Advancing and Validating Models of Cognitive Architecture

Carl Vogel
Trinity Centre for Computing and Language Studies
Trinity College Dublin, the University of Dublin
Dublin, 2, Ireland
Email: vogel@tcd.ie

Anna Esposito
Università degli Studi della Campania "Luigi Vanvitelli"
Caserta, Italy
and
International Institute for Advanced Scientific Studies (IIASS)
Vietri sul Mare, Italy
Email: iiass.annaesp@tin.it

***Abstract*—We present a methodology for proposing and evaluating components of cognitive architectures comprising complementary approaches: hypothesizing and behavioral mechanisms, evaluating the aptness of those mechanisms in providing accounts of human behaviors; embed models of hypothesized mechanisms within simulation systems and observe synthesized model behaviors. We illustrate these theoretical approaches with examples.**

I. INTRODUCTION

The demand for sophisticated and functional man-machine interfaces originates from emergent social needs, such as aging of the world population and the consequent difficulty to maintain social welfare because of the cost associated with human labor. The achievement of human level automaton intelligence raises the exigency for more precise approaches for advancing and validating models of cognitive architectures and pose the following imperative challenges:

1) Identify pre-processing algorithms able to capture invariant features from multimodal social signals;
2) Infer simple and fast computational models able to detect or classify with an approximation comparable to humans features for the maintenance of objects hierarchically structured, time dependent and reciprocally connected through complex relations (such as a set of complex emotional feelings).

Robotic architectures may be designed for effectiveness with respect to task-related criteria without reference to theories of human architectures for cognition and interaction. However, an important strand within cognitive science that attends to simulated agents is exactly the efficacy the simulations exhibit as models of human cognition and interaction. We highlight advantages of pursuing a research agenda that explores human cognitive architecture while attending to simulated agents; some of the knowledge that results transfers usefully to task oriented robotic systems. One may subscribe to any of a number of available principled inventories or desiderata for cognitive architecture [1], [2], independently of whether one adopts the strategic directions proposed here for validating theories of cognitive architecture. The directions are empirical in providing a focus for collecting positive and negative evidence regarding hypotheses.

We note two complementary directions of research that yield useful information in this context. One approach hypothesizes behavioral mechanisms and identifies the extent to which those mechanisms provide apt descriptions of human behavior directly. A complementary approach locates such mechanisms in simulation systems, and provides evaluation of the extent to which the emergent features of those simulation systems match features of human communication and cognition. Of course, both of these approaches presuppose independent exploration from the perspective of cognitive psychology and linguistic analysis of the properties of human cognition and communication. In the present work, we unify the two directions by addressing a single primary behavioural mechanism (symmetry identification), both exploring the role of the mechanism within human language processing and in simulation systems.

In first direction, specific mechanisms are analyzed as hypothetical cognitive operations (for example, symmetry identification), and the locus of such mechanisms in a host of intelligent behaviors is identified. In the case of symmetry, this is highlighted within both visual processing and language processing. For example, general purpose abilities, such as symmetry perception, are robust across many forms of symmetry, such as copy identification, mirrored copy recognition, etc. [3]. Symmetry perception abilities have been invoked to explain the divergence between the computational complexity predictions made on the basis of asymptotic worst-case complexity of grammaticality assessment associated with classes of fully abstract formal languages [4] and actual human behaviors [5], [6], [7]. Symmetry production behaviors are also involved in dialogue synchrony [8], where repetition of self and of others in dialogue has been analyzed as a means for indexing interpersonal engagement in dialogue as well as correlating with proxy measures of mutual understanding, at least in task-based dialogues [9], [10], [11].

In the other direction, simulation systems are configured and analyzed with reference to the degree of fit between properties of emergent interactions and properties of, for example, linguistic interactions among humans or languages in general. For example, language evolution simulations have been constructed, with agents operating in idealized versions of the external world (including pessimistic assumptions about the degree to which agents may have a shared perspective on the world), with the history of interaction behaviors yielding systematic properties that can be evaluated in relation to how well they map onto properties of human interactions [12],

[13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. Accordingly as the system parameters yield interaction systems that either fit or do not fit with human interactions or human languages, one may find support or falsification of theories of the key parameters and configurations that gave rise to human interaction as it is.

