated or mimicked), while other ranges of values will not. In this way, the continua corresponding to tract variable dimensions can be intrinsically partitioned into discrete modes or ranges that afford attunement. We will now describe a simulation of attunement that illustrates this idea.
>
> The idealized relation between a constriction dimension and (some dimension of) the sound that it produces can be seen in Figure 1, adapted from Stevens (1989). The relation between the constriction parameter values and the acoustic result shows different kinds of regions—stable and unstable. In stable regions (labelled I and III), small changes in the constriction parameter result in only very small changes to the acoustics, while in the unstable region (labelled II), small changes to the constriction parameters result in large acoustic changes.
>
> (Browman and Goldstein 2002: 13)

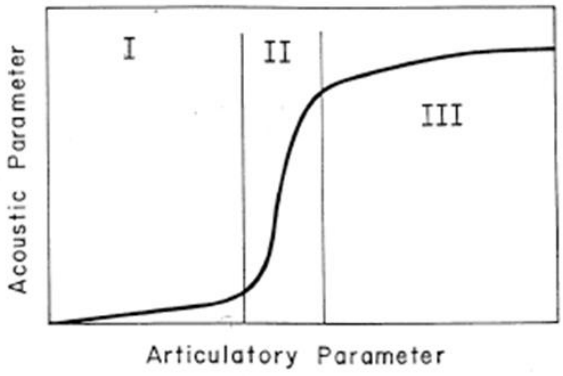


Figure 1: Non-linearity in the articulatory-acoustic map (after Stevens, 1989)

Browman and Goldstein discuss experimental results that demonstrate that the partitioning of constriction degree into discrete regions (e.g. closure, critical, narrow, mid, wide), as shown in their Figure 1 above, is "an intrinsic consequence of the attunement constraint, and the non-linear map relating articulation to acoustics." (2002: 15). Discrete categories of constriction parameters arise in the processes of reciprocal attunement that occur in a population of speakers. Browman and Goldstein model gestural structure in vocal tract activity in terms of two parameters: (1) the common distinct constriction degrees (closure critical narrow mid wide); and (2) common distinct constriction locations (labial dental alveolar palatal velar

uvular pharyngeal). The descriptors used here are no more than labels to approximate the "ranges of values of these parameters that are typically used contrastively." (2002: 16). Contrastive values are set up in a particular language when its speakers converge on these values through their reciprocal attunement to a stable constriction parameter (see above). These parameters are value-weighted such that their intrinsic tendency towards partitioning into discrete modes affords reciprocal attunement whereas unstable regions that do not lend themselves to mimicking do not. Contrastive values of gestural units in a given language arise through self-organizing processes of reciprocal attunement of speakers to one another's vocal tract gestures constrained by the shared biomechanics of vocal articulation and the coupling constraints of the acoustic medium (and to a lesser extent other media such as the optic and the haptic).

Mutual attunement is possible in the first instance because of the adaptation of phonetic gestures to the achievement of interpersonal parity (Fowler 2010: 292; Section 15). The achievement of interpersonal parity through the developmental processes of attunement referred to above means that organ-specific phonetic gestures are bonded or combined to form, in utterance-activity, gestural scores that have the capacity to selectively organize and present information in ways that foster awareness between persons (Reed 1996a: 171). Discrete organ-specific vocal tract gestures combine to form phonological units in which the discreteness of each of the gestures that combine to form a given unit is retained (Goldstein and Fowler (2003: 2). The gestures that combine in this way are potentially contrastive in the given phonological unit. For example, *bin*, *vin*, and *win* all begin with gestures of the lips, but the gestures contrast in the degree (state value) of Lip Aperture (or degree of lip constriction). *Bin* has the highest degree of lip constriction; *vin* has less lip constriction; and *win* the least (Goldstein and Fowler 2003: 35). The variable degrees of lip constriction exhibited along the continuum of Lip Aperture in the three examples are points along a scale of topological-continuous variation (of degree of lip constriction). The question arises as to how these continua are partitioned into contrasting states or values (of the vocal tract).

According to Goldstein and Fowler (2003: 35), it is the public, observable status of vocal tract gestures that provides the basis of the explanation. Vocal tract gestures, rather than being the imperfect realization of the abstract categories of an internal phonological code, are observable public events that are shaped by and attracted to the phonological norms of a particular speech community. Phonological forms are socially distributed second-order constructs with respect to the first-order phonetic gestures which speakers produce in their vocal tract activity. Phonological units and phonological structure in languages – the phonological forms of a specific language – are abstract structures with respect to the articulatory and acoustic properties of actual phonetic gestures as they are articulated and perceived in context-dependent vocal tract activity. Phonological forms are relatively more abstract and context-independent structures that are decomposable into combinations of discrete, organ-specific vocal tract gestures that are potentially contrastive, as noted above. They are low-dimensional attractor spaces with respect to the high-dimensional articulatory and perceptual detail of actual phonetic gestures in acts of speaking and listening. As low-dimensional attractor spaces, phonological forms are quasi-causal, mechanism-independent

constraints on real-time vocal tract activity. They are emergent normative constraints on the vocal tract activities of a population of interacting agents. As pointed out above, Goldstein and Fowler (2003: 35-39) show that mutual attunement is the causal mechanism that enables the emergence of phonological forms at the population level:

