

Introduction: concepts of development, learning, and acquisition¹

KATRIN LINDNER AND ANNETTE HOHENBERGER

‘Learning’, in a very broad sense of cognitive psychology, is defined as “a relatively permanent change in an organism’s potential for responding that results from prior experience or practice” (Gordon 1989: 6). This special issue will deal with learning in a more narrow sense, with learning in human beings and, more specifically, with language learning.

Due to its association with early behaviorism, ‘language learning’ has long been a frowned upon notion. It was substituted with the term ‘language acquisition’ by researchers with a generativist agenda. In opposition to the empiricist tradition, they established new notions along with a new linguistic paradigm — generative grammar. ‘Learning’ has not only been largely banned from the linguistic literature but it has also received less attention in cognitive psychology, due to the “cognitive revolution” (Baars 1986) instigated by Chomsky (1957, 1959). One of the consequences of this paradigmatic turn was that new information processing models emphasized knowledge representation and de-emphasized the learning process (Glaser 1990; Reber 1993: 4).

In this special issue, we take a fresh look at the old controversy in the light of novel and challenging approaches which have arrived recently in the field — implicit modes of learning, bootstrapping, optimality theory, connectionism, usage-based and emergentist approaches, self-organizing and dynamic systems. In this introductory paper we will set the scene for the articles to come: In Section 1 we will start out with a clarification of basic terms like development, learning, and acquisition.² In Section 2 we will briefly recapitulate the various stages in the debate between rationalists and empiricists up to now. In Sections 3 and 4 more recent approaches will be characterized before, in Section 5, the outline of the special issue including the individual papers will be summarized.

1. Development — learning — acquisition

Learning has often been discussed in relation to ‘development’ and to acquisition. Of these three terms, development is the most neutral one, embracing a whole range of scenarios, from the unfolding of a predetermined genetic plan up to the active interaction of the organism with the environment. In the psychological literature, the labels ‘development’ and ‘learning’ (Savelsbergh 2003)³ as well as ‘learning’ and ‘acquisition’ (van Geert 1986: 26–27; Reber 1993: 5) are often used interchangeably. In most of the linguistic literature, however, ‘learning’ and ‘acquisition’ are seen as a dichotomy.

If ‘development’ and ‘learning’ are distinguished, as in van Geert (1986: 24), they are differentiated in terms of intentionality. In ‘learning’, what is learned is “viewed as the goal of intentional acts carried out by the learner”, whereas in ‘development’, “it is the nonintentional result either of something that occurs to the subject, such as biological maturation, or of intentional activities of the subject himself”. Since learning is considered to be intentional, it is equated with ‘explicit learning’. However, this is only one type of learning: Another type is ‘implicit learning’ which has been the topic of more recent studies (Reber 1993; Stadler and Frensch 1998; Cleeremans 1997; Cleeremans et al. 1998; see also Weinert this issue, and Section 3.2). Implicit learning is considered to be unintentional, is based on associative learning, and leads to tacit knowledge.

As for the relationship between learning and development, Elman (2005) points out that a developmental theory that relies only on learning does not capture the whole picture. ‘Maturation’, i.e. primarily endogenous changes, has to be taken into account as well. Learning and maturation interact in that learning is constrained to certain time windows in the maturational processes (Elman et al. 1996). Changes from simple to complex behavior (“the scaling problem”) can only be explained via development. Complex behavior is the product of a causal chain of developmental stages. “In other words, development may be Nature’s solution to

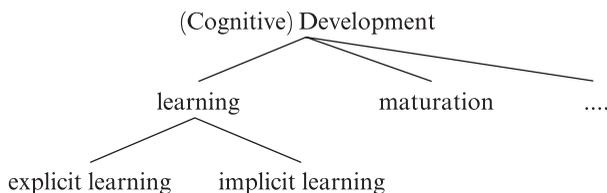


Figure 1. *The relation of learning and development*

its own scaling problem.” (Elman 2005: 116). Figure 1 provides a rough outline of the relationship between learning and development.

Thus a learning theory has to be embedded into a developmental theory, which also includes maturation. Fig. 1 does not yet show the complete picture. In order to fill in the blank spaces, let us turn to the relationship between learning and acquisition.

1.1. *Language learning vs. language acquisition*

From a semantic point of view, both expressions — ‘to learn something’ and ‘to acquire something’ — seem to indicate a process, yet the processes differ with regard to the time span and the characteristics of the end state involved: while learning implies a process without a prespecified, let alone successful end state, acquisition, in contrast, presupposes that after a short time the process comes to a successful end.⁴ Furthermore, in the case of ‘to acquire’ the ‘acquired X’ seems to be or become an object, compressed and not spaced out in time (cf. [1], [2]):

- (1) He acquired/*learned a fortune/ shares of firm X/ the right to vote/ a client/ competence in speaking language X.
- (2) He acquires/ learns sign language.

All these properties of ‘acquisition’ characterize precisely the concept proposed by generative approaches.

If the term ‘learning’ is used in the generative literature, then it is used in a negative way as in the ‘learnability problem’ (Gold 1967; Wexler and Culicover 1980), i.e., language cannot be learned. Therefore, it is claimed that relevant mechanisms (learning algorithms) and linguistic categories are part of children’s biological endowment (for more details see Eisenbeiß this issue).

Given that acquisition and implicit learning imply a nonintentional learner and lead to implicit/tacit knowledge, can ‘acquisition’ be identified with ‘implicit learning’? Krashen’s Acquisition-Learning Hypothesis in second language acquisition (Krashen 1987, 1988) seems to identify the two. According to him, second language learners may both acquire and learn a language. ‘Learning’ is the conscious, effortful study of the forms of language. ‘Acquisition’ is the unconscious, intuitive picking up of a language through meaningful interaction in the target language, much like in first language acquisition. While both systems in themselves are incommensurable, they can and should be combined, with acquisition as the more important mechanism taking the lead and learning acting as the monitor or editor of the planned utterance in the second language.

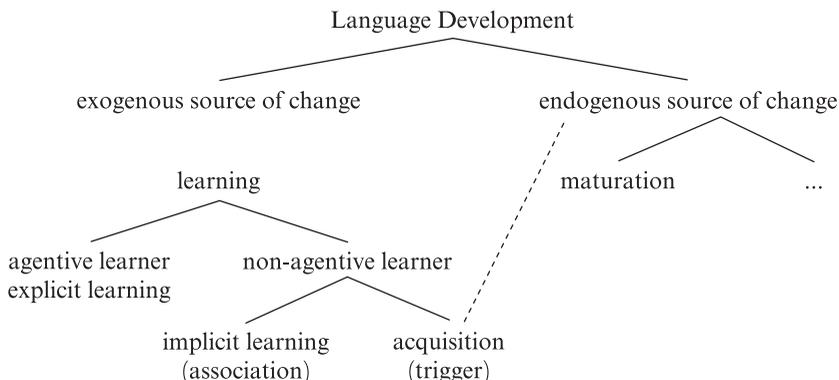


Figure 2. *The relation between development, types of learning, acquisition, and maturation with regard to language acquisition*

However, ‘acquisition’ cannot readily be equated with implicit learning from a generative perspective, since generativists deny associative learning as a core learning mechanism and instead claim that the learning mechanism is triggering. Figure 2 summarizes the concepts and underlying learning mechanisms discussed so far.

Note that in Figure 2, ‘acquisition’ has a twofold connection: it is related to the concept of a ‘nonagentive learner’ under the branch of ‘exogenous source of change’ and to ‘maturation’ under the branch of ‘endogenous source of change’, reflecting the fact that ‘acquisition’ requires exposition to the Primary Linguistic Data (henceforth PLD, i.e., positive evidence) on the one hand and, on the other hand, presupposes a rich innate endowment which, for some researchers, means maturation (Wexler 1999).

2. Historical background: two epistemological positions in combat

2.1. *Empiricism vs. rationalism*

The main controversy about language being learned or acquired has evolved in waves of philosophical dispute between the empiricist and rationalist tradition (cf. also Jordan 2004; Russell 2004). This controversy is also known as the “nurture-nature” debate (Elman et al. 1996; Tomasello and Slobin 2004, among others).