The span of research into the underlying linguistic and psychological facts about human cognition and interaction is too vast to encompass. However, in the context of cognitive architectures for robotics, it is important to dwell on recent research with focus on sensory modalities in isolation and in their fusion [24], [25], [26], [27]. This is important because while there are biological mechanisms that support information fusion from distinct modalities, there is ample evidence for gender based and culture based divergences in processing visual and aural cues [28], [29], [30]. This entails that autonomous robotic systems developed for the purpose of embodying theories of human cognition and interaction must also have a role for learning in the fusion of input signals [31], [32]. On the other hand, we accept lessons from cognitive neuroscience in the form of argument that embodiment and grounding are not essential to progress in understanding the neural substrates of human cognitive architecture [33].

Research in these complementary directions, and the underpinning direct elucidation of properties of human cognition and communication must continue. Research successes can be highlighted with examples. Example successes of this approach include apparent support for some basic theoretical assumptions in applied linguistics, such as the necessity of Pinker's assumption that humans must be endowed with an innate notion of predicate argument structure, such as that assumed by lexical functional grammar, on which human language acquisition during childhood development is bootstrapped [34]. The evident support for this hypothesis follows from a surprising range of simulation paradigms in which some version of this hypothesis is essentially hard-wired. Such examples provide argument that the research method endorsed here continues to yield knowledge about cognitive architectures that may advance the state of the art in robotics.

II. ADAPTABLE FUNCTIONS

A robust programme of research is set by the cognitive faculty approach to psychology. The modularity of mind hypothesis [35] posits a number of special purpose input modules. Functional specification of modules, in this sense, include speed of processing, involuntary processing, information encapsulation, and so on. Given functional specification, a research agenda is set in relation to determining the input-output properties that individuate faculties and discovery of faculty-internal processes that determine the input-output properties.

Horizontal faculties (perhaps, memory) are not content specific; for example, one might argue that there is no specific faculty of event memory as distinct from object memory. This does not, as Fodor notes, entail that for any individual memory performance is identical across domains. Demands within any domains may differentially require other horizontal faculties. For example, memory of music may well be more robust for some individuals than memory of utterances. In contrast, vertical faculties [35, p. 21], such as vision, are taken to be specific to functional domains, genetically endowed, located in distinct neural substrate and are computationally autonomous (do not share or compete for resources like memory, processing, intelligence, attention, etc.). Such a model is arguably compatible with the subsumption architecture in robot design inasmuch this model argues against "shared memory between computational elements" [36, p. 3], given mutual entailment of information encapsulation.

	Horizontal	Vertical
Resource Competition	Yes	No
Neural Specificity	No	Yes
Information Encapsulation	No	Yes
Putative Example	Memory	Vision

TABLE I. SALIENT PROPERTIES OF HORIZONTAL AND VERTICAL COGNITIVE FACULTIES

Noting that Fodor does not suppose human cognitive architectures to be best explained by either a horizontal or vertical inventory of faculties exclusively, a synthesis of the horizontal and vertical models of cognitive faculties may also be explored in consistency with this paradigm. The resulting models that mix features of horizontals and verticals may be thought of as diagnoals. These diagonals may involve functions that operate across domains, but with differential importance in each, for example – sound perception associated with language, music, or scene monitoring of predators. Evidence exists that concepts associated with functions afforded by objects (e.g. "decorate self", in relation to, for instance, a necklace or comb) may demonstrate locality effects in neural regions distinct from those associated with the artifacts that may be manipulated to achieve the functions [37]; however, the separability of neural substrate involved in recognizing function from that involved in manipulating objects does not entail that reasoning about function and reasoning about object manipulation constitute distinct vertical faculties. Further, diagonal inventories of faculties may involve information encapsulation according to domain (perhaps via location registration of memory in neural substrate), but also sensitivity to competition for computational resources with other processes. It is constructive to dwell further on possible properties of the diagonals in the analysis of cognitive faculties. The next section addresses a possible candidate for a diagonal in symmetry perception and learning.