> Mutual attunement must be accomplished primarily through the acoustic medium. Because the relation between constriction parameters and their acoustic properties is nonlinear (Stevens 1989), certain regions of a vocal tract continuum will afford attunement, while others will not. Thus, the categories we observe could represent just those values (or regions) of the tract variable parameters that afford attunement. They are an example of self-organization through the public interaction of multiple speakers.
>
> (Goldstein and Fowler 2003: 36)

Building on the findings of Stevens (1989) referred to in the above quotation, Goldstein and Fowler (2003: 36) conducted preliminary simulations of agents attempting to attune their actions:

> In a preliminary simulation designed to investigate the partitioning of a tract variable constriction continuum into discrete regions, agents interacted randomly under the following three conditions: (a) Agents attempt to attune their actions to one another. (b) Agents recover the constriction parameters used by their partners from the acoustic signal, and that recovery is assumed to be noisy. (c) The relation between constriction and acoustics is nonlinear. The simulation investigated an idealized constriction degree (CD) continuum and how it is partitioned into three categories (corresponding to stops, fricatives, and glides). Figure 4 (in referenced document) shows the function used to map constriction degree to a hypothetical acoustical property, which could represent something like the overall amplitude of acoustic energy that emerges from the vocal tract during the constriction. The crucial point is that the form of the nonlinear function follows that hypothesized by Stevens (1989) for constriction degree and several other articulatory-acoustic mappings. Regions of relative stability (associated with stops, fricatives, and glides) are separated by regions of rapid change. The constriction degree continuum was divided into 80 equal intervals.
>
> (Goldstein and Fowler 2003: 36)

In recovering the constriction parameters from a noisy environment in which the relation between constriction and acoustics is nonlinear, agents in effect overcome structural constraints that enable them to channel energy via structure such that attunement between agents in a given population arises and, on that basis, dialogically coordinated means of mediating the environment in languaging behaviour. The developmental findings cited by Goldstein and Fowler (2003: 38) lend support to this:

> While children's early words are consistent in the oral constriction organ employed, and match the adult models in this regard, they are quite variable in within-organ properties,

> such as constriction degree (or constriction location). The simulations suggest that within-organ categories emerge only from attunement, which presumably takes some time. This conclusion is further bolstered by the recent perceptual findings with infants 10-12 months of age (Best and McRoberts in press), showing reduced discrimination for within-organ contrasts, even when the contrasts can be found in the language the child is about to acquire. At this age, infants have only begun to attune their vocal behavior to the language environment (de Boyssson-Bardies et al. 1992), and therefore partitioning of within organ categories is expected to be incomplete.
>
> (Goldstein and Fowler 2003: 38-39)

Children learn in time to control bodily dynamics, including vocalizing, in ways that control perceptual input. They do so through reciprocal attunement with others and through the honing of "within-organ properties." Initially, the infant has little control over this input, sending random commands to the muscles. In time, he or she learns to correlate certain random commands and within-organ properties with specific responses in the world, especially responses from other persons. This correlation is established on the basis of the consistencies that are established between the motor command and the perceptual input. In time, the infant can elicit desired responses in others by calling up the appropriate motor command (e.g. proto-imperatives). He or she can control vocal and other bodily behaviours in concert with others (and later solo) that establish a consensual domain of consistent motor-sensory relationships. These motor-sensory relationships bias perception in value-weighted ways that lead to higher-order behavioural control. Thus, control of vocal tract and other gestural activity mean that the gesture can be used to get others to fulfil one's needs and wants. The use of the gesture for higher-order control of this kind just is, from the child's point of view, the meaning of the gesture.

Mutual attunement via the acoustic medium to each other's vocal tract gestures is afforded by those more stable regions of vocal tract activity that, in time, give rise to the emergence of the phonological values of a particular population of speakers. In their articulatory and perceptual activities, speakers and listeners actively seek after and orient to (phonological) values in each other's vocal tract gestures that have the potential to alter their cognitive and affective relationships to their environment. The contrastive values that may be articulated and/or detected in vocal tract gestural activity are information that is specific to the affordances of that activity. The detection of such information has the functional capacity to change the agent's relationship to some aspect of its environment (Reed 1996a: 97). Moreover, the pick up of this information through the listener's exploratory activity of the highly structured energy field – the soliton-like wave – that acts on an excitable medium (air, body, brain) covaries with the agent's awareness of a change in the agent's relationship to the environment (see Reed 1996a: 98).

