Francis Lowenthal · Laurent Lefebvre
Editors

Language and Recursion

Springer

Language and Recursion

Francis Lowenthal • Laurent Lefebvre
Editors

Language and Recursion

Springer

Editors
Francis Lowenthal
Department of Cognitive Sciences
University of Mons
Mons, Belgium

Laurent Lefebvre
Department of Cognitive Sciences
University of Mons
Mons, Belgium

ISBN 978-1-4614-9413-3 ISBN 978-1-4614-9414-0 (eBook)
DOI 10.1007/978-1-4614-9414-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2013953263

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Are There Any Relations Between Recursion and Language?

Hauser, Chomsky & Fitch (2002) claimed that the faculty to use recursive processes is the unique difference between human and animal communication. This position is contested by certain authors (Pinker & Jackendoff, 2004; Gervain, Macagno, Cogoi, Peña, & Mehler, 2008; Hochmann, Azadpour, & Mehler, 2008) but approved by others (Premack, 2004). Fitch and Hauser (2004) showed that tamarin apes and human adults react in the same way to violations of a Finite State Grammar (FSG—a grammar based on concatenation) but that tamarin apes do not react to violations of a Phrase Structure Grammar (PSG—a grammar based on non-tail recursive hierarchies) and claimed that this was evidence that nonhuman animals do not have access to recursive processes. Gentner, Fenn, Margoliash, & Nusbaum (2006) worked with songbirds and claimed the contrary, but Corballis (2007) considered that none of these papers provided us with conclusive evidence of the presence of recursion in nonhuman animals.

Nevertheless, the above mentioned authors did never specify which exact definition of "recursion" they use.

All these data suggest that the possible relation between recursion and language deserves more than a little attention. This attention should also be used to try to agree on a common definition of recursion.

Friederici and her colleagues (Opitz & Friederici, 2004; Frederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Makuuchi, Bahlmann, Anwander, & Friederici, 2009) showed also that violations of PSG do not trigger exactly the same brain regions as violations of FSG grammars.

Other authors contest the "innate approach" used by Chomsky and choose an approach based on the principle that "language is acquired" (Saffran, Aslin, & Newport, 1996). Some of them (Seidenberg, 1997; Altmann & Dienes, 1999) suggested that the best model to describe language acquisition is not Chomsky's

Standard Theory model but a connectionist approach and they suggest using neural networks. This point of view can be associated to the point of view adopted by those working on Implicit Learning. Here again it must be noted that during the learning phase, recursion plays an important role.

Independently, we developed the nonverbal communication device (NVCD) approach (Lowenthal, 1983, 1986) which can be associated to Implicit Learning. In this framework we created exercises that imply the manipulation of recursive structures, at least the manipulation of embedded sequences. Young subjects using this approach acquire a richer language structure or reacquire such a structure in the case of cerebral lesions (Lowenthal & Saerens, 1986). This led us to formulate the hypothesis that these approaches favor the emergence of new cerebral networks in patients with cerebral lesions (Lowenthal, 1999). Lefebvre, Baleriaux, Paquier, & Lowenthal (2006) proved a milder version of this hypothesis: they showed, using an fMRI observation, that the use of an NVCD-type approach can have an influence on the basal ganglia in language tasks. We suggest that the exercises we used lead to the elaboration of a regular language, which does not require full recursion. If this claim is confirmed by further researches combining the use of NVCD-like approaches, neuropsychological observations, psychometric methods, and neural networks, it might reconcile Chomsky's theory with psychological observations.

For all these reasons, it appeared useful to organize a **multidisciplinary** conference on language and recursion where all the main researchers in the field would be invited. The invited participants belonged to very different domains: from computer sciences to neuropsychology, from mathematics to linguistics, and from animal cognition to neural networks.

The aim of the organizers of this 3-day conference was less to give to each speaker the opportunity to present his or her point of view, but to provoke long and fruitful exchanges and discussions.

This book is the result of this conference. It is **not** a "book of proceedings," but a group of chapters written, several months after the conference was finished, by selected participants to the conference. It presents thus the present state of their researches, inspired by the long and rich discussions we had during the conference. These chapters do represent what the authors think **now** would be relevant, in some cases taking in account the most recently published results in the literature (e.g., Martin-Ordas, Berntsen, & Call, 2013).

This set of chapters is completed by a synthesis of the main discussions held during the 3 days of the conference, during the official sessions, the coffee breaks, and the late evening bar conversations: as we said before, these discussions represented the most important part of our meeting.

Mons, Belgium Francis Lowenthal

References

Altmann, G. T. M., & Dienes, Z. (1999). Rule learning by seven-month-old infants and neural networks. *Science, 284*, 875.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science, 31*(4), 697–704.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a non-human primate. *Science, 303*, 377–380.

Frederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*(7), 2458–2463.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204–1207.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences, 105*(37), 14222–14227.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve ? *Science, 298*, 1569–1579.

Hochmann, J.-R., Azadpour, M., & Mehler, J. (2008). Do humans really learn A^nB^n artificial grammars from exemplars? *Cognitive Science, 32*, 1021–1036.

Lefebvre, L., Baleriaux, D., Paquier, P., & Lowenthal, F. (2006). Basal ganglia: A crossroad between verbal and non-verbal activities? *Acta Neurologica Belgica, 106*, 65–66.

Lowenthal, F. (1983). Strategy games, winning strategies, and Non-Verbal Communication Devices (at the age of 8). In R. Hershkowitz (Ed.), *Proceedings of the seventh International Conference for the Psychology of Mathematics Education* (pp. 364–368). Weitzmann Institute of Science: Rehovot.

Lowenthal, F. (1986). NVCDs are structuring elements. In L. Burton & C. Hoyles (Eds). *Proceedings of the tenth International Conference for the Psychology of Mathematics Education* (pp. 363–368). London.

Lowenthal, F. (1999). Can handicapped subjects use perceptual symbol systems? *Behavioral and Brain Sciences, 22*(4), 625–626.

Lowenthal, F., & Saerens, J. (1986). Evolution of an aphasic child after the introduction of NVCDs. In F. Lowenthal & F. Vandamme (Eds.), *Pragmatics and education* (pp. 301–330). New York: Plenum.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences, 106*, 8362–8367.

Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology, 23*, 1–4.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neurosciences, 24*, 8436–8440.

Pinker, S., & Jackendoff, R. (2004). The faculty of language: What's special about it? *Cognition, 95*, 201–236.

Premack, D. (2004). Is Language the key to human intelligence? *Science, 303*, 318–320.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Sciences, 274*, 1926–1928.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

About the Editors

Dr. Francis Lowenthal received his PhD from the Massachusetts Institute of Technology, and is the current head of the Cognitive Sciences Department at the University of Mons, Belgium. He is the director of the Cognitive Sciences Laboratory, and is the Chairman of the Council of the Institute of Language Sciences.

Dr. Laurent Lefebrve is an Assistant Professor in the Cognitive Sciences Department at the University of Mons, and Chair of the Graduate Studies in the School of Psychology and Education, University of Mons.

Contents

Contributors

Michael C. Corballis University of Auckland, Auckland, New Zealand

Z. Dienes University of Sussex, Brighton, United Kingdom

Daniel M. Dubois HEC-ULg, Management School of the University of Liege, Liege, Belgium

ASBL CHAOS, Centre for Hyperincursion and Anticipation in Ordered Systems, Institute of Mathematics B37, Liege, Belgium

William Tecumseh Fitch Department of Cognitive Biology, University of Vienna, Vienna, Austria

Sandiway Fong Departments of Linguistics and Computer Science, University of Arizona, Tucson, AZ, USA

Philippe Fortemps Computer Science, Management and Engineering, University of Mons, Mons, Belgium

Pierre Frath Université de Reims Champagne-Ardenne, Reims, France

Robert Freidin Princeton University, Princeton, NJ, USA

Angela D. Friederici Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Q. Fu Institute of Psychology, Chinese Academy of Sciences, Beijing, China

Judit Gervain Laboratoire Psychologie de la Perception, CNRS and Université Paris Descartes, Sorbonne Paris Cité, France

X. Guo Shanghai Key Laboratory of Magnetic Resonance and School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

Laurent Lefebvre Department of Cognitive Sciences, Research Institute for Health sciences and Technology, University of Mons, Mons, Belgium

Alban Lemasson Animal and Human Ethology, Université de Rennes 1 – CNRS, Rennes, France

Stephen C. Levinson Max Planck Institute, Nijmegen, The Netherlands

Donders Institute, Radboud University, Nijmegen, The Netherlands

Francis Lowenthal Department of Cognitive Sciences, Institute of Language Sciences, University of Mons, Mons, Belgium

Maurício Dias Martins Department of Cognitive Biology, University of Vienna, Vienna, Austria

Language Research Laboratory, Lisbon Faculty of Medicine, Lisbon, Portugal

M. Rohrmeier Department of Linguistics & Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA

Roger Vergauwen Centre for Logic and Analytic Philosophy, Institute of Philosophy, Katholieke Universiteit Leuven, Leuven, Belgium

Vincent Wautié Computer Science, Management and Engineering, University of Mons, Mons, Belgium

Klaus Zuberbühler School of Psychology and Neuroscience, University of Saint Andrews, Fife, Scotland, UK

Cognitive Science Centre, University of Neuchâtel, Neuchâtel, Switzerland

List of Figures

List of Table

Part I
What Is Recursion?