The empiricist tradition started off with Aristotle in the ancient world and was revived in the Middle Ages by St. Thomas Aquinas. In the 17th

century it inspired empiricism and sensualism of Locke and Hume, who, in turn, are considered the predecessors of behaviorism. Aristotle held that physical objects were fully grasped through experience.⁵ The most concise quote by which the doctrine of empiricism is captured comes from St. Thomas Aquinas — “Nihil est in intellectu quod prius non fuerit in sensu” [Nothing is in the intellect which was not first in the senses]. Locke coined the dictum that the infant’s mind is comparable to a *tabula rasa*⁶ and is thus devoid of any innate ideas.

The rationalist tradition started out with Plato. In his dualist metaphysics he held that our perceptual experiences allow only indirect access to the ultimate real forms, namely the ideas and universals (compare his famous ‘allegory of the cave’). This line of thinking was revived in Descartes’ dualist thinking in the 16th century and by Leibniz’ proposal of innate ideas (god, perfection, eternity) in the 17th century. Chomsky (1981, 1986), as a contemporary representative of the rationalist tradition, asks how a child can acquire so much sophisticated linguistic knowledge in such a short time and without conceivable effort, given only limited and imperfect input. The answer to the “logical problem of language acquisition” (Hornstein and Lightfoot 1981; Baker 1979), which is also called “Plato’s problem,” lies in the assumption that knowledge of language is innate.

2.2. Behaviorism and connectionism vs. generativism

At the beginning of the 20th century, learning theories became the dominant paradigm in psychology, starting with Watson (1913, 1930) and Thorndike (1913, 1932), and followed by Pavlov (1928) and Skinner (1938). They claimed that all learning involves some very fundamental and easy learning principles, namely classical, instrumental and operant conditioning.⁷ In classical conditioning *sensu* Pavlov, an organism (e.g., a dog) learns through experience to associate an unconditioned stimulus (UCS, e.g., a piece of meat) with a conditioned stimulus (e.g., a tone). After repeated exposure to this contingency the UCS acquires the power to elicit a conditioned response (CR, production of saliva) which is similar to the spontaneous unconditioned response (UCR) to the UCS. In instrumental or operant conditioning *sensu* Skinner, the conditioned response (e.g., pecking on a response lever) is emitted voluntarily by the organism (e.g., a pigeon) which experiences some consequence of its behavior (reward or no reward). If the consequence is positive, the behavior becomes reinforced; if the consequence is negative, the behavior becomes extinct. Central to operant conditioning is the stimulus-response (S-R)

connection formed by associative bonds. According to the “law of effect” (Thorndike 1913), the strength of these bonds is contingent upon the consequences: rewards strengthen the bonds, whereas lack of reward weakens them.

In his *Verbal Behavior*, Skinner (1957) applied the learning mechanism of operant conditioning to language, claiming that “verbal behavior is similar to any other operant response, controlled by reinforcement and punishment.” (Mowrer and Klein 2001: 13). Operant conditioning also acquired explanatory force in functionalist approaches to language learning (Halliday 1975). Functionalists argued that in order to communicate successfully with the environment and thus receive reinforcement for their verbal behavior, children speak like their environmental models do. Furthermore, contingencies and transitional probabilities between words may account for word order and particular constructions and can be modeled e.g., by Markov chains.

Following the criticism of Skinner’s *Verbal Behavior* (1957) by Chomsky (1959),⁸ behaviorism was renounced as a viable theoretic account of language in particular and cognitive science in general. For language, Chomsky argued that structure-dependency, syntactic transformations and structural ambiguity could not possibly be learned by any general problem-solving mechanism such as differential reinforcement, association, imitation, analogy or generalization (Piattelli-Palmarini 1989). Rather, he invoked innate constraints on the format of possible human languages, known as Universal Grammar (UG). UG makes available to infants only a subset of languages, from which the proper choice is triggered by a minimum amount of PLD.

Although Chomsky claims that structural properties of language are not learned in one of the major generative language acquisition models, the Principles and Parameters theory, there remains a “learning residue”. Atkinson locates this residue in the interaction of the open parameters with perceptual experience and in the minimal induction process involved in the formulation and testing of hypotheses concerning the parameter (Atkinson 1986: 121 and 126).

In connectionism, as in behaviorism, learning also plays a pervasive role. Connectionism is related to behaviorism in two aspects. Firstly, experience — here the linguistic input — is the dominant factor in learning and secondly, association based on spatio-temporal contiguity, similarity, and analogy is the underlying learning mechanism.⁹ The major methodology of connectionism is simulation of various cognitive processes in neural networks (Elman et al. 1996; Plunkett 2001, Westermann et al. this issue, among others). Over the past twenty years, connectionist models have developed from a simple 1:1 mapping from input to output

to very sophisticated network architectures. The major improvements came from the addition of hidden layers between input and output and from recurrent connections (Elman et al. 1996; McLeod et al. 1998). The current connectionist models are biologically inspired and can model critical periods or time windows in development and neural plasticity (Elman et al. 1996). With regard to the history of science in the field, connectionist models (as well as other models proposing emergent structures, see below) have taken up the challenge of falsifying claims which formed the basis of the generative approach such as the ‘poverty of stimulus’ (see Section 2.2).

The properties of current connectionist models can most easily be characterized in contrast to the “classical metaphor of cognition” in symbolist frameworks (Cleeremans 1997). As Bates and Elman (1992) argue, the classical metaphor holds that

1. Representations are discrete.
2. Rules tend to be absolute.
3. Learning is viewed as programming or memorizing (in line with the computer metaphor). New knowledge is due to hypothesis testing.
4. There is a classical distinction between hard- and software (in line with the computer metaphor).

The alternative view of connectionism, as reported by Cleeremans (1997), holds that

- 1'. Representations are more flexible due to processing of partial information and may thus be graded.
- 2'. Organisms are sensitive to regularities and patterns in the input. The postulation of absolute rules which either apply or fail to apply is not necessary.
- 3'. Learning is due to experience and happens continuously.
- 4'. Instead of the classical distinction between hard- and software, the machine and what it knows is basically the same.

Since experience plays a dominant role (see 3'), language input is rehabilitated as a major source that drives language learning.

The debate between generative and connectionist approaches has led to a diversification of approaches in both camps and to combinations of ideas and methods from the two. Some approaches have claimed the “radical middle” (e.g., Hennon et al. 2000). Some are closer to one or the other pole. Closer to the generative position are studies in the framework of Optimality Theory (see Fikkert and de Hoop this issue) and bootstrapping (see Höhle this issue). Closer to the connectionist position

are studies in the framework of usage-based (Tomasello 2000a, 2000b, 2003) and emergentist theories (MacWhinney 1999; see also Behrens this issue). An integration of statistical learning with rule-governed behavior is found in Marcus (2001) and with UG is found in Yang (2004).

2.3. *Major issues in the current debate between generativism and connectionism*

The discussion between generativism and connectionism is one between ‘deductive and inductive models’ (Broeder and Murre 2000) and concerns the ‘logical problem of language acquisition’ in particular the “poverty of the stimulus” argument (Pullum and Scholz 2002; Pullum 2003), the availability of negative or positive evidence and general constraints on learning grammar (MacWhinney 2004).

Gold’s (1967) proof that nonfinite languages could not be learnt by mere *text presentation* (only positive evidence) has been used by linguists and philosophers as “a powerful argument for the existence of innate knowledge of universal grammar that can assist in learning” (Pullum 2003). Gold proved that nonfinite grammars are not learnable without negative evidence which helps learners to recover from incorrect overgeneralizations. Only through *informant presentation* (exposure to sentences plus parental grammaticality judgments of the child’s speech) would nonfinite languages become learnable (cf. Atkinson 1992; Pullum 2003). Gold’s proof was supported by the unavailability of negative feedback. It was found that (i) parents did not correct their offspring’s utterances grammatically (at most its content) and (ii) children did not even make use of repairs when they were provided. The only kind of negative evidence that was available throughout the course of language acquisition was indirect negative evidence. From absent structures in the input the child can “deduce” that they are not part of the language to learn.