In my view, meaning, as distinct from value, is definable as just this reflexive awareness of the interpreted change that the detection of a given value brings about in the agent's (self-)awareness of its relationship to the relevant environment. The processes of mutual attunement described by Goldstein and Fowler (2003: 35-39) show how a given population of

languaging agents has evolved both biologically and culturally to be motivated to orient to and to relate to the specific affordances of vocal tract activity in particular culturally shaped and socially distributed ways that have specific values for that population (see Cowley 2008: 339-340; Section 15). Phonological values are, then, ecological facts, as defined by Reed (1996a: 100-103). They do not exist as a private mental code of stored phonological categories, but as normative, socially distributed constraints on vocal tract behaviour – behaviour that is public and observable. Infants are induced and rewarded by caregivers to seek out in their own and others' vocal tract activity the culturally salient phonological patterns that serve to promote culturally valued meanings and activities that may bring about modifications of agents' relationships to their environment.

19. The Second-order Reification of Vocal Tract Gestures as Speech Sound Segments

Mutual attunement as discussed above provides a plausible foundation, developmentally speaking, for the emergence of conventional resources that enable persons to solve the many coordination problems encountered by them in the different spheres of social life. Coordinated vocal tract gestural activity between persons in languaging behaviour requires a specific organization of this behaviour qua action systems that will tend to persist and recur at the population level as a consequence of the values they have for agents in the situations in which they serve to coordinate social interaction between persons. Phonological units and structures are, on this view, the result of the bonding of organ-specific gestures to form thermodynamically more stable action systems, or what I shall call *phonosemantic operators*, in utterance-activity between persons. Phonosemantic operators consist of vocal tract gestural scores in utterance-activity.

The term *phonosemantic operator* serves to remind us that (1) vocal tract gestural scores are not the arbitrary vehicles for the encoding of a separate, more abstract semantic or cognitive content by a sender and its subsequent decoding by a receiver; they are intrinsically meaningful (Section 4); (2) they function as operators on situations, their conventions and the understandings that persons have of the given situation and its associated conventions; (3) they are normatively constrained and culturally shaped and standardized resources that have the functional capacity to evoke situation conventions that enable them to solve coordination problems between persons in particular situations; (4) their conventionalization entails the typification of utterance-activity as differentiated kinds that are iterable across different situations and on different occasions such that they become abstracted from particular situations and standardized as cultural types through processes of 'grammaticalization'.

Iterability arises when writing systems serve to model the topological-continuous character of first-order dynamics in terms of discrete elements such as phonemes that are repeatable across speakers and occasions. One consequence of iterability in literate cultures is the standardization of speech sounds by means of institutionalized repetition in order to achieve error-free transmissibility as distinct from imitation of the topological-continuous aspects of speech (Harris 1991: 388). Harris explains this as follows:

> Transmissibility without error-compounding requires not only that the elements of the

> material to be transmitted be discrete, but also that they be pre-set in sender and receiver. When the speaker and the hearer are referring to a set of elements known to both, the hearer need receive only enough of a signal to distinguish a particular element – phoneme or word – in contrast to all other elements that could occur there. When the hearer then transmits (repeats) the utterance, he pronounces his own rendition of the pre-set (i.e. known) elements which he has distinguished. This means that both must learn to recognize a set of grammatical elements, primarily particular phonemes (or phonetic distinctions) and secondarily vocabulary (morphemes, words), in respect to which they speak and perceive utterances. It is this public institutionalization that makes the transmission of an utterance a repetition, whereas an attempt to redo or transmit something whose elements are continuous or not pre-set is an imitation.
>
> (Harris 1991: 388)

Pre-set elements are established in a population of speakers as norms of, for example, pronunciation and phonological spelling based on the presumption of the objective existence in the speech signal of segmental speech units, e.g. phonemes, which are taken as normative phonological units. Harris, in keeping with the overall formalist-structuralist premises of twentieth century linguistics, nonetheless preserves the idea that these segments are real features of speech sounds rather than the artefacts of a particular way of doing meta-language. He also assumes, incorrectly, in my view, that the continuous or non-discrete dimensions of vocal tract gestures are not also subject to processes of conventionalization. As Port (e.g. 2007: 351) shows, this view is a consequence of using the discrete, serially ordered graphic elements of alphabetic notation to transcribe and 'represent' the very different dynamics of talk as if the dynamical characteristics of talk can be attended to in terms of segments such as phonemes, syllables, and so on. The presumption by many linguists of the objective existence of segmental speech units belies the scientific evidence concerning the rapid perceptual learning that characterizes the infant's early encounters with the speech patterns of the ambient language (Port 2007: 353).