Chapter 1
Pragmatics as the Origin of Recursion

Stephen C. Levinson

Introduction

Hauser, Chomsky, and Fitch (2002) speculate that perhaps the sole feature of language that may be domain specific is the recursive nature of syntax. The implication is that it was the evolution of this syntactic ability that accounts for the species-unique character of human language. This chapter sets out a rival possibility, namely, that the focal type of recursion—understood here as centre embedding—has its natural home in principles of language use, not language structure.

The different senses of 'recursion', and the formal characterization of each of them, are amply discussed elsewhere in this volume. One notion in particular has played a central role in discussion, namely, centre embedding where one clause is embedded within another, as in *The rat the cat killed ate the malt* (Chomsky & Miller, 1963), a pattern isomorphic with a mirror language like ABBA or ABCCBA where there are nested dependencies. Like the 'counting language' (AAABBB), nested dependencies are—if of unrestricted depth—the stigmata of context-free languages. From a comparative linguistic point of view, central embedding is particularly interesting compared to edge recursion (as in *John thought he'd come*) since it is more easily distinguished from strings of sentences without an embedding relation (a practical problem rehearsed below). Hence centre embedding will have a central place in these remarks.

It is worth emphasizing that no string set can be assigned unambiguously to the context-free languages unless there is evidence of indefinite recursion.

S.C. Levinson (✉)
Max Planck Institute, Nijmegen, The Netherlands

Donders Institute, Radboud University, Nijmegen, The Netherlands
e-mail: Stephen.Levinson@mpi.nl

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_1,

Besides, Chomsky soon realized that string patterns themselves are not the proper objects of study: what is psychologically interesting is the *structure* we map on the strings, which is nearly always underdetermined by the strings alone.

Pragmatics and Embedded Construals

At the outset of generative grammar (Chomsky, 1957), the close relation was noted between sentence (1a) and the mini-discourse in (1b)—indeed the proposal was that (1a) was derived from (1b) by syntactic 'generalized transformation':

(1a) The ship which was the largest of its kind sank in high seas.
(1b) The ship sank in high seas. It was the largest of its kind.

That model of syntax is long gone, but the insight that strings of sentences are construed in complex ways remains and the striking systematicity of some of this has been explored in Gricean pragmatics (e.g. Levinson, 2000) or rhetorical structure theory (Mann & Thompson, 1988). For example, (2a) is naturally construed as (2b) and (2c) as (2d):

(2a) Buy a ticket. Win a thousand dollars!
(2b) If you buy a ticket, you can win a thousand dollars!
(2c) Sue screamed. Bill left.
(2d) Because Sue screamed, Bill left.
(2e) John may have been partly at fault. But Bill must take the blame.
(2f) Although John may have been partly at fault, Bill must take the blame.

Note how these construals mirror the more complex syntax of English, with relative clauses (as in (1)), conditionals (as in 2b), causal subordination (as in 2d) and concessives (as in 2f). We rely on these interpretations all the time in our understanding of texts (consider *veni*, *vidi*, *vici*).

Although no one has done the careful work to actually establish this, these kinds of construals seem universally available. Certainly I have encountered them in languages as diverse as Guugu Yimithirr, Tzeltal, Tamil or Yelî Dnye. It is the widespread availability of such understandings that makes it possible for many languages to simply not provide a conditional or relative clause (e.g. Guugu Yimithirr) or a disjunction (Tzeltal) or a concessive or causal connective (Yelî Dnye). We turn now to the case of languages that seem to offer quite restricted possibilities of embedding in their syntax.

Languages with Restricted Embedding

Linguistic typologists are well aware that many languages show little evidence of indefinite embedding.[1] Recently, Pirahã has been a focus of debate, with the original fieldworker (Everett, 2005) claiming no evidence at all for recursive structures

[1] See, for example, the discussion of Amele in Comrie & Kuteva (2008).

and generativist reanalysis suggesting that embedding may in fact be evidenced (Nevins, Pesetsky, & Rodrigues, 2009). Analysis hinges on the distinction between embedding and parataxis and on whether (3) should be analysed as (a) (Everett) or (b) (Nevins et al.):

(3) *Hi xob-aʹaxaʹıʹ. Hi kahaıʹ kai-sai.*
He see-well. He arrow make-OLD.INFO
a.Everett: ‘He is really smart. He makes arrows (as we were saying)’
b.Nevins et al. ‘He is really good [COMP at making arrows]’

What is not in doubt, however, is that embedding is very limited and at most seems capped at one level deep.