In the 1980s the focus in the generative debate shifted to the unavailability of positive evidence in ‘error-free learning’ based on production data. (MacWhinney 2004: 888–895). If it were true that children produced grammatical utterances with respect to some constructions without any prior positive evidence, only nativism could yield a coherent explanation (see Eisenbeiß this issue). The major arguments concerned structural dependency, extraction from complex NPs, binding principles, etc. Yet, MacWhinney (2004) argues that it has never been shown that this learning has taken place in the absence of positive evidence (p. 895).¹⁰

Pullum (2003) and Scholz and Pullum (2006) address another aspect of Gold’s classical learnability proof, namely that the grammar has to be

learned 100 percent (compare [2] in the list below). They propose that, alternatively, “probably approximately correct” or “(PAC) learning” (Scholz and Pullum 2006: 70; cf. also Pullum 2003) may be invoked, under which grammar learning is possible if the goal of explaining grammar learning is weakened to less than 100 percent.

MacWhinney (2004) refutes the logical problem of language acquisition by proposing a mixture of seven solutions which, operating together, “support full and successful language learning” (pp. 897–909):

1. *Limiting the class of grammars* to those which are learnable on the basis of positive data, namely finite state grammars.
2. *Revising the end-state criterion* of learning to probabilistic and not absolute identification (see Pullum 2003).
3. *Conservatism*: A *conservative* learner will not overgenerate and hence needs no corrective feedback.
4. *Competition* along with conservatism will strengthen correct representations based on rote learning (e.g., *went*), against the pressure to generate analogous forms (*goed*).
5. Through *cue construction* the child generates features which block overgeneralization.
6. By *monitoring* and correcting her own faulty output in the light of the PLD the child converges on the correct form.
7. *Indirect negative evidence*.

Given these seven solutions, MacWhinney denies that recovery from overgeneralizations is impossible. Thus, the ‘logical problem’ is obsolete. Alternative mechanisms can account for language learning. They will be described in the next section.

3. More recent concepts of learning

3.1. *Distributional, statistical, or probabilistic learning*

While in the literature ‘learning’ has always been accepted as the way to build up vocabulary, it has only been in the last 15 years that researchers have begun to explore children’s skills in detecting distributional, i.e. basic statistical properties of the language to learn.

‘Distribution’ is the result of a ‘discovery procedure’ known since structural linguistics; i.e. the distribution of *x* is the occurrence of *x* in a particular linguistic unit (in a syllable, a word, a sentence or an utterance). Most often *x* is seen in relation to the co-occurrence of *y*: while *x* may be a phoneme, a syllable, a word etc., *y* may be another phoneme, the onset,

nucleus or coda of a syllable, a word, a prosodic cue at the end of a sentence, an object in the real world or the like. For instance, in English the determiner *the* is positioned in front of a noun (or an attributive adjective + noun) and it is only observed with these constituents. Thus it is possible for the learner to detect words that “go together”. These word groups then lead to larger constituents which in turn facilitate the finding of a hierarchical structure. The determiner, moreover, allows for prediction of which word class the next word belongs to (given *x* then *y* is to be expected).¹¹ In this way it is possible to look at the statistical properties of language and carry out probabilistic ‘reasoning’ in terms of occurrences or covariation and in terms of hierarchical structure.

Statistical learning of this kind has been observed in young children with regard to various domains: most often it is explored with regard to the discovery of prosodic regularities (e.g., Jusczyk et al. 1993a, 1993b, 1994), to word segmentation with or without prosodic cues (e.g., Brent and Cartwright 1996; Christiansen et al. 1998; Saffran et al. 1996a; Aslin et al. 1998; for adults Saffran et al. 1996b), to word sequencing and hierarchical structure (in children Gomez and Gerken 2000 and in adults Saffran 2001), to inflectional morphology like verb or noun paradigms (e.g., Nakisa and Hahn 1996), word classes (Höhle et al. 2004) and lexical semantics (e.g., Redington and Chater 1997). Statistical learning is also involved in a number of processes summarized under “bootstrapping” (cf. Höhle this issue).

It has to be kept in mind that distributional learning is not limited to auditory information; it is also relevant for visual information (Canfield and Haith 1991) and for the learning of motor sequences (Hunt and Aslin 2001). Moreover, it is not restricted to human learners, as comparisons of young children with cotton-top tamarin monkeys show (Ramus et al. 2000; Hauser et al. 2001). Thus this type of learning seems to be rather basic, robust and neither species nor domain specific.

In a number of studies, such distributional cues have been analyzed with regard to mutual support or competition (see Weinert this issue). A very prominent psycholinguistic model based on such a view is the Competition Model by Bates and MacWhinney (e.g. 1982, MacWhinney and Bates 1989).¹² In this model prosodic, morphosyntactic, semantic and/or pragmatic cues compete with one another in sentence interpretation and production. On this account the meaning of a sentence is the emergent result of such competition.

Many studies about statistical learning of languages are concerned with learning an artificial language (cf. for instance, studies by Saffran and Aslin and their collaborators or by Gomez and Gerken 2000). Language learners — infants, children, and adults — are able to discover the under-

lying regularities in the data without learning them explicitly (cf. Weinert this issue). Thus Aslin et al. (1998) and others have also called this implicit learning. Interestingly, researchers in this field tend to be quite careful in the use of the term ‘rule’ (cf. Gomez and Gerken 2000). There is an as yet unresolved dispute on the terminology and interpretation of the very same phenomena. Connectionists would talk about patterns and regularities (of varying complexity and abstraction) while generativists would cling to the traditional notion of a rule as a symbolic object which can be considered and manipulated consciously (cf. the discussion in Marcus 1999; Marcus et al. 1999; Eimas 1999; Seidenberg and Elman 1999a, 1999b; McClelland and Plaut 1999).

Distributional analysis has been used by researchers who prefer a formal theory (e.g. Marcus et al. 1999; Hauser et al. 2002; Yang 2004) as well as by those who prefer a connectionist framework (Christiansen et al. 1998; Nakisa and Hahn 1996; Redington and Chater 1997). For the first group it is an addition to other methods and does not seem to contradict concepts like modularity etc. For the second group it is the major methodology for going about learning.¹³

3.2. *Implicit learning*

As mentioned before, Aslin, Saffran, and colleagues call distributional/statistical learning ‘implicit learning’, a concept which was first explored by Reber (1967), in particular with regard to learning an artificial grammar. Reber (1993) defines implicit learning as follows:

Implicit learning is the acquisition of knowledge that takes place largely independently of conscious attempts to learn and largely in the absence of explicit knowledge about what was acquired. (Reber 1993: 5)

Reber (1993) combines classical learning theory (namely conditioning) with modern cognitive psychology:

[...] the basic principles of cognitive induction and abstraction on the one hand and conditioning and associative learning on the other share a common process — the detection of covariation between events. (Reber 1993: 4)

In contrasting implicit and explicit learning systems, Reber holds that implicit learning is the default learning mechanism. It is phylogenetically older and therefore more robust and resilient than the more recent mechanism of explicit learning. Yet he suggests that implicit and explicit learning are not two totally separate ‘systems’, but that there are transitions

between the two. For a discussion of the relationship between the two based on empirical data from artificial grammar learning by children and adults, see Weinert (this issue).

In the last ten years the defining features of ‘implicit learning’ have received quite a bit of attention (e.g., Berry 1997; Stadler and Frensch 1998; Jiménez 2003 and numerous papers in journals like *Trends in Cognitive Science*): be it the number of learning systems to be assumed, the role of symbolic rules, the nature of representations, or the degree of consciousness. A summary of this debate including the suggested models is impossible here. Yet one position, the one by Cleeremans, should be mentioned, since it provides valuable criticism of the current discussions about processing and representations and opens up new perspectives.

Cleeremans (1997: 196) argues that these discussions in the field of implicit learning prevail since researchers are still tempted to think in terms of a symbolist theory of cognition: that there is a processor that takes ‘knowledge’ from one module, processes it and sends it off to another module for further processing.

One reason why the classical concept cannot accommodate implicit learning is that the processor and the representations are distinct; i.e., the processor manipulates the representations so that the representations can be the object of representations themselves.