III. SYMMETRY

Copy identification – for example, recognizing two objects as of the same type, or a sequence repetition – is a special case of symmetry awareness [3]. The copy facility is a function that cuts across desiderata for cognitive architectures [2] – perception as one and declarative and procedural memory, as another, internal simulation as a third. The copy identification function has been invoked specifically in the processing of language [6]. [6] addressed predictions of formal language theory that certain natural languages should engage for their native speakers heavier cognitive processing loads and error rates than those required by native speakers of syntactically less complex languages. These predictions, however, failed to be demonstrated under empirical testing [5]. The formal language theoretic prediction is that natural languages that require more than context free sensitivity [38], [39], [40], [41] should require more resources than languages that require only context-free expressivity. The argument of [6] reconciled the difference between the analytical complexity predictions and empirical facts with appeal processing functions that are orthogonal to the usual formal language theoretic constructs [4].

The essential property of context-free expressivity is in nested dependencies, arbitrarily deep – essentially, recursive center embedding – and bracket-matching and string reversals are canonical examples of these: see (1).[1] An element of this language ($ababbaba$; where $w = abab$ and $w^r = baba$) is shown with its nested dependencies highlighted in (2); the same string is used to highlight the nested dependencies using indexed bracketing in (3). More complexity is required by cross serial dependencies in natural language, and a canonical string-set idealization of these is the string copy languages: see (4), (5) and (6), for the string $abababab$ within the language given by (4), (where $w = abab$). A semantic version of these are exemplified by the "respectively" constructions of English. However, the relevant processing facts relate not to semantic dependencies, but syntactic (and morphosyntactic) constraints, such that relevant candidate sentences making use of the context-free or mildly context sensitive embeddings (to arbitrary depth) are either grammatical or not according to whether the constraints are satisfied (and not just bearing unintended interpretations, but corresponding to grammatical sentences).[2] Natural languages, whether essentially context-free or indexed in expressivity, contain constructions additional to those that are homomorphic to the canonical string-sets described in these examples. However, the fact that languages contain sentences other than those homomorphic to the constructions that pick out key expressive properties of the language family does not diminish the need to account for these canonical constructions.

(1) $\{ww^r|w \in \{a,b\}^*\}$

(2)
```
ababbaba
||||||||
|||ii|||
||j--j||
|k----k|
l------l
```

(3) $({}_l a({}_k b({}_j a({}_i bb)_i a)_j b)_k a)_l$

(4) $\{ww|w \in \{a,b\}^*\}$

(5) $({}_l a({}_k b({}_j a({}_i ba)_l b)_k a)_j b)_i$

(6)
```
abababab
||||||||
l---l|||
 ||| |||
 k---k||
  ||  ||
  j---j|
   |   |
   i---i
```

The conclusion of [5] on the basis of the empirical data comparing comprehension error rates among speakers of Dutch and German on nested subordinate clauses is that the push-down automaton cannot be the functional basis of human language processing architecture. The hypothesis of [6] is that the architecture must be supplemented by a specific meta-grammatical process: *expect a copy*. The capacity to recognize that a sequence is repeated is related to the *expect a copy* process, in that being able to recognize that a copy of a sequence is necessary to being able to satisfy the expectation that a copy of the sequence will occur. This may also be generalized with local complementation operators: if the local complement of an a is an e and if a local complement of a b is an f, then the expectation of a complemented copy of the sequence $aabbb$ is satisfied by the sequence $eefff$. In the case of linguistic phenomena, the local complement of an accusative-marked nominal may be an accusative-marking verb. If applied at a suitable level of abstraction and with access to category-relevant complementation (for example, whereby a dative-marked nominal category matches a dative-marking verbal category), this operation can supplement context-free grammars (or regular grammars) to recognize mildly context-free constructions. Use of this *expect a copy* process then explains the empirical linguistic processing facts revealed by [5] that are otherwise difficult to reconcile with the processing complexity facts associated with fully abstract families of languages using the terms of formal language theory [4]: processing the reversal languages (1) is intuitively harder than processing the copy languages (4).[3] Further, this is taken to be a language specific adaptation of a general purpose cognitive function in symmetry recognition. The *expect a copy* function is one that we hypothesize as a diagonal faculty: on that we imagine to operate across functional domains, but without information encapsulation. Additional to the known empirical facts that the function can be used to explain, empirical predictions follow: for example, as noted by [6], finding differential "string-copy disfluencies" among speakers of natural languages that have robust cross-serial dependencies in relation to speakers of languages that do not have such constructions. Independent appeal to metagrammatical processes in relation to other syntactic phenomena provide indirect support for this approach [43], [44]. Another way to think of metagrammatical approaches is as gerrymandering the Chomsky hierarchy in a manner that yields regions which may provide more apt characterization of the expressive requirements of natural language syntax than that made available by the levels of expressivity that define the Chomsky hierarchy.[4]

IV. SIMULATION SYSTEMS

Evolutionary accounts of language development among first users is a necessary complement to the emphasis on development in cognitive architectures that characterize altricial species [2], [45], even if it is not the case that ontogeny replicates phylogeny. It is necessary to explain why natural

---

[1] In this example, the r operator denotes the reversal of the string to which it is applied. In this case, w is an arbitrary sequence of symbols, a or b (in any order).