As discussed above, it is the unlimited character of nested dependencies that is relevant for the theoretical issues. But in lacking evidence of indefinite recursion, Pirahã is not unique at all. The Australian languages provide a wealth of well-documented cases. As Hale (1976) pointed out,

> *In a large number of Australian languages, the principal responsibility for productive recursion in syntax is shouldered by a structure which I will refer to as the adjoined relative clause. It is typically marked as subordinate in someway, but its surface position with respect to the main clause is marginal rather than embedded—hence the locution ‘adjoined’. Typically, but not invariably, it is separated from the main clause by a pause.*

A further property is that these juxtaposed sentences with the structure S1 + (particle) S2 function with a wide array of possible interpretations as relatives, temporal clauses, conditionals, etc. Hale (1976) pointed out that the Warlpiri sentence in (4) allows any of the indicated readings (the square brackets in the examples below indicate the putative embedded clause):

(4) *Ngajulu-rlu kapi-rna maliki rluwa-rni, [kaji-ngki yarlki-rni nyuntu].*
1-ERG AUX dog shoot-NP COMP?-AUX bite-NP you
a. ‘I will shoot the dog, if / …when it bites you.’
b. ‘I will shoot the dog that bites you / …that is going to bite you.’

Although Warlpiri has a particle that may be analysed as a complementizer, many Australian languages do not. It then becomes a completely live issue as to whether we are dealing with structural dependence or parataxis with ‘subordinate’-like construals. Consider the following Wambaya sentence (Nordlinger, 2006):

(5) *[Ilarri irri ngarabi] daguma irri-ngg-i.*
grog.I(ACC) 3.PL.A(NP) drink hit 3.PL.A-RR-FUT
a. ‘They’ll drink grog (and then) they’ll fight’ (coordinate construal)
b. ‘When they drink grog, they’ll fight’ (subordinate construal)

Nordlinger argues that the ‘subordinate’ construal may be forced by prosody, but as Hale noted there is often a pause between clauses of these types in Australian languages generally. It will not be easy then to come to a definitive conclusion either way, just as in the Pirahã case.

Many Australian languages nevertheless have some cases of relatively clear subordination. But in these cases indefinite embedding is hard to support, because the embedded verb typically takes a nominal case, for example, a purposive. This often constrains further embedding. Consider Kayardild (Evans, 1995) which adds an oblique case (COBL) to each of the subordinate constituents, as in (6). This case is terminal, so no further subordination is possible:

(6)	*Dan-da*	*banga-a*	*[kakuju-ntha*	*ngijuwa*	*raa-jarra-ntha*	*walbu-nguni-nj]*
	This-NOM	turtle-NOM	uncle-COBL	3rdSUB.COBL	spear-PST-COBL	raft-INSTR-COBL

‘This is the turtle [uncle speared from the raft]’

It is thus not possible to add, say, a relative clause to ‘the raft’. Kayardild consequently systematically blocks recursion at one level deep. In general, polysynthetic languages show very restricted levels of embedding (see Evans & Levinson, 2009). And, in the opposite direction, languages with very limited morphology often offer no clear evidence for subordination at all (see, e.g. Englebretson (2003) on Indonesian). Pirahã is thus not an isolated case.

A frequent response to these sorts of findings is to invoke the metaphor of UG as a ‘toolkit’ whose tools may not be all deployed (as in Jackendoff, 2002): ‘the putative absence of obvious recursion in one of these languages is no more relevant to the human ability to master recursion than the existence of three vowel languages calls into doubt the human ability to master a five- or ten-vowel language’ (Fitch, Hauser, & Chomsky, 2005). But this sits uncomfortably with the claim (Hauser et al., 2002) with which we began, namely, that ‘recursion’ (understood as embedding) may be the one crucial domain-specific feature of linguistic ability: such a crucial design feature ought to be evidenced in any language system.

Centre Embedding in Syntax

It has long been noted that there are comprehension problems associated with repeated centre embeddings. Chomsky and Miller (1963) said of the sentence *The rat [the cat [the dog chased] killed] ate the malt* that it is ‘… surely confusing and improbable but it is perfectly grammatical and has a clear and unambiguous meaning’. They assumed that such sentences are licensed grammatically but run up against performance processing difficulties. There have been numerous theories since about why exactly the processing is difficult, but all revolve around short-term memory limitations (Folia et al., 2011; Gibson, 1991, 1998; Kimball, 1973; Perfors, Tenenbaum, Gibson, & Regier, 2010; Weckerly & Elman, 1992). Gibson (1998; Gibson & Thomas, 1999), for example, suggested that the problem not only involves keeping track of a number of unfulfilled dependencies but also follows a locality metric: hence nested dependencies three or more deep are more difficult than cross-serial dependencies (Bach, Brown, & Marslen-Wilson, 1986), where the dependencies are serially and more locally discharged (see De Vries, Petersson, Geukes, Zwitserlood, & Christiansen, 2012 for recent confirmation). These studies repeatedly show severe performance difficulties at three levels of embedding or higher (Marks, 1968), allowing a connectionist account of performance (Christiansen & Chater, 1999).