This, so Cleeremans (1997), cannot hold for implicit learning. He defines it as:

At a given time [i.e., a particular context], knowledge is implicit when it can influence processing without possessing in and of itself the properties that would enable it to be an object of representation. Implicit learning is the process by which we acquire such knowledge.¹⁴ (Cleeremans 1997: 199)

Furthermore, so Cleeremans, the nature of these representations is not suitable for representing implicit learning: the representations are considered to be static and available for outside inspection, this flexibility of access being a defining feature of symbols. And they are compositional in nature. In symbolic systems elementary constituents are concatenated. Following van Gelder (1990) Cleeremans assumes (1997: 201) that “concatenation, by definition, preserves the tokens of complex expression’s constituents and their relationships in the expression itself.” Thus, the elements in these representations “covary with the things they represent” (p. 201).

These properties of representations in symbolic systems — flexibility in access and compositionality — are incongruent with the concept of representing implicit knowledge since such knowledge, although it influences processing, seems to be unavailable to outside inspection (p. 203).

Cleeremans (1997) discusses four principles which provide another framework for thinking about implicit knowledge. In particular, these four principles evade the tacit assumption of direct mapping between behavior and representation which is assumed in symbolic systems.

1. Sensitivity to some regularity does not imply that this regularity is represented as a manipulable object of representation in the learner's mind; e.g. a child's utterance that an adult would analyze as a chunk (in her mature language system) may not be represented as a chunk in the child's language system.
2. Architectural modularity has been suggested to explain dissociations in symbolic systems. Such a view presupposes independent modules. Yet there is empirical evidence, in particular from experiments which show that one variable may be selectively affected, but not the other. Thus modules are only functionally and not architecturally distinct (cf. Plaut 1995; Farah 1994; Pulvermüller 2001, 2005).
3. Experimental measures may not affect just one variable or one particular process, but a number of other variables or processes. Thus there may be no tasks that exclusively examine one process. This is particularly pertinent to experiments about implicit learning where one task concerns awareness, the other implicit learning (cf. e.g., Dunn and Kirsner 1988; Jiménez et al. 1996).
4. The dichotomy of controlled and automatic processes does not seem to hold. Rather, one has to consider a continuous dimension. 'Abstraction' or 'awareness' may in fact be graded and continuous rather than discrete and dichotomous. (cf. the findings by MacLeod and Dunbar 1988; Cohen et al. 1990).

These principles are most congruent with connectionist models. For instance, learning is not driven by incremental memorization, but by processing. There is no distinction between processing and learning. Neither is there a distinction between the processor and the building of representations. "The machine and what it knows are one and the same." (Cleeremans 1997: 199). Moreover, in connectionist models knowledge is represented neither symbolically, as a manipulable object, nor compositionally. Instead knowledge is conceptualized as patterns of activation distributed over many processing elements. Because of this distributional nature representations are not seen as compositional, i.e. concatenative.

Cleeremans compares implicit knowledge to linguistic presuppositions, i.e., when we process a sentence like *John paid the bill*, we may assume that John is in a restaurant, that he is male etc., but we may not be aware of all of the presuppositions associated with this sentence. Yet knowledge that we are not aware of comes to bear in processing this sentence.

3.3. Neurophysiology of learning

3.3.1. *The neurocognitive bases of learning.* Progress in the understanding of the neurocognitive and neurophysiological processes in the mature brain of adults and the developing brain of infants and children increasingly shapes our conceptualization of learning (Johnson 2005; Kuhl 2000, 2004). Neither a narrow localizationist nor a modular view of the brain nor a mere associationist or holistic view can accommodate the complex findings about the architectural and functional organization of the brain (Pulvermüller 1997: 25). A strict modular view (Fodor 1983), on the one hand, claims that the brain is innately organized into modules with a specific architecture and specific computations which therefore process data of a particular stimulus domain in a domain-specific way right from the outset. A holistic view such as Hebb's (1949), on the other hand, claims that the brain is nothing but a huge associative memory machine which builds up connections between neurons based on their pattern of co-activation in the past (Pulvermüller 1997: 26). Basically, everything can be connected with everything else. Hebb's concept of synaptic co-activation in cell assemblies comes closest to the connectionist ideal of a brain that becomes crafted by processing the structured information from the outside.¹⁵

While, no doubt, there are specific areas that are involved in language processing such as Broca's and Wernicke's area, the pars opercularis, and further subcortical structures sustaining language (Brown and Hagoort 1999), it has been found that nonlinguistic areas are involved in language processing as well. In a series of experiments, Pulvermüller (2001, 2005) has shown that in the processing of action words (such as *kick*, *pick* and *lick*), specific motor areas (for the leg, arm and tongue) are co-activated while in the processing of perceptual words the visual areas are co-activated. The best way to conceive of these findings is by postulating distributed "word webs" in which several areas of the brain — language, motor, perception, etc. — are temporarily bound together in a neuronal cell assembly and serve a particular function. The formation of specific neuronal cell assemblies is the result of associative learning as defined by Hebb.

The neurophysiological mechanisms underlying the formation of neuronal cell assemblies have been specified by Singer and Gray (1995) and by Pulvermüller (1999, 2001), amongst others. This mechanism of temporal binding has been referred to as "self-organization" by Singer and Gray since a higher level "order parameter" (the cell assembly) emerges through the cooperative activity of many local units (the individual neurons) in response to a stimulus. Self-organization refers to the shaping

of the brain's functionality through the interplay between broadly pre-specified computations and a biasing architecture of brain areas on the one hand and experiences the organism makes during ontogeny on the other hand. In this respect, neuronal self-organization has become a powerful dynamic concept which can capture the interaction 'nature' and 'nurture'.

3.3.2. *Developmental neuroscience of language.* New methods such as event-related potentials (ERP), magneto-encephalography (MEG), near-infrared spectroscopy (NIRS) and functional magnetic resonance imaging (fMRI) allow for the safe study of young children's brains during language tasks.

In a recent survey of ERP studies with infants and children, Friederici (2005) states that young language learners have the same ERP components as adults do. At two months they show a mismatch negativity (MMN) in a phoneme discrimination task, at nine months a closure positive shift (CPS) at intonational phrase boundaries, by twelve months a negativity (N 400) in a lexical semantic task, and by 32 months the two markers of syntactic processing, an early left anterior negativity (ELAN) and a positivity (P600). Thus there is continuity in the neurobiological processing of language. Qualitatively, the ERP components are the same, they differ only quantitatively in terms of temporal delay, strength, and expansion. With magneto-encephalography (MEG), Imada et al. (2006) found evidence for involvement of the left inferior temporal cortex (Broca's area) in the processing of syllables at around six months. With NIRS, Pena et al. (2003) showed that neonates already process forward speech in the left temporal regions, whereas no such dominance was found in backwards speech or silence (for converging results from an fMRI study, see Dehaene-Lambertz et al. 2002). The very early processing of speech in the left hemisphere is also corroborated in an fMRI study of Dehaene-Lambertz et al. (2006). In their study, three-month-old infants, who listened to 2-s-long sentences repeatedly, manifested a characteristic spatio-temporal activity band whose phase shifted systematically across the temporal lobe bilaterally until it reached Broca's area in the left frontal region.

Does the early left-hemispheric dominance in the processing of language in healthy infants prove that language is innately located in the brain? Johnson (2005) abstains from interpreting development in terms of a "causal epigenetic pathway" but suggests "a more probabilistic epigenetic view" in his "interactive specialization view" (Johnson 2005: 140–141). In typical development, small variation in the architecture of a particular brain area and/or an earlier timing of maturation biases this

area to process stimuli from a particular stimulus domain slightly better than other areas. For example, the left temporal lobe is slightly more apt at processing rapidly changing acoustic information. Thus it is more probable that this brain area will host language.

This view is compatible with nativist and empiricist approaches as long as they are nonexclusive. It takes seriously the dynamic interplay between genetic predispositions, early brain development, and the interaction with the environment.

4. Dynamic Systems Theory

In the 1990s a new framework emerged: Dynamic(al) Systems Theory (henceforth DST).¹⁶ Some conceive of DST as a novel paradigm (Beer 2000; van Gelder 1998) while others argue that DST goes together with connectionism, the two having only different histories (Thelen and Bates 2003). Still others conceive of both as complementary (Smith and Samuelson 2003; see Westermann et al. this issue). The fact that some aspects of DST have been integrated by both nativism and connectionism — a process which has led to some conceptual hybridization — reflects the attractiveness of this novel concept.