Infants learn to perceive and to model the very rapid patterns of co-articulated vocal tract gestures in ways that bear no resemblance to the much slower time-scales implied by the unnatural segmentation of co-articulated speech gestures into discrete grapheme-like segments, i.e. phonemes. Through statistical learning, infants achieve within the first year the ability to discriminate the salient articulatory and auditory patterns of the ambient language through constant habituation to the rapid rate of articulation of normal speech (Kuhl 2007). The mutual attunement of languaging agents in early language development to those regions of the vocal tract that afford partitioning into discrete regions and their repetition by speakers sets up an early perceptual focus on organ-specific contrasts that are amenable to their subsequent reification as segments on analogy with the graphic units of alphabetic notation.

Writing is a form of external cognitive and cultural scaffolding that affords the development of the skills of institutionalized reification of, attending to, repetition of, and transmission of sequences of discrete elements such as phonemes and syllables. These elements compress socially distributed and standardized information that enables languaging agents to recognize

determinate sequences of discrete elements as grammatical units such as words, phrases, sentences, and so on. Such units are repeatable and are, as Harris puts it, "uniquely appropriate for error-free transmission of utterances." (1991: 388-389).

Organ-specific vocal tract gestures are potentially contrastive. In the first instance, it seems likely that conventionalized proto-utterances or phonosemantic operators, consisting of holistic vocal tract gestural scores of co-articulated contrastive features, were associated with different aspects of experience through processes of statistical learning in a population. Organ-specific phonetic gestural contrasts were the basis of the discrimination or partitioning of the environment by increasingly standardized gestural scores into largely implicit representational topologies that formed the basis for the collective pooling of experience. Organ-specific contrasts proved amenable to reification as the basis on which co-articulated gestural scores were re-described on analogy with units of written notation as frozen segments of speech sounds. These segments form the basis of second-order phonological categories. The latter serve the meta-linguistic purpose of distinguishing between different classes of gestural score and subjecting these to increasing pressures of conventionalization. In this way, the reified aspects of standardized gestural scores and their distinctions are decomposed and then preserved as segmental distinctions in sound (e.g. pronunciation) that can be preserved under diverse conditions of transmission across persons and situations. Once discrete segmental distinctions emerge in this way, it follows that such distinctions, serially ordered in determinate sequences as norm replicating linguistic objects, would soon give rise to further meta-linguistic criteria for the segmentation of gestural scores into proto-word like objects that were defined according to the ways in which sound segments were combined. Moreover, the principle of decomposition at work here also means that these early word-like objects could be modified by the affixing of further standardized sequences of sound segments for the purpose of differentiating categories of experience such as time, plural, animate, gender, and so on.

Phonological forms, as we saw above, are socially distributed cultural types that constrain and shape vocal tract gestural activity. They provide standardized solutions to the problems of coordination that arise in emergent social situations. The conventionalization and standardization of phonosemantic operators leads to their re-categorisation as lexicogrammatical patterns (Sections 16, 25). The mutual attunement of speakers' vocal tract gestures afforded by the acoustic medium is an inherently and implicitly reflexive and self-organizing process in which persons seek to coordinate their diverse, not always congruent or harmonious, points of view. The achievement of interpersonal parity that is consequent upon mutual attunement in language development means that persons become increasingly sensitized to the context-dependent nature of different utterance-types qua phonosemantic operators and therefore to the ways in which utterances can have different effects and consequences, depending on the context. This goes hand-in-hand with the increasing understanding that different utterance-types have the power to evoke, operate on, and transform an increasing diversity and range of social situations.

In the first instance, utterances operate on implicit situation conventions and their reflexivities

that are social, but not linguistic. This is what infant protolanguage (Halliday 1975) (cf. Reed's 1996a: 167 "indicational language") achieves, i.e. the capacity to act upon and to transform participants' awareness and understanding of aspects of the non-linguistic situation that are the focus of proto-linguistic topics or arguments in, for example, joint attention-sharing routines. However, the point is not confined to infant protolanguage: predicational language also operates on non-linguistic social realities. The re-categorisation of first-order phonosemantic operators as second-order phonological and lexicogrammatical patterns and types is a further normative emergence such that second-order language is constituted not only as a means for operating on non-linguistic social conventions, but also as a means of conventionally creating and constituting social situations and realities through the grammaticalized resources of predication in second-order language. Moreover, if we see grammar not as a further level of formal linguistic encoding/decoding, we can better understand it as a further layer of reflexivity -- a fractal meta-linguistic process that views and refracts first-order languaging behaviour through the prism of second-order cultural constructs and their (always partial) perspectives. In other words, 'grammar' qua meta-language provides partial models of and perspectives on the shape of first-order languaging dynamics, usually in the service of particular social, institutional and cultural projects.