Karlsson (2007) examined corpora in seven European languages (English, German, Finnish, French, Latin, Swedish, Danish). He found that in the Brown corpus of English written texts, 57 % of clauses have embeddings, of which 76 % were final, 13 % were initial and 11 % were centre embeddings (mostly relative clauses). This seems to be the general pattern at least for familiar languages of similar word order, but polysynthetic languages show a much lower incidence of embedding (e.g. 7 % for Mohawk, 6 % in Gunwinggu and just 2 % in Kathlamet; Mithun, 1984). Centre embeddings can be classified as degree 1 (one embedding), degree 2 (embedding within an embedding) and degree 3 (embedding within an embedding within an embedding). The following gives a (simplified) example of Karlsson's coding:

(7) Karlsson (2007)

1	*"If*	← degree 1 subordinate clause
	2 *as often happened*	← degree 2 center-embedding
1	*she asked him*	
	2 *to tell her about it*	← degree 2 complement
0	*she thought*	← matrix-clause (degree 0)
1	*that he*	← degree 1 complement
	2 *who had been so kind*	← degree 2 center-embedding
1	*would understand* "	

No examples of degree 3 embedding were found in corpora, although from hand-annotated historical texts from his and other earlier compilations, a total number of 13 cases have been found in the whole of Western literature. He therefore observes that the maximal degree of multiple centre embedding is three in written language. For spoken language, no cases at all have been found, and only three cases of degree 2 have been found, from which he concludes that degree 2 is the upper bound for spoken language. These findings are of course interesting, since they undermine the idea that natural languages are not regular and necessarily context-free or higher—it remains an interesting question whether treating, say, English as regular (with large numbers of simple rules) is more complex than treating it as context-free (with less more complex rules; see Perfors et al., 2010 on such trade offs).

The psycholinguistic findings and the corpus findings converge: after degree 2 embedding, performance rapidly degrades to a point where degree 3 embeddings hardly occur.

Centre Embedding in Interactive Discourse

We are now in a position to appreciate a rather surprising phenomenon.[2] There are embeddings in interactive discourse that have the same basic properties exhibited in sentential syntax, but which are distributed over two or more speakers. Yet in this case there is no similar limit on embedding—multiple embeddings seem in principle indefinite, certainly at least of degree six.

The basic phenomenon is illustrated in 8. (Examples will be drawn largely from interaction in service encounters, e.g. from Merritt (1976), as these l end themselves to brief exposition.)

(8) A: "May I have a bottle of Mich?"
B: "Are you twenty one?" ← degree 1 center-embedding
A: "No" ← response at degree 1
B: "No"

Clearly, in this interchange the second question leaves the first unanswered until a preliminary question is addressed, which then allows the answer to the first question to be subsequently provided. The question–answer pair in the middle forms an island over which a discontinuous dependency is maintained. In these kinds of insertion sequences, paired utterances are embedded at the same level together. We have here a nested dependency just as in *The boy the horse kicked has a broken leg*. Sequences of this type $Q_1Q_2A_2A_1$ belong squarely in the class of the counting or mirror languages, the prototypes of context-free languages. A context-free grammar that would generate strings Q^nA^n indefinitely might have the rules Q&A $\rightarrow$ Q (Q&A) A, Q&A $\rightarrow$ Q A.

Like nearly all the demonstrations of context-freeness in syntax, the assignment of structure to utterances in these cases is relative to a construal. In this case the construal depends not on the syntax and semantics so much as the speech act or illocutionary force: regardless of form or semantic content, the dependencies hold across utterances paired by function—across 'adjacency pairs' in the terminology of conversation analysis (Schegloff, 2007).

How deep do such embeddings go? Consider (9).

---

[2] The observations are not new, but take on a new significance in the light of recent discussion. They were made early in conversation analysis (e.g. Sacks, 1995 [1967]; see Schegloff (2007) for review), and I even pointed out their significance in terms of the Chomsky hierarchy 30 years ago (Levinson, 1981), noting however a number of non-syntax-like properties. See also Koschmann (2010). Merritt (1976) also discussed a range of discourse structures in services encounters, including embeddings.