[2] For example, the fact that Swiss-German requires more than context-free expressivity can be idealized by $\{w|w = a^i b^j c d e^i f^j g, i,j \geq 0\}$, where the i iterations of a correspond to accusative-marked nominals and of e, to matched accusative marking verbs, while the j iterations of b correspond to dative-marked nominals and of f, to dative-marking verbs – if the iterations of dative nominals do not match the dative marking verbs or if the iterations of accusative nominals do not match the accusative marking verbs, the corresponding sentences are ungrammatical and not merely non-sensical [40].

[3] The context-free reversals are explained as harder than the mildly context-sensitive copies by virtue of requiring an additional meta-linguistic process (reversal) to apply before checking the copy. However, a review by [42] notes that not all experimental tests meet this robust prior expectation.

[4] However, the resulting gerrymandered regions may not support closure under union, intersection with regular sets, etc., as define fully abstract families of languages.

language processing is a function managed within the human cognitive architecture, and as part of this, it is necessary to determine what gives rise to the distinctive properties of natural languages. Natural language is less efficient than smell or telepathy as a means of communicating mental states. Natural language is less efficient than telepathy in communicating narrative histories or future oriented plans. This is because ambiguity in natural language is rife. While natural language is an excellent medium for thought, it is not obvious as a tool for effective communication, apart from the nuances of ambiguity that it positively enables, such as deception or vagueness. Therefore, it becomes important to model the first steps of natural use and to assess the parameters of evolution in terms of how quickly stable (if ambiguous) conventions of use emerge. If natural language was an effective innovation in communication or entertainment, it must have proven so very quickly, else it is extremely unlikely to have endured as it has. These considerations have influenced, a programme of research that has been developed around the language evolution workbench (LEW) for modelling interacting individuals and groups of agents as they develop symbolic communication systems [46], [47], [21], [22], [23], [48].

The LEW platform enables experimentation with pessimistic assumptions about early communication. In particular, the system does not require the assumption that early communications were successful in achieving positive transfer of information nor positive documentation of information sharing (cf. [49], [34], [15]).[5] On the contrary, presupposing agents able to verify information sharing presupposes perfect communication, effectively telepathy, to begin with, and in this situation, it is difficult to see why natural language would emerge for the purpose of communication if communication were already perfect by other channels. The LEW platform enables specification of values for a range of parameters relevant to natural language: number of agents, group structures, number of discriminable phonemes, number of distinct event types, the degree to which agent memory may be deemed imperfect, feedback potential, and so on. The model of communication assumes that agents share joint access to (models of) events that occur in the world, but allows that they may have distinct perspectives on those events. The model assumes that speakers construct utterances that label each component of the event as they construe it, and that hearers hear speakers' utterances without noise, but that they may segment speakers' phoneme sequences into meaning bearing units distinctly to the way speakers bundle phonemes into meaning bearing units. Hearers attach meanings to each segment of the utterance, generally anchored in the hearer's perception of the shared event. There is the additional possibility of innovation in terms of a hearer interpreting parts of an utterance in relation to hearer memory of past associations between phoneme sequences and units of meaning. If feedback is enabled, this models task-based dialogue in which speaker and hearer may detect that their communication has been successful or not, but if not, without any indication of the components of the communication that were misunderstood. The experimenter may probe the system to measure degree of understanding achieved among interlocutors. It may be evident that there is ample room for miscommunication.