DST has roots in very different domains — mathematics, chemistry, meteorology, physics, biology, in particular motor systems, robotics, philosophy — and comes under a variety of notions such as self-organization, chaos theory, theory of fractals, nonlinear dynamic systems, dissipative systems, catastrophe theory and synergetics (see Hohenberger 2002).

Van Gelder (1998: 621–622) characterizes a dynamic(al) approach as follows:

1. Emphasis on change, not state.
2. Focus on the position of a state with respect to other states, not on structure itself.
3. Structure is laid out temporally.
4. The timing of events is more important than the order of particular events.
5. Dynamic systems operate in parallel, not serially.
6. Processes are ongoing rather than having a fixed beginning and end.
7. Coupling between the environment and the system.
8. Representations are not static but can be graded (cf. also Cleermans and Jiménez 2002).
9. Dynamic(al) systems are not necessarily representational.

In general, a dynamic(al) system approach looks at the development of variables in time (van Geert 1997). In nonlinear dynamical systems, input and output are related in a nonlinear way, i.e. big changes in input may engender small changes in output, and, most interestingly, small changes in input may engender big changes in output. Most biological systems are nonlinear dynamical systems. They have the capability to self-organize, i.e., they become irreversibly more complex through the cooperative activity of their subsystems. It is this latter property, emergence, through which novel qualities can arise. Self-organization and emergence are novel concepts which change our understanding of the cognitive and language development in young children. Crucially, DST suspends Fodor's claim (1980) that nothing new may ever occur in development if it is not innate. Note that nativism derived its attraction mainly through the impossibility of explaining language development through other mechanisms, in particular learning mechanisms (Reber 1993: 5). Self-organization, however, is a powerful device for explaining exactly how, without external or internal controller, macroscopic order arises through local interaction of microscopic entities. A pertinent example is syntax as an emergent property of the lexicon having surpassed a critical threshold (Bates and Goodman 1999; Marchman and Bates 1994). In this special issue, two contributions — Hockema and Smith as well as Hohenberger and Peltzer-Karpf — present varieties of DST applied to the development of language. The contribution by Westermann et al. shares a number of assumptions with DST.

5. The purpose of this special issue

As the preceding sections have shown, the various new approaches to language learning can no longer be strictly allocated to the two big paradigms, empiricism and nativism (or generativism and connectionism). Today we have to orientate ourselves in a much more diversified field where various approaches overlap and hybrid models have emerged, as between distributional learning and generative grammar in bootstrapping approaches (see the contribution of Höhle); between self-organization and connectionist theory (as in biologically inspired connectionism; see Elman et al. 1996 and the contributions of Hockema and Smith as well as Westermann et al. in this issue); self-organization and generative grammar (see the contribution of Hohenberger and Peltzer-Karpf); between connectionism and generative grammar, as in Optimality Theory (see the contribution of Fikkert and de Hoop); or between usage-based approaches and connectionism (see the contribution of Behrens).

Since 'learning' is widely used in the field, it is high time to clarify the following questions:

- What are these concepts of learning?
- Where do they differ or converge?
- What are the implications?

and finally

- What is learning?

5.1. *Headers*

To facilitate comparison between the approaches for the reader, we established the following headers as orientation for the contributors.

The underlying concept of learning

(e.g., What are the basic assumptions and presuppositions? In which paradigm is the approach embedded?)

The application of this concept

(e.g., Has the concept been applied in the field of lexical learning, phonology, morphology syntax etc. or in studies of L2 learning and multilingualism?)

Speculations about the further development of concepts of language learning

(e.g., Is there a possibility for convergence with or for divergence from concepts form another approach? Which are/will be the major issues in the future?)

Strengths and weaknesses / pros and cons of the concept

The application of this concept to atypical language development

(e.g., specific language impairment, Williams Syndrome?)

Language learning and language change

(e.g., How is language learning related to the evolution of language phylogeny?)

5.2. *Short summary of the individual papers*

As demonstrated in Section 3 of this introduction, implicit learning is highly relevant to the topic of language learning. Therefore we would

like to start this special issue with a paper on implicit learning contrasted with explicit learning.

Sabine Weinert discusses these two ways of learning based on findings from experiments with miniature languages learned by children and adults. She starts out with a brief characterization of implicit, nonconscious and automatic learning on the one hand and of explicit hypothesis testing and problem solving on the other. She points out that both kinds of learning are highly dependent on the structure of the input. One of her recent findings is that in the implicit mode the participant succeeds in learning more complex morphophonological regularities and word categories if and only if the input contains additional features which covary with the word categories and rules. Furthermore, she finds that implicit learning is modality specific. Correlated phonological cues but not visual cues facilitate the learning of morphophonological regularities. In contrast, explicit learning seems to focus more on visual-conceptual cues. In the last section, Weinert discusses the consequences of her findings for current language acquisition theories. She concludes her paper with a differentiated developmental view on the interrelationship between implicit and explicit learning.

Sonja Eisenbeiß gives a survey of previous and current generative approaches to language learning focusing on core concepts such as mentalism, representation, nativism, and modularity. In GG, language development is conceptualized as a highly selective process which is better called acquisition rather than learning. In her own feature-driven and lexicalist structure-building approach, she argues that the child instantiates grammatical features in accordance with universal restrictions on dimensions (e.g., number) and form (e.g., binarity) and integrates these features into lexical entries guided by the general cognitive principle of specificity. These recent developments in GG towards minimizing the role of domain-specific innate predispositions in favor of more general domain-unspecific principles may facilitate an approximation between GG and various other approaches to language learning which are represented in this special issue, in particular connectionist, situated and embodied approaches.

Paula Fikkert and Helen de Hoop outline an optimality-theoretic approach to language learning. In OT the learning task of children is to find the correct constraint ranking in their native language as exemplified by an optimal form or interpretation. It is widely assumed that initially markedness constraints outrank faithfulness constraints so that children start out with unmarked (and frequent) forms and only later become more and more faithful to marked (and infrequent) forms too. While their main focus lies on phonology (prosodic and segmental structure), the

authors also discuss lexical, syntactic and semantic phenomena (content of lexical representations, TNS/AGR, word order/animacy, indefinite articles and pronouns). They consider perceptual as well as production studies, arguing that it is the directionality of optimization — unidirectional from form to meaning (hearer perspective) and from meaning to form (speaker perspective) or bidirectional (taking into account both perspectives) — that explains whether comprehension precedes production or vice versa. Classical controversies on the nature of constraints (universal, gradient, functional) as well as the source of constraints (innate, learned, emergent) also exist in OT, reflecting its dual ancestry in generative and connectionist accounts and its striving for a resolution.

Barbara Höhle provides an overview of another group of approaches that developed out of generative concepts about 20 years ago: bootstrapping approaches. Researchers in this field are interested in identifying the mechanisms that help the child to get into the language system. The term *to bootstrap a system* or, abbreviated, *to boot a system* again originates from computer terminology (see Section 2.1 above) and denotes the process where an operator system is needed to load a more complex program. In language acquisition, such a bootstrapping mechanism takes cues from the input language or some knowledge previously established about the language to be learned in order to acquire further knowledge within the same domain or in a different domain. Yet the nature of these mechanisms is conceptualized in different ways, as Höhle shows. She reviews the empirical evidence for a number of bootstrapping mechanisms. In particular she focuses on prosodic bootstrapping. Since bootstrapping has highly restricted functions because it serves to get the child started, one of Höhle's conclusions is that these mechanisms can only be considered as part of a more comprehensive model of language learning, for instance of a model of distributional learning or of a model based on dynamic systems theory.

Distributional learning is also the basic type of learning for the next four groups of approaches in this special issue: the usage-based or emergentist approaches, the connectionist approaches, as well as two kinds of dynamic system approaches. Moreover, all of them share the assumptions that new structures emerge from more simple structures or via reorganization and that linguistic knowledge emerges from the child's interaction with the ambient language.