(9) Merritt (1996)

C: Do you have master carbons? ← Q(0) (Pre-Request)
S: (pause) Yes, I think we do ← A(0)
What kind do you want? ←Q(1)
C: How many kinds do you have? ← Q2
S: Well, there are carbons for gelatin duplicators, and
carbons for spirits ← A2
C: Well I'll take the carbons for spirits, please ← A(1)
S: ((goes to get)) ← Action(0)

Here we have a Q–A pair embedded within a Q–A pair embedded within a request–compliance pair and thus an embedding of degree 2—a depth that occurs vanishingly rarely in spoken language syntax, but which in spoken discourse is routine. As the bracketing makes clear, this is a pushdown stack, responses climbing back up the stack. That is because speech acts tend to come in 'adjacency pairs', so that a question expects an answer in next turn; where the adjacency criterion is not met, an answer is nevertheless still due.

There are a range of reasons for these 'insert sequences', but typically the inserted adjacency pairs deal with a prerequisite for handling the initial action (Schegloff, 2007). One prerequisite is hearing or understanding the prior turn. Thus (10) is an example of a different type involving other initiation of repair, with a further repair initiation on the first repair initiator. It takes us to degree 3 embedding, well beyond the attested bound for recursive embedding in spoken language:

(10) Merritt (1976)
S: Next ← Request to order (0)
0 C: Roast beef on rye ← Order (0)
1 S: Mustard or mayonnaise? ← Q(1)
2 C: Excuse me? ←Repair Initiator (2)
3 S: What? ← Repair (3)
3 C: Excuse me, I didn't hear what you said ← Repair (3)
1 S: Do you want mustard or mayonnaise? ← Repair(2)
C: Mustard please ← A(1)
0 S: ((provides)) ← Compliance with order (0)

Finally, consider (11), where a series of queries within queries takes us to degree 4, exceeding any attested depth of embedding in natural language syntax.

(11) Abbreviated from Levinson (1983)

C: .. I ordered some paint… some vermillion…..

And I wanted to order some more, the name's Boyd ←Pre-Order (0)

R: Yes how many tubes would you like sir? ←Q(1)

C: .. What's the price now with VAT? ←Q(2)

R: I'll just work that out for you ←Hold(3)

C: Thanks ←Accept Hold(3)

(10.0)

R: Three pounds nineteen a tube sir ←A(2)

C: Three nineteen is it= ←Q(3)

R: Yeah ←A(3)

C: That's for the large tube? ←Q(4)

R: Well yeah it's the 37 ccs ←A(4)

C: I'll just eh ring you back I have to work out how many I'll need ←Hold (2) for A(1)

((call-back with order and acceptance)) ← (0)

Human subjects performing psycholinguistic tests in an artificial-grammar learning paradigm show large degradation in performance at and after degree 3 embedding—'whereas two nested dependencies are still within our processing limits, three nested dependencies appear to be beyond what we can process' (De Vries et al., 2012). In contrast, the deepest attested nesting of centre-embedded insertion sequences seems to be of at least degree 6 (see Levinson (2013) and citations there).

Discussion

It has been argued here that recursive embedding in syntax is not necessarily a prominent feature of languages—in some large class of languages (yet to be exactly determined, but including many Australian languages), it is either not clearly evidenced or capped at a very shallow level. These languages provide no evidence, therefore, that a core element of language design is indefinite embedding of the kind produced by a context-free grammar. On the other hand, whether or not languages have clear syntactic embedding, they always seem to make use of 'pragmatic embedding' as it were—that is uncoded construals that understand clauses as if they were complements, relative clauses or temporally subordinate. The two facts together suggest that 'recursion' understood propositionally (as relations between propositions) is not so much a universal property of grammar as a property of human psychology, most evident in language use.

Examining the patterning across turns in interactive discourse (dialogue in most of the cases examined), we find a curious analogue of the recursive embedding that has so much exercised linguists. Turns at talk are tied to each other as responses to prior speech acts, typically across adjacency pairs like question–answer, request–compliance and offer–acceptance. When so construed, we see that pairs of utterances may be embedded within other pairs of utterances, apparently with little effort and to a much great depth than is exhibited in syntax. Once again, pragmatics outplays syntax.

This phenomenon raises a central question: how can we explain that what is apparently cognitively impossible in syntax (namely, indefinite centre embedding) is so straightforward in the pragmatics of dialogue? The dialogue facts seem to rule out the idea that there is an absolute performance barrier due to short-term memory limitations; they also seem to undermine the idea that the difficulty found in syntax is based on holding dependencies over a lot of intermediate material (Gibson's, 1998 locality effects). In the dialogue case, exactly the same pushdown stack structure is involved, and the range over which these dependencies have to be held in memory can be immense (see Levinson (2013) for an example spanning 80 turns).