The enforced reification of segments through formal and informal teaching and learning in a given population sets up the possibility of predication. In the first instance, the selective analogizing of phonological contrasts associated with discrete organ-specific vocal tract gestures to discrete graphic notational units sets up the premises for the reification of the former as discrete segments of sound (quite contrary to the empirically determined evidence of the spatiotemporal overlap of vocal tract gestures in co-articulation). The resulting reified segments are, then, objectified as formal properties of the articulatory-auditory dynamics of speaking and listening in languaging behaviour that fulfil the criteria of error-free repetition and transmissibility discussed above. They therefore serve as normative constraints on how speakers are induced to perceive and articulate the far more complex dynamical properties of acts of speaking. In this sense, speech sound segments (phonemes, syllables) and their graphic analogues in written notation are operators that can operate on first-order languaging dynamics so as to transform aspects of the dynamics into recognizable serial orderings of static, non-overlapping, and context-independent phonemes or phonological distinctions that are, in turn, re-categorised as grammatical words, phrases, sentences, etc. The latter are abstracted away from the dynamics such that they are amenable to their transcription into written notation.

Importantly, these serial orderings of phonemes therefore take on a grammatical interpretation precisely because the serial ordering of phoneme segments corresponds to a grammatical meaning. Because the reified segments are conventional (normative) constraints that operate on first-order dynamics, we can say that the process of composing and generating serially ordered strings of such segments is productive. This is so in the sense that an unbounded set of strings or serially ordered sequences of phonemes can generate the conditions in which second-order language emerges as a conventionalized means of operating on and transforming first-order dynamics. At this point, second-order language emerges as a

conventional cultural affordance for operating on, structuring, and transforming (1) non-linguistic social conventions in situations; and (2) language itself as a conventionalized resource for operating on, structuring, and transforming itself qua system of social conventions (see Bickhard 2004) in ways that can transcend situation-bound first-order dynamics (Linell 2009).

20. Operator-Argument Dependency Relations

Once hierarchically organized serial orderings of discrete, de-contextualized formal objects are postulated as the basis of second-order language and its composition, then specialized notation systems such as writing systems, formal grammar, logic, and computer languages can be shown to have emerged as further technological and cultural elaborations of this basic principle, i.e. that language is a second-order 'code' consisting of discrete, standardized, serially ordered formal tokens on which operations of symbolic manipulation are performed. In this way, the operator-argument relation, rather than co-occurrence relations or probability of co-occurrence per se, defines the various ways in which formal tokens are organized in dependency relations such that an operator word, say, does not occur in a sentence unless a word of its argument set also occurs (Harris 1991: 332). The dependency relation expresses a meaningfulness that is more than the sum of the individual words that are combined in the dependency relation. Operators are dependent on their arguments and occur only in the presence of their argument words. Operators predicate something about the argument, i.e. they say something about it. The operator-argument dependency relation in its many forms is, then, the basis of grammatical units such as the clause. The dependency relations formed on the basis of co-occurrence relations of words yield a new kind of grammatical unit, the clause, based on the meaning-relation called predication. In this way, the clause constitutes, linguistically speaking, a unit of information about some referent situation that can be asserted, claimed, denied, argued about, refuted, and so on (Halliday 2004 [1985]: 106-121). Dependency relations are not confined to the clause and encompass morpheme- to discourse-level relations. Some examples follow:

1. The morpheme suffix *–er* in *teacher* is not just a suffix indicating the meaning 'agency', but an operator on the base morpheme *teach* such that the latter is transformed into an argument of further predication, e.g. the nominalization of the base morpheme Process. Thus, the nominalization specifies an agent that can function as the argument of a further predication, as in the clause *The teacher taught the students English*;

2. The adjective *phonetic* in the nominal group *phonetic symbol* is an operator on its argument, i.e. *symbol*. The operator predicates of the semantic class of Thing specified by *symbol* a particular sub-classification according to type;

3. In the clause, *The teacher taught the students English*, the Subject, *The teacher* and

the Finite [TEACH + PAST] are operators on the argument (predication) of the clause, i.e. *TEACH + the students English*. The Subject operator ties the argument (predication) to a deictically specified third-person non-speech participant, i.e. the referential noun phrase, *The teacher*. In other words, the predication is tied to and is about the Subject, which is contextually-sensitive to the local requirement that the Subject is correctly associated with its referent and that the referent is retrievable from the relevant local representational topology context (its reference is locally resolvable or interpretable). In other words, the entity specified by the Subject is immediately accessible to the interactants (first and second person) in the interaction (Davidse 1997). The Finite is marked for tense: it temporally grounds the argument in past time with respect to the here-now grounding of the utterance. The two operators therefore operate on and transform the argument of the clause into a fully grounded proposition, i.e. a particular dependence relation among the grammatical units composing the clause that constitute a unit of information founded on the predication relation, e.g. the Finite operator PAST operates on and transforms the (abstract) argument *The teacher TEACH the students English* into a unit of information that does not directly reflect any given real-world situation though it does relate to what we know and experience about English teachers and their students irrespective of the clause we have before us. Rather, the clause has the capacity to assert information about a situation – information that can be taken up, responded to, and transformed by the speaker or by others in various ways in the form of further operations on it. In this minimal sense, the clause has the capacity to operate on a given situation and to transform the situation and participants' understandings of the situation. It is this predicational capacity of language to operate on and to transform both non-linguistic **and** linguistic social realities that affords the emergence of virtual cultural entities that are constituted by means of the predicational capacity of language to operate on and to constitute virtual second-order social and cultural realities that exist only in off-line non-perceptual experience.