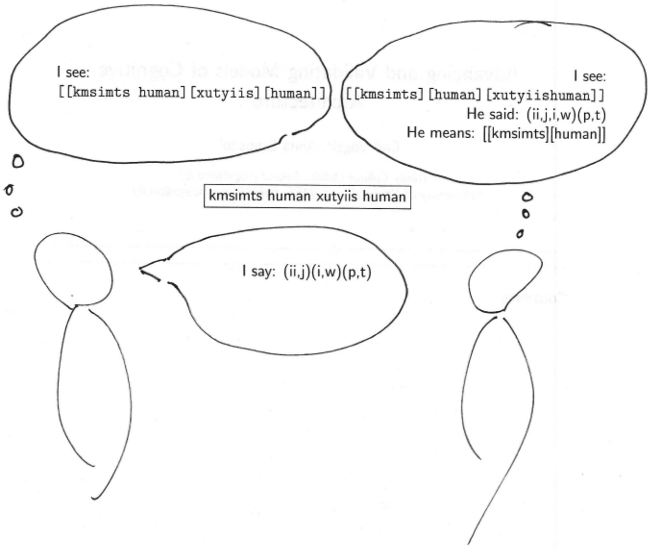

Fig. 1. Communication scenario between two simulated agents inventing a linguistic system

The basic scenario is depicted in Figure 1: showing an interaction event between two agents. The agents in the model have access to events that happen, but possibly distinct perspectives on those events. In the example, the event generated is modeled by the sequence of atoms, "`kmsimts human xutyiis human`". Events may be arbitrarily complex in terms of embedding relations embedded within relations, and this entails that there is not a finite space of possible meanings. Perspectives on events are modelled by distinct bracketing sequences (e.g. "`[kmsimts, human][xutyiis][human]`". vs "`[kmsimts][human][xutyiis][human]`"). Speakers label with phoneme sequences each bracketed item that they individuate within their own perspective on the shared event, and utter that sequence of phonemes (e.g. "`(ii, j)(i,w),(p,t)`"). As this is about the first uses of language, speakers initially have to innovate mappings between phoneme sequences and meaning chunks; however, they have an inclination to re-use past associations. This represents an influence of the imperative to identify and re-use copies (as discussed in Section III). Hearers have access to the same sound stream, but may segment that stream differently from the speaker (e.g. "`(ii,j,i,w),(p,t)`"), and anchor within their own perspective of the event a meaning chunk that corresponds to each phoneme sequence in what they have just heard. Hearers also have an inclination to copy past associations, but also have a potential to innovate in their interpretation, just as speakers have this potential in their

---

[5] For example, in the work of [49], there is a finite space of possible meanings. Further, agents have the capacity to sample what agents say in the context of meanings to be expressed. Thus, agents have direct access to the intended meanings as they develop a language for conveying those meanings. In the context of development, consider assumptions made by [34, p. 29, emphasis in the original]: "A third possible source of input that I *will* exploit ... depends on the assumption that the child can infer the meaning of adults' utterances from their physical and discourse contexts and from the meanings of individual words in the sentence.... Thus the input to the acquisition mechanism might be a pair consisting of a sentence and its meaning."

production. Depending on parameter settings, speaker and hearer may or may not obtain feedback on the overall accuracy of their communication, but never obtain direct feedback on which elements of their communication are correct or incorrect (this approach models task based communication in which shared tasks may be successfully accomplished even without mutual understanding of the language used to negotiate the task). Using these models, even with many agents, the "copy imperative" is sufficient to lead *quickly* to emergent systems of communication that exhibit significant connunicative success. What is remarkable about simulations using the LEW system is that even with pessimistic assumptions about communicative success, interesting levels of understanding can be achieved.

Simulations show that where the world is more structured (when possible event types vary in likelihood, following a Zipfian distribution) then agents develop communication systems more quickly and show greater coordination than when the world is random [47]. It can be shown that having a propensity to share phoneme segmentation patterns improves coordinations on meaning in the overall system [21]. Where agents may be involved in sub-communities, small groups are more likely to achieve more meaning convergence than larger fully interconnected groups with the same population size [22]. Where agents obtain partial feedback on success (such as described above) the systems converge on greater levels of meaning coordination than where such feedback is unavailable [23]. Where groupings of agents evolve in accord communication success among the agents within those groups, the communication systems are all the more viable [48]. Thus, the LEW (and systems like it) can be used to test theories about the interaction of various parameters and the potential for the resulting communication codes to mimic properties of natural language (such as low synonymy [50] or emergence of linguistic conventions within groups [51] and depending on the level of participation [52]). The relevance of group size and more successful information exchange among small groups rather than larger groups of agents is predicted by "the Data Processing Theorem" which states that the output of any processing system cannot contain more information than the input signal, suggesting that as more agents process an event, the information loss is commensurately greater [53], [54].