Heike Behrens reviews usage-based and emergentist approaches to language learning which have become prominent in the last 15 years. They state that language can be learned from language use by means of social skills like intention reading and a powerful generalization mechanism. Usage-based approaches go back to assumptions and findings from Cog-

nitive Linguistics, in particular they share the view that a speaker's linguistic system develops as a generalization over concrete usage events. All linguistic units are seen as form-function pairings in that there is no difference in importance between formal or semantic-pragmatic factors as driving forces for language learning. With growing experience, entrenchment leads to the automatization of frequent routines on the one hand. On the other hand, the child compares utterances to one another, finds similarities and dissimilarities, and thus categorizes the input and forms increasingly abstract linguistics. The generalization mechanism for language-specific as well as for more general patterns (rule-like behavior) is assumed to be the same. This property as well as the concept that children "start small" and make use of learning mechanisms that are attested in other domains also links the usage-based and emergentist approaches to connectionist approaches and dynamic systems theories.

Connectionist approaches to language learning are discussed by Gert Westermann, Nicolas Ruh, and Kim Plunkett. Learning in this framework is considered in terms of adapting weights of connections within a network. Since knowledge is encoded in the strengths of the connections there is no separation between knowledge and memory. Input, of course, plays a decisive role. Connectionist models allow for testing of hypotheses via simulation. The authors review a number of models that have been constructed to examine findings and hypotheses from babbling to learning syntax. The advantage of these models is that, on the one hand, these simulations allow for a high degree of specificity in testing a particular theory and that, on the other hand, the child's input and environment can be studied in great detail. The authors point out similarities and differences to related approaches, like Bayesian inference and the dynamic(al) system theory. Whether or not the differences between connectionism and DST are of a principled nature is left to future research.

This special issue comprises two dynamic systems approaches to language learning. The first one by Stephen Hockema and Linda Smith explores language learning from two directions: inside-out and outside-in. The authors review how children extract phonological, lexical, semantic and syntactic knowledge from multiple statistical and distributional regularities, including internal resources (perceptual, cognitive, developmental) and external ones (social, contextual, bodily, spatial). In the domain of word learning, the authors show how the child works through successive levels of higher-order generalizations, learning about words *as* they learn. The dynamic systems approach applied to language evolution leads to the view that individual languages lie on multiple trajectories across the language space rather than being prespecified final states. Hockema and Smith especially emphasize the unique role of the

body as the interface between the internal mental life and the external world.

In the second dynamic systems approach, Annette Hohenberger and Annemarie Peltzer-Karpf give an account of language learning against the background of developmental neuroscience. While linguistic dispositions (domain-specific and domain-general) are broadly genetically predetermined, selection and modularization result from the interplay with language experience, in accordance with principles of neural self-organization. Hence, self-organization plays a mediating role between nature and nurture. Central concepts in their approach are neural and cognitive language growth, emergence of structural representations, as well as synergy of microscopic local processes leading to macroscopic global change. Evidence for their dynamical approach comes from a chaotic itinerary to language with phases of stability and intermittent turbulence which the authors trace in the areas of lexical, morphological, and syntactic development in first- and second-language learners.

*University of Munich
Middle East Technical University, Ankara*

Notes

1. We thank Heike Behrens, Barbara Höhle, Wolfgang Klein, and Sabine Weinert for their helpful comments on an earlier version of this paper. Maura Bresnan-Enders kindly corrected our English. Correspondence address: Katrin Lindner, Germanistische Linguistik, Ludwig-Maximilians-Universität München, Schellingstr. 3, 80799 München, Germany. E-mail: katrin.lindner@germanistik.uni-muenchen.de.
2. We will not be able to summarize all the theoretical distinctions common in developmental psychology, so the interested reader is therefore referred to Goswami (2002). However, this introductory paper is still written from a more (psycho-) linguistic stance and thus attempts to summarize the discussions in this field rather than in developmental psychology. For ease of reading, concepts are indicated with single quotation marks only when they are first discussed.
3. In terms of concepts however, as Sabine Weinert (personal communication) pointed out to us, development and 'learning' have been kept distinct. In a more narrow definition of development only universal, species-specific and directed change is included. This narrow concept, however, has been modified in the developmental psychology of life spans (cf. also Oerter and Montada 2002).
4. Compare the following sentences:
 - (i) He did not acquire/ never acquired the language.
He did not learn/ never learned the language.
 - (ii) He is *acquiring/ learning a language.
 - (iii) He acquired the language within a week/ ?? within four years/ *for three years but never felt he could speak it.
He learned the language within a week/ within four years/ for three years but never felt he could speak it.

5. However, being a student of Plato, Aristotle maintained that sensual data has to be worked upon by the rational mind (through language and logic) in order to gain form and conceptual coherence. In this respect Aristotle was also a rationalist. Yet empiricists usually consider him their ancestor.
6. A loose translation of *tabula rasa* is the ‘blank slate’ as (Pinker 2002: 5) calls it. In his book with the identical title Pinker criticizes this concept which gave rise to another wave of discussion in the nature-nurture literature.
7. For further details compare the discussion in Mowrer and Klein (2001).
8. In the preface to a reprint of the review Chomsky (1967: 142f) states that he intended the review not specifically as a criticism of Skinner’s approach to language but rather as a more general critique of behaviorists’ assumptions on higher mental processes. Critical responses to Chomsky’s review are extremely rare. For detailed criticism see MacCorquodale (1970). For a critical review of Chomsky’s positions up to 1980 from a behaviorist perspective cf. Palmer ([1981] 2000, 2000). We thank Wolfgang Klein for pointing this reference out to us.
9. According to Gentner (2003) these underlying learning mechanisms are the reason “why we’re so smart”.
10. Error-free learning, if it exists at all, is most likely to occur with universal principles such as structural dependency. Parameterized dimensions such as binding, pro-drop, government direction, etc. are likely to exhibit at least a certain amount of error given their dependency on the PLD that the child has to scrutinize for relevant clues for the parameter. MacWhinney (2004: 895) argues that a cascade of parametric decisions is sometimes tantamount to a learning account. In this case, learning and parameter-setting become indistinguishable. See also Atkinson’s (1986) “learning residue” in the Principles and Parameters approach (Section 2.2).
11. Cf. also the calculations of transitional probability in Aslin et al. (1998: 321).
12. Within this framework a wide range of cross-linguistic comparisons have been made (compare for instance the papers in MacWhinney and Bates 1989) which show that learners solve the competition in language-specific ways. This model has also been applied successfully to data of late talkers (Thal and Flores 2001) and children with SLI (Evans and MacWhinney 1999, Lindner 2003).
13. Yet it may not be the only solution. Gentner (2003) and Tomasello (2003), for instance, also rely on analogy.
14. Along with Searle (1992) Cleeremans (1997: 199) holds that all knowledge learners possess is at least “potentially accessible to consciousness”. Therefore implicit knowledge may be so at some particular time or in some specific context.
15. In the meantime a number of other concepts have been proposed. For instance, following Callebaut (2005) modules may interact in constrained ways, yet they are assumed to be functional (and possibly also physical) units unto themselves.
16. In the literature both adjectives, *dynamic* and *dynamical* are variably used. We therefore use the “mixed” terminology *dynamic(al)*.

References

- Aslin, Richard N., Jenny R. Saffran & Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9. 321–324.
- Atkinson, Martin. 1986. Learning and models of development. In Paul van Geert (ed.), *Theory building in developmental psychology*, 105–139. Amsterdam: North Holland.