21. Languaging as Recursively Self-Maintenant Process

Languaging behaviour is a recursively self-maintenant process (Bickhard 2004, 2011; Thibault 2005a). It consists of a universe of objects that enter into constructive relations. As the examples show, albeit informally, a given trajectory through the object-space has the capacity to induce self-maintenant structures that are characterized by invariant patterns of transformation. Buss and Fontana (1996: 20) further point out that self-maintenant systems of this kind possess properties such as "regeneration, structure-dependent extension, complex substructure, capacity for hierarchical nesting" – properties that are characteristics of living systems (see also Abler 1989). The imposition of different boundary conditions on the object-space in the form of different meta-linguistic criteria generates different levels of organization, and a diversity of organizations within each level. Table 3 sets out a possible way of thinking about a flat ontology of languaging behaviour and the role of different fractal scales of meta-linguistic constraints. This can be compared to the more usual levels account in Table 4.

In Table 3, Level 0 is defined by self-copying expressions or simple ensembles of copying expressions, i.e. organ-specific vocal tract gestures. Level 1 refers to self-maintaining organizations of utterances (gestural scores) composed of Level 0 organizations. Level 2 is defined by self-maintaining meta-organizations (e.g. second-order lexicogrammar) composed of Level 1 organizations composed of Level 0 organizations.

Time scale		**Constraints**	**Relationship between Scales**	
Cultural-historical	Lexicogrammatical forms and discourse categories and patterns as second-order re-descriptions of standardized phonology and combinations as words, combinations of words, etc.	Quasi-causal constraints and attractors from cultural-time-scales: population-level dynamics		
Cultural-historical	Low-dimensional phonological forms as population-level socially distributed and normative standardizations of phonetic gestures influenced by pronunciation norms and writing	Quasi-causal constraints and attractors from cultural time-scales: population-level dynamics	Time-invariant, contextually-invariant, low-dimension (sparse)	Level 2: self-maintaining meta-organizations composed of Level 1 organizations composed of Level 0 organizations
Real-time events between persons	Rich high-dimensional phonetic gestures of first-order languaging behaviour	Real-time bio-physical processes coordinating persons-in-interaction	Time-varying, contextually-variant, high-dimensional (rich)	Level 1: self-maintaining organizations of utterances composed of Level 0 organizations
	Organ-specific phonetic gestures as primitives; the smallest kinetically persistent and self-maintaining entities of the phonological repertoire of types specific to a given language	Molecular constellations formed by combinations or syntheses of organ-specific gestures as kinetically persistent phonetic organizations		Level 0: simple ensembles of self-copying expressions

Table 3: A flat ontology of languaging behaviour and constraints on different scales

Language Strata: Hjelmslev	**Language Strata: Halliday**	**Relationship between strata**	
Content substance	semantics		
Content form	lexicogrammar		
Expression form	phonology	Abstract phonological categories and forms and their combinations	Phonological form realizes/is realized by
Expression substance	phonetics		

Table 4: A levels account: Stratification of language according to Hjelmslev and Halliday

Infant protolanguage (Halliday 1975; c.f. Reed's 1996a: 158-161 indicational language) operates on a more restricted range of conditions in the immediate situation: non-predicational or proto-linguistic utterances indicate topics or arguments without explicitly predicating something about the argument. Topics typically indicate objects, properties, events in/of situation and function to make the infant's interlocutor aware of the infant's focus of interest or attention; elementary meanings operate on situations and transform them in this way. Predicational language means that the operator is something that is said about or predicated of its argument. The operator does not occur in a clause or sentence except in relation to a particular class of argument words that are also present (Harris 1991: 332). In other words, the operator can select amongst a potentially large set of potential arguments about which it can predicate something according to the demands of a wide variety of actual and anticipated situations. This capacity means that languaging behaviour is recursively self-maintenant because it can vary how it contributes to the condition of self-maintenance of languaging agents when they engage in such behaviour. The morpheme suffix -ER in *teacher* (see above) already shows this principle at work. As we saw, -ER is an operator on its base (TEACH); it recursively operates on the base and predicates something of it. The resulting nominal is a normative emergence that itself can function as an argument in ways that can be recursively operated on by multiple potential operators in response to a wide range of environmental conditions and situations that may or may not be supported by the given environment (Bickhard 2004; Freeman 2000: 97; Section 8). Grammatical forms do not encode aspects of the environments in which they occur; they specify linguistically constituted information that can partition or differentiate the environment in ways that may or may not be judged as appropriate, correct, truthful, and so on, in the situation.