V. CLOSING REMARKS

Section III suggested an example of a means of advancing and validating aspects of models of cognitive architecture: seeking robust processes, cognitive faculties, that may have useful functions of varying importance across cognitive domains (and as such, perhaps qualify neither as horizontal nor vertical cognitive faculties in Fodor's sense, as discussed in Section II). Section IV suggested that systems of interacting agents may be modelled using abstract models of embodiment. This does not underestimate the value of simulations that depend upon physical embodiment [55], [13], [16], [17]; however, it does emphasize that the essence of cognition is abstraction over embodiment. This section emphasized simulation of communication systems, but as hinted in the introduction, comparable models that allow modelling of other dimensions of human existence and interaction, such as emotional thought and communication, also have scope for enhancing understanding of the parameters that influence those dimensions. In both of these approaches, we have emphasized the importance of exploring "diagonal" faculties – symmetry identification and production. This is not because we think that the "copy imperative", even if linked to neural substrate in mirror neurons, provides a "silver bullet" explanation of intelligent behaviours, but because we think that this is a human ability(one among many) whose role is as yet incompletely mapped out. We suggest here a methodology for pursuit of this exploration for this and comparable putative functions within models of cognitive architecture. In sum, we think that robust process models and simulation systems are crucial in advancing and validating models of cognitive architecture.

ACKNOWLEDGMENT

This research is supported by the Science Foundation Ireland (grants 12/CE/I2267 and 13/RC/2106) through the CNGL programme and the ADAPT Centre.

REFERENCES

[1] J. L. Krichmar, "Design principles for biologically inspired cognitive robotics," *Biologically Inspired Cognitive Architectures*, vol. 1, pp. 73–81, 2012.

[2] D. Vernon, C. von Hofsten, and L. Fadiga, "Desiderata for developmental cognitive architectures," *Biologically Inspired Cognitive Architecures*, vol. 18, pp. 116–127, 2016.

[3] C. W. Tyler, Ed., *Human Symmetry Perception and Its Computational Analysis*. VSP International Science Publishers, 1996.

[4] J. E. Hopcroft and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley Publishing Co., Reading MA., 1979.

[5] E. Bach, C. Brown, and W. Marslen-Wilson, "Cross and nested dependencies in german and dutch: A psycholinguistic study," *Language and Cognitive Processes*, vol. 1, no. 4, pp. 249–262, 1986.

[6] C. M. Vogel, U. Hahn, and H. Branigan, "Cross-serial dependencies are not hard to process," in *Proceedings of the 16th International Conference on Computational Linguistics*, 1996, pp. 157–62, cOLING'96, Copenhagen, Denmark.

[7] G. Westphal-Fitch, L. Huber, J. C. Gómez, and W. T. Fitch, "Production and perception rules underlying visual patterns: effects of symmetry and hierarchy," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, no. 1598, pp. 2007–2022, 2012. [Online]. Available: http://rstb.royalsocietypublishing.org/content/367/1598/2007

[8] A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt, Eds., *Development of Multimodal Interfaces: Active Listening and Synchrony*. Berlin: Springer, LNCS 5967, 2010.

[9] D. Reitter, F. Keller, and J. Moore, "Computational modeling of structural priming in dialogue," in *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*. Association for Computational Linguistics, 2006, pp. 121–124.

[10] D. Reitter and J. Moore, "Predicting success in dialogue," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Association for Computational Linguistics, 2007, pp. 808–815.

[11] C. Vogel, "Attribution of mutual understanding," *Journal of Law & Policy*, vol. 21, no. 2, pp. 377–420, 2013.

[12] J. Hurford, "Biological evolution of the saussurean sign as a component of the language acquisition device," *Lingua*, vol. 77, pp. 187–222, 1989.

[13] L. Steels, "Synthesizing the origins of language and meaning using co-evolution, self-organization and level formation," in *Evolution of Human Language*, J. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge University Press, 1997, pp. 384–494.

[14] J. R. Hurford and S. Kirby, "Co-evolution of language-size and the critical period," in *New Perspectives on the Critical Period Hypothesis and Second Language Acquisition*, D. Birdsong, Ed. Lawrence Erlbaum, 1998.