- Atkinson, Martin. 1992. *Children's syntax: An introduction to principles and parameters theory*. Cambridge, Oxford: Blackwell.
- Baars, Bernard J. 1986. *The cognitive revolution in psychology*. New York: Guilford Press.
- Baker, Carl L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533–81.
- Bates, Elizabeth & Jeffrey L. Elman. 1992. *Connectionism and the study of change*. Center for research on language, University of California, San Diego, CRL Technical Report 9202.
- Bates, Elizabeth & Judith C. Goodman. 1999. On the emergence of grammar from the lexicon. In Brian McWhinney (ed.), *The emergence of language*, 29–79. Mahwah, NJ: Erlbaum.
- Bates, Elizabeth & Brian MacWhinney. 1982. Functionalist approaches to grammar. In Eric Wanner and Lila Gleitman (eds.), *Language acquisition: The state of the art*, 203–367. New York: Cambridge University Press.
- Beer, Randall D. 2000. Dynamical approaches to cognitive science. *Trends in Cognitive Sciences* 4(3), 91–99.
- Berry, Diane (ed.). 1997. *How implicit is implicit learning?* Oxford: Oxford University Press.
- Brent, Michael R. & Timothy A. Cartwright. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61, 93–125.
- Broeder, Peter & Jaap Murre. 2000. *Models of language acquisition: Inductive and deductive approaches*. Oxford: Oxford University Press.
- Brown, Colin M. & Peter Hagoort. 1999. *The neurocognition of language*. Oxford: Oxford University Press.
- Callebaut, Werner. 2005. The ubiquity of modularity. In Werner Callebaut & Diego Rasskin-Gutman (eds.), *Modularity: Understanding the development and evolution of natural complex systems*, 3–28. Cambridge, MA: MIT Press.
- Canfield, Richard L. & Marshall M. Haith. 1991. Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology* 27, 198–208.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Chomsky, Noam. 1959. A review of B. F. Skinner's *Verbal behavior*. *Language* 35, 26–58.
- Chomsky, Noam. 1967. A review of B. F. Skinner's *Verbal behavior*. In Leon Jakobovits & Murray S. Miron (eds.), *Readings in the psychology of language*, 142–171. Englewood Cliffs, NJ: Prentice Hall.
- Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Dordrecht: Foris.
- Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Christiansen, Morton H., Joseph Allen & Mark S. Seidenberg. 1998. Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes* 13, 221–268.
- Cleeremans, Axel. 1997. Principles for implicit learning. In Diane Berry (ed.), *How implicit is implicit learning?*, 195–234. Oxford: Oxford University Press.
- Cleeremans, Axel, Arnaud Destrebecqz & Boyer, Maud. 1998. Implicit learning: News from the front. *Trends in Cognitive Sciences* 2(10), 406–416.
- Cleeremans, Axel & Luis Jiménez. 2002. Implicit learning and consciousness: A graded, dynamic perspective. In Robert M. French and Axel Cleeremans (eds.), *Implicit learning and consciousness: An empirical, philosophical and computational consensus in the making*, 1–40. New York & Hove: Psychological Press.

- Cohen, Jonathan D., Kevin Dunbar & James L. McClelland. 1990. On the control of automatic processes: A parallel distributed account of the Stroop effect. *Psychological Review* 97. 332–361.
- Dehaene-Lambertz, Ghislaine, Stanislas Dehaene & Lucie Hertz-Pannier. 2002. Functional neuroimaging of speech perception in infants. *Science* 298. 2013–2015.
- Dehaene-Lambertz, Ghislaine, Lucie Hertz-Pannier, Jessica Dubois, Sébastien Mériaux, Alexis Roche, Mariano Sigman & Stanislas Dehaene. 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Science* 103. 13240–14245.
- Dunn, John C. & Kim Kirsner. 1988. Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review* 95. 91–101.
- Eimas, Peter. 1999. Do infants learn grammar with algebra or statistics? *Science* 284. 435–436.
- Elman, Jeffrey L. 2005. Connectionist models of cognitive development: Where next? *Trends in Cognitive Sciences* 9. 111–117.
- Elman, Jeffrey L., Elizabeth Bates, Mark Johnson, Annette Karmiloff-Smith, Dominique Parisi & Kim Plunkett. 1996. *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.
- Evans, Julia & Brian MacWhinney. 1999. Sentence processing strategies in children with expressive and expressive-receptive specific language impairments. *International Journal of Language and Communication Disorders* 34. 117–134.
- Farah, Maratha J. 1994. Neuropsychological inference with an interactive brain: A critique of the ‘locality’ assumption. *Behavioral and Brain Sciences* 17. 43–104.
- Fodor, Jerry A. 1980. On the impossibility of acquiring “more powerful” structures. In Massimo Piattelli-Palmarini (ed.), *Language and learning: The debate between Jean Piaget and Noam Chomsky*, 142–162. Cambridge, MA: Harvard University Press.
- Fodor, Jerry A. 1983. *The modularity of mind*. Cambridge, MA: MIT Press.
- Friederici, Angela D. 2005. Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences* 9. 481–488.
- van Geert, Paul. 1986. The concept of development. In Paul van Geert (ed.), *Theory building in developmental psychology*, 3–50. Amsterdam: North Holland.
- van Geert, Paul. 1997. Que será, será: Determinism and nonlinear dynamic model building in development. In Alan Vogel, Maria C. D. P. Lyra & Jaan Valsiner (eds.), *Dynamics and indeterminism in developmental and social processes*, 13–38. Mahwah, NJ: Lawrence Erlbaum.
- van Gelder, Tim. 1990. Compositionality: A connectionist variation on a classical theme. *Cognitive Science* 14. 355–384.
- van Gelder, Tim. 1998. The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences* 21. 615–665.
- Gentner, Dedre. 2003. Why we’re so smart. In Dedre Gentner and Susan Goldin-Meadow (eds.), *Language in mind*, 195–235. Cambridge, MA: MIT Press.
- Glaser, Robert. 1990. The reemergence of learning theory within instructional research. *American Psychologist* 45. 29–39.
- Gold, E. Mark. 1967. Language identification in the limit. *Information and Control* 10. 447–474.
- Gomez, Rebecca L. & LouAnn Gerken. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4. 178–186.
- Gordon, William C. 1989. *Learning and memory*. Pacific Grove, CA: Brooks/Cole.
- Goswami, Usha. 2002. *The Blackwell handbook of childhood cognitive development*. London: Blackwell.

- Halliday, Michael A. K. 1975. *Learning how to mean: Explorations in the development of language*. London: Arnold.
- Hauser, Marc D., Elissa L. Newport & Richard N. Aslin. 2001. Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition* 78. B53–B64.
- Hauser, Marc D., Daniel Weiss, & Gary Marcus. 2002. Rule learning by cotton-top tamarins. *Cognition* 86. B15–22.
- Hebb, Donald O. 1949. *The organization of behavior: A neuropsychological theory*. New York: John Wiley.
- Hennon, Elizabeth, Kathy Hirsh-Pasek, & Roberta Michnick Golinkoff. 2000. Die besondere Reise vom Fötus zum spracherwerbenden Kind. In Hannelore Grimm (ed.), *Enzyklopädie der Psychologie*, Band 3.3, *Sprachentwicklung*, 41–103. Göttingen: Hogrefe.
- Höhle, Barbara, Jürgen Weissenborn, Dorothea Kiefer, Ante Schulz & Michaela Schmitz. 2004. Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy* 5(3). 341–353.
- Hohenberger, Annette. 2002. Functional categories in language acquisition: Self-organization of a dynamical system. (Linguistische Arbeiten 456). Tübingen: Niemeyer.
- Hornstein, Norbert & David Lightfoot. 1981. *Explanation in Linguistics: The Logical Problem of Language Acquisition*. London: Longmans.
- Hunt, Ruskin H. & Richard N. Aslin. 2001. Statistical learning in a serial reaction time task: Access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General* 130. 658–680.
- Imada, Toshiaki, Yang Zhang, Marie Cheour, Samu Taulu, Antti Ahonen, & Patricia K. Kuhl. 2006. Infant speech perception activates Broca's area: A developmental magnetoencephalography study. *NeuroReport* 17(10). 957–962.
- Jiménez, Luis (ed.). 2003. *Attention and implicit learning*. Amsterdam: John Benjamins.
- Jiménez, Luis, Cástor Méndez & Axel Cleeremans. 1996. Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 948–969.
- Johnson, Mark H. 2005. *Developmental cognitive neuroscience*, 2nd edn. Malden, NJ: Blackwell.
- Jordan, Geoff. 2004. *Theory construction in second language acquisition*. Amsterdam: John Benjamins.
- Jusczyk, Peter W., Anne Cutler & Nancy Redanz. 1993a. Preference for the predominant stress patterns of English words. *Child Development* 64. 675–687.
- Jusczyk, Peter W., Angela D. Friederici, Jeanine M. I. Wessels, Vigdis Y. Svenkerud, & Ann Marie Jusczyk. 1993b. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32. 402–420.
- Jusczyk, Peter W., Paul A. Luce, & Jan Charles-Luce. 1994. Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* 33. 630–645.
- Krashen, Stephen D. 1987. *Principles and practice in second language acquisition*. London: Prentice-Hall International.
- Krashen, Stephen D. 1988. *Second language acquisition and second language learning*. London: Prentice-Hall International.
- Kuhl, Patricia K. 2000. A new view of language acquisition. *Proceedings of the National Academy of Science* 97. 1185–1187.
- Kuhl, Patricia K. 2004. Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5. 831–843.
- Lindner, Katrin. 2003. The development of sentence interpretation strategies in monolingual German-learning children with and without specific language impairment. *Linguistics* 41. 213–254.