As Bickhard (2004) has shown, the interaction may or may not be supported by the given environment. If it is, it will succeed; if not, it will fail. Interactional success and failure are judged with respect to the anticipated outcomes – what could or should happen – given the situation. These outcomes are available to agents as feedback that can guide future learning. On the other hand, the judging of linguistically constituted information that is asserted about some situation as appropriate, truthful, believable, correct, sincere, etc. with respect to some current locus of cognitive or perceptual processing in the situation gives rise to a normative

notion of linguistic information. Information in this sense is concerned with what is predicated about the given locus of cognitive or perceptual processing qua linguistically constituted argument (see Allen and Bickhard 2011: 109). Languaging behaviour is one means whereby persons can vary how they contribute to their own condition of self-maintenant being-in-the-world. Language evolved above all as a further extension and augmentation of the possibilities afforded by the emergence of consciousness in many living species, i.e. the optimization and management of the life process of the self and the self's responses to its environment (Damasio 2010: 267). Damasio makes the following pertinent observation concerning the internal images that provide more precise information about the environment of the self:

> The lion's share of the advantage, I suspect, comes from the fact that in a conscious mind the processing of environmental images is *oriented* by a particular set of internal images, those of the subject's living organism as represented in the self. The self focuses the mind process, it imbues the adventure of encountering other objects and events with a motivation, it infuses the exploration of the world outside the brain with a *concern* for the first and foremost problem facing the organism: the successful regulation of life. That concern is naturally generated by the self process, whose foundation lies in bodily feelings, primordial and modified. The spontaneously, intrinsically feeling self signals directly, as a result of the valence and intensity of its affective states, the degree of concern and need that are present at every moment.
>
> (Damasio 2010: 267-268; italics in original)

Language is a socially and culturally distributed resource for maintaining the socio-cultural homeostasis of individuals and social groups. Many theories of language emphasise above all its relationship to 'mind.' Damasio's focus on the self and its role in focusing the mind process draws attention to the way in which languaging behaviour orients selves in their worlds in value-weighted ways. I have already argued that lexicogrammatical differentiators are normative, value-weighted patterns that attract first-order languaging dynamics to their basins of attraction. They enable selves to orient to normative cultural objects, events, etc. – actual and virtual – that are productively differentiated by lexicogrammatical differentiators. By the same token, they also relate selves and their actions to norms. Linguistic norms therefore have to do with the relationship of the norm both to its objects and to its subjects (selves) (see Macherey 2009: 74).

As Allen and Bickhard point out, self-maintenant systems, more generally speaking, have a normative stake in maintaining themselves (Allen and Bickhard 2011: 108). Languaging behaviour cannot be separated from languaging agents and treated as disembodied 'text' (Section 1) without jettisoning any prospect of explaining the language-person system as a class of recursively self-maintenant system in the sense defined here. In this perspective, second-order language is a culturally distributed form of 'global order' that is emergent from the local order of self-organizing first-order languaging dynamics. By the same token, the emergent organization of second-order language is a further normative emergence such that lexicogrammatical differentiators in utterance-activity specify and predicate information

about (aspects of) situations that may or may not be true or false, etc., as discussed above.

Dialogically coordinated interactional behaviour between persons elicits interpretation (Cowley 2011). Vocal tract and other gestures (facial expression, hand-arm gestures) compress information on account of the dynamical co-articulated properties. Information is used here in Gibson's (1986 [1979]) sense. Information is specificational. Thus, the information in vocal tract gestures specifies something about both the person articulating the gesture (propriospecific) and about some aspect of the situation and the person's orientation to the situation (exterospecific). Information is not the same as meaning. Meanings, on the other hand, are created by the unique, experience-grounded neural constructions of each individual person. Vocal tract and other gestures elicit and prompt for interpretation in the situations in which they occur; they do not contain or encode meanings or serve to convey or transmit these meanings from one person to another. When we concert our meanings with others in dialogically coordinated languaging activity, we make an effort, to varying degrees and with varying levels of motivation and commitment, to assimilate the meanings we interpret in others' gestures to our own viewpoints, as constructed and articulated by own bodily and neural dynamics (Freeman 2000: 15).

Seen in this light, the second-order lexicogrammatical patterns that we detect in the ambient energy fields created by vocal tract and other gestures are forms of cultural scaffolding to which interactants orient in order to augment the assimilation of each others' meanings in ways that may lead to, though they do not guarantee, common understanding. Grammar is a normative emergence in the sense discussed above that serves to solve coordination problems of precisely this kind. The normative character of lexicogrammatical constraints on languaging behaviour also serves another important purpose. The very rapid time-scales of first-order languaging activity simply are not fully available, perceptually speaking, to interactants. We do not have full awareness of the complexities of languaging behaviour, though they are usually under voluntary control and are intentional. Many aspects of the dynamics of languaging behaviour are rapid, continuous and not accessible to conscious monitoring (see also Port 2007: 352). However, we obviously do consciously attend to the articulatory and auditory properties and processes of vocal tract behaviour. We can also recall aspects of these.