[15] S. Kirby, "Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners," in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy, and J. Hurford, Eds. Cambridge University Press, 2000, pp. 303–323.

[16] L. Steels and F. Kaplan, "Bootstrapping grounded word semantics," in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, T. Briscoe, Ed. Cambridge University Press, 2000, pp. 53–73.

[17] L. Steels, F. Kaplan, A. McIntyre, and J. V. Looveren, "Crucial factors in the origins of word-meaning," in *The Transition to Language*, A. Wray, Ed. Oxford University Press, 2002, ch. 12, pp. 252–271.

[18] A. Cangelosi, "Evolution of communication and language using signals, symbols, and words," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 2, pp. 93–101, 2001.

[19] A. Cangelosi and D. Parisi, Eds., *Simulating the Evolution of Language*. London: Springer, 2002.

[20] D. Parisi and A. Cangelosi, "A unified simulation scenario for language development, evolution and historical change," in *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds. Springer, 2002, pp. 255–275.

[21] C. Vogel, "Group cohesion, cooperation and synchrony in a social model of language evolution," in *Development of Multimodal Interfaces: Active Listening and Synchrony*, A. Esposito, N. Campbell, C. Vogel, A. Hussain, and A. Nijholt, Eds. Springer, LNCS 5967, 2010, pp. 16–32.

[22] M. Bachwerk and C. Vogel, "Modelling social structures and hierarchies in language evolution," in *Research and Development in Intelligent Systems XXVII*, M. Bramer, M. Petridis, and A. Hopgood, Eds. London: Springer, 2010, pp. 49–62.

[23] ——, "Establishing linguistic conventions in task-oriented primeval dialogue," in *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues*, A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelauchaud, and A. Nijholt, Eds. Heidelberg: Springer, 2011, pp. 48–55.

[24] R. Bryll, F. Quek, and A. Esposito, "Automatic hand hold detection in natural conversation," in *IEEE Workshop on Cues in Communication*, 2001.

[25] A. Esposito and M. Marinaro, "What pauses can tell us about speech and gesture partnership," in *The Fundamentals of Verbal and Non-Verbal Communication and the Biometrical Issue*, A. Esposito, E. Keller, M. Marinaro, and M. Bratanic, Eds. Amsterdam: IOS Press, 2007, pp. 45–57.

[26] A. Esposito, "The perceptual and cognitive role of visual and auditory channels in conveying emotional information," *Cognitive Computation*, vol. 1, no. 3, pp. 268–273, 2009.

[27] N. Bourbakis, A. Esposito, and D. Kvaraki, "Extracting and associating meta-features for understanding peoples emotional behaviour: Face and speech," *Cognitive Computation*, vol. 3, no. 3, pp. 436–448, 2011.

[28] K. Scherer, R. Banse, and H. Wallbott, "Emotion inferences from vocal expression correlate across languages andcultures," *Journal of Cross-Cultural Psychology*, vol. 32, no. 1, pp. 76–92, 2001.

[29] A. Esposito, M. T. Riviello, and N. Bourbakis, "Cultural specific effects on the recognition of basic emotions: A study on italian subjects," in *HCI and Usability for e-Inclusion: 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2009, Linz, Austria, November 9-10, 2009 Proceedings*, A. Holzinger and K. Miesenberger, Eds. Springer Berlin Heidelberg, 2009, pp. 135–148.

[30] M. T. Riviello, A. Esposito, and K. Vicsi, "A cross-cultural study on the perception of emotions: How Hungarian subjects evaluate American and Italian emotional expressions," in *Cognitive Behavioural Systems*, A. Esposito, A. M. Esposito, A. Vinciarelli, R. Hoffmann, and V. C. Müller, Eds. Springer, LNCS 7403, 2012, pp. 424–448.

[31] A. Esposito, A. M. Esposito, and C. Vogel, "Needs and challenges in human computer interaction for processing social emotional information," *Pattern Recognition Letters*, vol. 66, pp. 41–51, 2015.

[32] A. Esposito and L. C. Jain, "Modeling social signals and contexts in robotic socially believable behaving systems," in *Toward Robotic Socially Believable Behaving Systems Volume II - Modeling Social Signals*, A. Esposito and L. C. Jain, Eds. Springer International Publishing Switzerland, 2016, pp. 5–13.