- MacLeod, Colin M. & Kevin Dunbar. 1988. Training and stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology* 14. 126–135.
- MacWhinney, Brian (ed.). 1999. *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, Brian. 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language* 31. 883–914.
- MacWhinney, Brian & Elizabeth Bates. 1989. *The cross-linguistic study of sentence processing*. New York: Cambridge University Press.
- Marchman, Virginia & Elizabeth Bates. 1994. Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language* 21. 339–366.
- Marcus, Gary F. 1999. Connectionism: With or without rules? *Trends in Cognitive Sciences* 3. 168–170.
- Marcus, Gary F. 2001. *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.
- Marcus, Gary F., Sugumaran Vijayan, Shoba Bandi Rao & Peter M. Vishton. 1999. Rule learning by seven-month-old infants. *Science* 283. 77–80.
- MacCorquodale, Kenneth. 1970. On Chomsky's review of Skinner's *Verbal behavior*. *Journal of the Experimental Analysis of Behavior* 13. 83–99.
- McClelland, James L. & David C. Plaut. 1999. Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences* 3. 166–168.
- McLeod, Peter, Kim Plunkett & Edmund T. Rolls. 1998. *Introduction to connectionist modelling of cognitive processes*. Oxford: Oxford University Press.
- Mowrer, Robert R. & Stephen B. Klein. 2001. The transitive nature of contemporary learning theory. In Robert R. Mowrer and Stephen B. Klein (eds.), *Handbook of contemporary learning theories*, 1–21. Mahwah, NJ: Lawrence Erlbaum.
- Nakisa, Ramin C. & Ulrike Hahn. 1996. Where defaults don't help: The case of the German plural system. In G. W. Cottrell (ed.), *Proceedings of the 18th annual conference of the cognitive science society*, 177–182. Malden, NJ: Lawrence Erlbaum.
- Oerter, Rolf & Leo Montada (eds.). 2002. *Entwicklungspsychologie*. 5th revised edn. Weinheim: Beltz.
- Palmer, David. 2000 [1981]. Chomsky's nativism: A critical review. *The Analysis of Verbal Behavior* 17. 39–50.
- Palmer, David. 2000. Chomsky's nativism reconsidered. *The Analysis of Verbal Behavior* 17. 51–56.
- Pavlov, Ivan P. 1928. *Lectures on conditioned reflexes*, trans. by W. H. Gantt. New York: International Publishers.
- Pena, Marcela, Atsushi Maki, Damir Kovacic, Ghislaine Dehaene-Lambertz, Hideaki Koizumi, Furio Bouquet & Jacques Mehler. 2003. Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences* 100. 11702–11705.
- Piattelli-Palmarini, Massimo. 1989. Evolution, selection, and cognition: From 'learning' to 'parameter-setting' in biology and the study of language. *Cognition* 31. 1–44.
- Pinker, Steven. 2002. *The blank slate: The denial of human nature in modern intellectual life*. London: Penguin.
- Plaut, David C. 1995. Double dissociation without modularity: Evidence from connectionist neuropsychology. *Journal of Clinical and Experimental Neuropsychology* 17. 291–326.
- Plunkett, Kim. 2001. Connectionism today. *Synthese* 129. 185–194.
- Pullum, Geoffrey K. 2003. Learnability. *The Oxford international encyclopedia of linguistics*, 2nd edn., 431–434. Oxford: Oxford University Press.
- Pullum, Geoffrey K. & Barbara Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19. 9–50.

- Pulvermüller, Friedemann. 1997. Aspects of language mechanisms: A Hebbian perspective. *European Review* 5. 23–37.
- Pulvermüller, Friedemann. 1999. Words in the brain's language. *Behavioral and Brain Sciences* 22. 253–336.
- Pulvermüller, Friedemann. 2001. Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5. 517–524.
- Pulvermüller, Friedemann. 2005. Brain mechanisms linking language and action. *Nature Review Neuroscience* 6. 576–582.
- Ramus, Franck, Marc D. Hauser, Cory D. Miller, Dylan Morris & Jacques Mehler. 2000. Language discrimination by human newborns and by cotton-top mandarin monkeys. *Science* 288. 349–351.
- Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6 (6). 855–863.
- Reber, Arthur. 1993. *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York & Oxford: Clarendon Press.
- Redington, Martin & Nick Chater. 1997. Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences* 1. 273–281.
- Russell, James. 2004. *What is language development?* Oxford: Oxford University Press.
- Saffran, Jenny R. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44. 493–515.
- Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996a. Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.
- Saffran, Jenny R., Elissa L. Newport & Richard N. Aslin. 1996b. Word segmentation: the role of distributional cues. *Journal of Memory and Language* 35. 606–621.
- Savelsbergh, Geert J. P. 2003. Development and learning: A tautology or dichotomy? *Infant Behavior and Development* 26. 439–400.
- Scholz, Barbara C. & Geoffrey K. Pullum. 2006. Irrational nativist exuberance. In Robert J. Stainton (ed.), *Contemporary debates in cognitive science*, 59–80. Malden, MA: Blackwell.
- Searle, John R. 1992. *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Seidenberg, Mark S. & Jeffrey L. Elman. 1999a. Do infants learn grammar with algebra or statistics? *Science* 284. 435–436.
- Seidenberg, Mark S. & Jeffrey L. Elman. 1999b. Networks are not 'hidden rules'. *Trends in Cognitive Sciences* 3. 288.
- Singer, Wolf & Charles M. Gray. 1995. Visual feature integration and the temporal correlation hypothesis. *Annual Review Neuroscience* 18. 555–586.
- Skinner, Burrhus F. 1938. *The behavior of organisms: An experimental analysis*. New York: Appleton-Century-Crofts.
- Skinner, Burrhus F. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts.
- Smith, Linda B. & Larissa K. Samuelson. 2003. Different is good: connectionism and dynamic systems theory are complementary emergentist approaches to development. *Developmental Science* 6(4). 434–439.
- Stadler, Michael & Peter Frensch. 1998. *Handbook of implicit learning*. Thousand Oaks, CA: Sage.
- Thal, Donna & Melanie Flores. 2001. Development of sentence interpretation strategies by typically developing and late-talking toddlers. *Journal of Child Language* 28. 173–193.
- Thelen, Esther; and Bates, Elizabeth. 2003. Connectionism and dynamic systems: are they really different? *Developmental Science* 6(4). 378–391.
- Thorndike, Edward L. 1913. *Educational psychology: The psychology of learning*. New York: Teachers College Press.
- Thorndike, Edward L. 1932. *The fundamentals of learning*. New York: Teachers College Press.

- Tomasello, Michael. 2000a. Do your children have adult syntactic competence? *Cognition* 74. 209–253.
- Tomasello, Michael. 2000b. First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics* 11. 61–82.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based account of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, Michael & Dan I. Slobin (eds.). 2004. *Beyond nature-nurture: Essays in honor of Elizabeth Bates*. Mahwah, NJ: Lawrence Erlbaum.
- Watson, John B. 1913. Psychology as the behaviorist views it. *Psychological Review* 20. 158–177.
- Watson, John B. 1930. *Behaviorism*, revised edn. Chicago: University of Chicago Press.
- Wexler, Kenneth. 1999. Maturation and growth of grammar. In William C. Ritchie & Tej K. Bhatia (eds.), *Handbook of language acquisition*, 55–109. Amsterdam: Elsevier.
- Wexler, Kenneth & Peter W. Culicover. 1980. *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
- Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8(10). 451–456.