Attending to and recalling vocal tract behaviour is a skill that can be improved though what Gibson (1983 [1966]) called the 'education of perception.' We are able to sample and bring to awareness for one's self and for others selected aspects of the process. The normative character of lexicogrammar thus provides abstraction amenable cultural resources for selectively sampling some aspects of the far more complex processes involved in languaging behaviour, e.g. when we use standard orthography to write down what someone has said. Such samples are second-order textual records of the primary data (Thibault 1994). Unfortunately, many scientific accounts conflate the primary (first-order dynamics) data with just such second-order samples and develop their theories accordingly (Section 1). Samples serve to bring some aspects of the process to awareness for particular, limited purposes. Moreover, these samples themselves can be embedded in and operated on in further acts of

interaction between persons.

Linguistic catalysis is, of course, a dialogical activity between persons. The capacity of the face, vocal tract, and hand-arm systems to differentiate very many fine-grained sensorimotor discriminations and their productive possibilities of combination and recombination meant that these body systems and their synergies came to be used for the selective orientation to and exploration of the environment both in collaboration with other perception-action systems and in concert with the other persons with whom one interacted. The increasingly delicate and discriminatory patterns of difference that could be articulated through these bodily resources meant that an increasingly wide diversity of meanings could be catalysed in relation to a widening range of situations. In the case of dialogically coordinated linguistic catalysis between persons, the boundary conditions include the higher-scalar ecosocial arrangements that constrain dialogically coordinated languaging behaviour between persons. Meanings emerge as a result of the interactions between the neural and bodily dynamics of individual persons and their ecosocial environments.

The emergence of higher-scalar boundary conditions entrains these dynamics to cultural patterns such that the vocal tract and other sensorimotor discriminations of a population of interacting agents are constrained by the same attractors in ways that enable the patterns of difference to be taken up and adaptively modified by others. In this way, individuals get linked to each other across time and space. These emergent patterns of differentiations are not encodings of meanings, ideas, or thoughts that are then transmitted to others, who in turn decode them (Section 4). Instead, these fine-grained discriminations and their possibilities of productive combination and recombination extend and augment the capacities of agents to explore, act on, interact with, experiment with, and analyse the emergent possibilities of their worlds (Thibault 2004a: 187).

In exploring their world in this way, agents assimilate their brain and body dynamics to it (Freeman 2000: 141). They do so by creating resonance patterns between their neural and bodily dynamics and selected aspects of their worlds through the discriminatory possibilities of this exploratory activity. As Freeman (op. cit.) points out, such activity is hypothesis testing – it seeks responses from the world that confirm or disconfirm its hypotheses, as also suggested by the notion of interactive success and failure discussed above. Languaging agents adapt to and assimilate to each other's dynamics through reciprocal exploratory activity and hypothesis testing (Freeman 2000: 141). In doing so, they "create transcendent social entities that enhance and empower the individuals." (Freeman 2000: 142).

Hypothesis testing is not based on encoding/decoding, but on the very different principle of differentiation. The many fine-grained differentiations and their productive combinations made by the face, vocal tract, and hand-arm systems are internal to the organism. In exploring and testing their environments, languaging agents interact with their environments. Differentiations do not encode/decode a meaning or content that is external to the individual (to the means of encoding??). Rather, through its interactions with its world, the agent differentiates or partitions the world by interacting with it and in the process selectively applying its own internal capacity for making fine-grained discriminations, e.g. in phonetic

gestures, in ways that can be tested against the environment itself as a differentiation that is appropriate, truthful, correct, relevant, and so on. If the agent-internal differentiation is of a kind that is appropriate to the given environment, it will be selected and confirmed by the relevant environment as one that is successful, appropriate, correct, and so on. If not, the differentiation will be de-selected and disconfirmed as unsuccessful, inappropriate, incorrect, and so on, and therefore judged to be unsuccessful, unwarranted, invalid, and so on.

The 'content' of a differentiation is not, then, arbitrarily and externally attached to a material sign-vehicle. In Halliday (2004) we saw that the content is an ecosocial experience that is coded by the coding of the semantics by lexicogrammar (Section 7). On this reading, the content – experience or social relation – derives from the external ecosocial environment and is explicitly encoded by the semantic categories of the language. Differentiations don't work like this. Phonetic gestures, for instance, are not the means for encoding/decoding experience as semantic content. Instead, they specify an environment that is presupposed as being one that is appropriate (or not) for the indicated differentiation. The phonetic gesture and its differentiations therefore specify the future-oriented interactive potentiality of the environment. Insofar as it is a hypothesis-testing act, a given phonetic gesture qua complex, multidimensional differentiator anticipates interaction possibilities that the phonetic gesture presupposes and explores and which the given environment may or may not support. The phonetic or other gesture is an explicit organization qua differentiator of interactive potential. It implicitly presupposes the contextual conditions that would support the given differentiator. By the same token, the given environment may not support these interaction potentialities such that the differentiation fails in the given environment.