Perhaps the mystery can be partly dissolved in the following way. Note that our action planning system in general needs to be able to hold a stack of subgoals, and check them off one by one—to make the tea may require calling the water-getting subroutine, which may require the kettle-finding subroutine, etc. Many aspects of language use are best explained in terms of joint-action planning (Clark, 1996; Levinson, 2013), so that language usage is able to draw directly on the cognition of our action systems in a way that syntax cannot. Note that the indefinite centre embedding in interactive discourse is construed over speech acts—actions in linguistic clothing. In addition, interactive language use is 'distributed cognition' par excellence, and this may somehow lower the processing load, although to participate effectively in such joint action, each party must nevertheless model the whole emerging structure. If action, and specifically joint action, is indeed the root of this ability to parse embedded structures, then the more abstract and removed from this domain a mental task is, the more restricted human processing of this kind may be expected to be. That might explain our limited prowess in syntax. But this is speculation.

When an ability is much more developed in one arena than another, it seems reasonable to surmise that it is primarily adapted for the more developed arena. The inference then is that syntactic embedding may have evolved out of our capacities in the dialogue arena, which in turn draws directly on joint-action abilities. There is just some general evidence for this in the discourse sources of complex constructions. Geluykens (1992), for example, has shown that complex constructions like left dislocations are often interactionally produced with a slot for a minimal response. Likewise, specialists have noted that in language genesis, in the progression from pidgin to creole, paratactic constructions give rise to subordination (Sankoff & Brown, 1976). Thirdly, in child language development, structures like conditionals seem to arise from a distribution of turns across speakers (De Castro Campos, 1981, following Jespersen's, 1940): there are thus at least three lines of evidence—from corpora, from creolization or language change and from child language—that may suggest an origin of complex syntax in interactive language use.

Acknowledgements I thank the participants at *The International Conference on Language and Recursion*, Mons, Belgium, 14 March 2011; my colleagues at the MPI for Psycholinguistics in Nijmegen (especially Penelope Brown, Kobin Kendrick, Pim Levelt and Karl-Magnus Peterssen); and Nick Evans for comments on this work. A fuller version of many of these ideas can be found in Levinson (2013). This work was conducted within the framework of the ERC Advanced Grant #269484 INTERACT.

References

Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes, 1*, 249–262.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 269–321). New York: Wiley.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*, 157–205.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Comrie, B., & Kuteva, T. (2008). Relativization on subjects. In M. Haspelmath, M. Dryer, D. Gil, & B. Comrie (Eds.), *The world Atlas of language structures online*. Max Planck Digital Library: Munich.

De Castro Campos, M. F. P. (1981). *On conditionals as dialogue constructs*. Paper presented at the International Encounter in the Philosophy of Language, Campinas.

De Vries, M., Petersson, K. M., Geukes, S., Zwitserlood, P., & Christiansen, M. H. (2012). Processing multiple non-adjacent dependencies: Evidence from sequence learning. *The Philosophical Transactions of the Royal Society B, 367*, 2065–2076.

Englebretson, R. (2003). *The problem of complementation in colloquial Indonesian Conversation. Studies in Discourse and Grammar 13*. Amsterdam: John Benjamins.

Evans, N. (1995). *A grammar of Kayardild*. Berlin: Mouton de Gruyter.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*(5), 429–492. doi:10.1017/S0140525X0999094X.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology, 46*, 621–646.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*, 179–210.

Folia, V., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2011). Implicit artificial syntax processing: Genes, preference, and bounded recursion. *Biolinguistics, 5*, 105–132.

Geluykens, R. (1992). *From discourse process to grammatical construction*. Amsterdam: John Benjamins.

Gibson, E. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown* (Ph.D. Thesis). Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*, 1–76.

Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes, 14*(3), 225–248.

Hale, K. L. (1976). The adjoined relative clause in Australia. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 78–105). Canberra, AIAS and New Jersey: Humanities Press.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jespersen, O. (1940). *A modern English grammar on historical principles*. London: George Allen and Unwin.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics, 43*, 365–392.

Kimball, J. (1973). Seven principles of surface-structure parsing in natural language. *Cognition, 2*, 15–47.

Koschmann, T. (2010). On the universality of recursion. *Lingua, 120*(12). doi:10.1016/j.lingua.2010.03.019

Levinson, S. C. (1981). Some pre-observations on the modelling of dialogue. *Discourse Processes, 4*(2), 93–116. doi:10.1080/01638538109544510.

Levinson, S. C. (2000). *Presumptive meanings*. Cambridge, MA: MIT Press.

Levinson, S. C. (2013). Recursion in pragmatics. *Language, 000–000*.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: A theory of text organization. *Text, 8*(3), 243–281.

Marks, L. (1968). Scaling of grammaticalness of self embedded English sentences. *Journal of Verbal Learning and Verbal Behavior, 5*, 965–967.

Merritt, M. (1976). On questions following questions in service encounters. *Language in Society, 5*, 315–357.

Mithun, M. (1984). The evolution of noun incorporation. *Language, 60*(4), 847–894.