[33] A. Leshinskaya and A. Caramazza, "For a cognitive neuroscience of concepts: Moving beyond the grounding issue," *Psychonomic Bulletin & Review*, vol. 23, no. 4, pp. 991–1001, 2016.

[34] S. Pinker, *Language Learnability and Language Development*. Harvard University Press, 1996 (1984).

[35] J. A. Fodor, *The Modularity of Mind*. Cambridge: MIT Press, 1983.

[36] R. A. Brooks, "How to build complete creatures rather than isolated cognitive simulators," *Architectures for intelligence*, pp. 225–239, 1991.

[37] A. Leshinskaya and A. Caramazza, "Abstract categories of functions in anterior parietal lobe," *Neuropsychologia*, vol. 76, pp. 27–40, 2015.

[38] A. V. Aho, "Indexed grammars—an extention to context-free grammars," *Journal of the Association for Computing Machinery*, vol. 15, no. 4, pp. 647–671, 1968.

[39] G. Pullum and G. Gazdar, "Natural languages and context-free languages," *Linguistics and Philosophy*, vol. 4, pp. 471–504, 1982.

[40] S. Shieber, "Evidence against the context-freeness of natural language," *Linguistics and Philosophy*, vol. 8, no. 3, pp. 333–43, 1985.

[41] C. Culy, "The complexity of the vocabulary of bambara," *Linguistics and Philosophy*, vol. 8, no. 3, pp. 345–51, 1985.

[42] W. T. Fitch and A. D. Friederici, "Artificial grammar learning meets formal language theory: an overview," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 367, no. 1598, pp. 1933–1955, 2012. [Online]. Available: http://rstb.royalsocietypublishing.org/content/367/1598/1933

[43] V. Dahl and M. McCord, "Treating coordination in logic grammars," *American Journal of Computational Linguistics*, vol. 9, no. 2, pp. 69–91, 1983.

[44] D. Milward, "Non-constituent coordination: Theory and practice," in *Proceedings of the 15th Annual Conference on Computational Linguistics*, 1994, pp. 935–41.

[45] D. Vernon, "Reconciling constitutive and behavioral autonomy. the challenge of modellling development in enactive cognition," *Intellectica*, vol. 1, no. 65, pp. 63–79, 2016.

[46] C. Vogel and J. Woods, "Simulation of evolving linguistic communication among fallible communicators," in *Proceedings of the Fourth International Conference on the Evolution of Language*, J. Hurford and T. Fitch, Eds. Harvard University, Cambrige, MA, 2002, p. 116.

[47] ——, "A platform for simulating language evolution," in *Research and Development in Intelligent Systems XXIII*, M. Bramer, F. Coenen, and A. Tuson, Eds. London: Springer, 2006, pp. 360–373.

[48] M. Bachwerk and C. Vogel, "Language and friendships: A co-evolution model of social and linguistic conventions," in *The Evolution of Language*, T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, and J. R. Hurford, Eds. Singapore: World Scientific, 2012, pp. 34–41, proceedings of the 9th International Conference (EvoLang9).

[49] J. Hurford, "The evolution of the critical period for language acquisition," *Cognition*, vol. 40, pp. 159–201, 1991.

[50] A. Carstairs-McCarthy, "Synonymy avoidance, phonology and the origin of syntax," in *Approaches to the evolution of language. Social and cognitive bases*, J. R. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge University Press, 1998, ch. 17, pp. 279–296.

[51] S. Garrod and G. Doherty, "Conversation, co-ordination and convention: an empirical investigation of how groups establish linguistic conventions," *Cognition*, vol. 53, pp. 181–215, 1994.

[52] P. G. T. Healey and G. J. Mills, "Participation, precedence and co-ordination in dialogue," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, 2006, pp. 1470–1475.

[53] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Son, 1968.

[54] A. Esposito, "The amount of information on emotional states conveyed by the verbal and nonverbal channels: Some perceptual data," in *Progress in Nonlinear Speech Processing*, ser. LNCS 4391, Y. Stylianou, M. Faundez-Zanuy, and A. Esposito, Eds. Springer-Verlag: Berlin, 2007, pp. 249–268.

[55] L. Steels, "The spontaneous self-organization of an adaptive language," in *Machine Intelligence*, S. Muggleton, Ed. Oxford University Press, Oxford, 1996, vol. 15.