Nevins, A., Pesetsky, D., & Rodrigues, C. (2009). Pirahã exceptionality: A reassessment. *Language, 85*(2), 355–404.

Nordlinger, R. (2006). Spearing the Emu drinking: Subordination and the adjoined relative clause in Wambaya. *Australian Journal of Linguistics, 26*(1), 5–29.

Perfors, A., Tenenbaum, J. B., Gibson, E., & Regier, T. (2010). How recursive is language? A Bayesian exploration. In H. van der Hulst (Ed.), *Recursion and human language* (pp. 159–175). Berlin: Mouton de Gruyter.

Sacks, H. (1995). *Lectures on conversation*. Oxford: Blackwell.

Sankoff, G., & Brown, P. (1976). The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language, 52*, 631–666.

Schegloff, E. A. (2007). *Sequence organization in interaction*. Cambridge: Cambridge University Press.

Weckerly, J., & Elman, J. L. (1992). A PDP approach to processing center-embedded sentences. In *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 139–156). Hillsdale, NJ: Erlbaum.

Chapter 2
Investigating Recursion Within a Domain-Specific Framework

Maurício Dias Martins and William Tecumseh Fitch

Introduction

Recursion has long been an object of interest and fascination from scholars in different fields such as mathematics, computer science, linguistics and visual arts, partly because recursion allows the generation of structures that are simple and complex at the same time. Recursive structures are complex because they can contain infinite hierarchical levels, yet simple because this infinity can be achieved and represented using very simple rules.

One famous class of recursive structures is the fractals (Fig. 2.1). Fractals are structures that display self-similarity (Mandelbrot, 1977), so that they appear geometrically similar when viewed at different scales. Fractals are produced by simple rules that, when applied iteratively to their own output, can generate complex hierarchical structures. Since the same kind of representation can be used at different levels of depth, simple rules suffice to represent the whole structure.

Humans have long understood this generative potential at an intuitive level. In ancient Egypt we find depictions of recursive structures representing the self-generating power of the universe (Eglash, 1997), and the Sierpinski triangles shown in Fig. 2.1 are found in the Anagni Cathedral, Italy (Wolfram, 2002). In other contemporary, pre-industrialized cultures, we find recursive depth as a symbolic representation of status or power (Eglash, 1998).

M.D. Martins (✉)
Department of Cognitive Biology, University of Vienna,
Althanstrasse 14, 1090 Vienna, Austria

Language Research Laboratory, Lisbon Faculty of Medicine,
Av. Professor Egas Moniz, 1649-028 Lisbon, Portugal
e-mail: mauricio.martins@univie.ac.at

W.T. Fitch
Department of Cognitive Biology, University of Vienna,
Althanstrasse 14, 1090 Vienna, Austria

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_2,

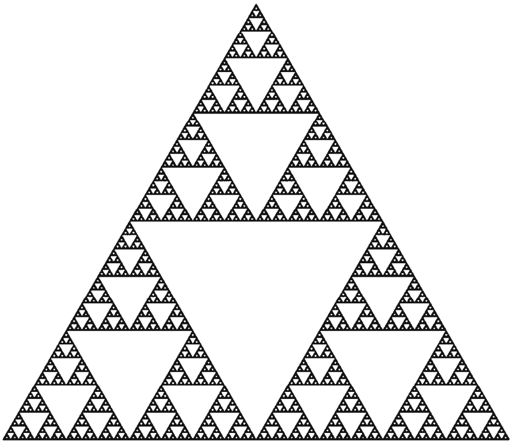


Fig. 2.1 Example of fractal structure—a Sierpinski gasket—exhibiting self-similarity

Due to these curious properties, several contemporary thinkers have proposed that recursion, when available to the human mind, could have been associated with the emergence of our unbounded creativity (Hofstadter, 1980; Penrose, 1989). When available to different domains such as language (Hauser, Chomsky, & Fitch, 2002), problem solving (Schiemenz, 2002) and spatial reasoning, recursion could have allowed an open-ended generative power.

Unfortunately the investigation of recursion as a cognitive ability has proven to be difficult for several reasons: First, multiple definitions of recursion exist in different fields of research. When used interchangeably in the literature, such discrepant interpretations hinder mutually consistent interpretation of empirical findings. Second, there are multiple levels of analysis at which recursion can be measured, at least including algorithm, structure and mental representation. Analysing one of these levels does not necessarily allow inferences about the other two. For example, both recursive and non-recursive algorithms can be used to generate recursive structures. Therefore we cannot conclude that systems able to generate recursive structures necessarily use a recursive algorithm. Furthermore, the structural properties attributed to an object depend on the representational abilities of the observer. It is certainly possible for an individual to generate structures without representing them mentally (e.g. one's own heartbeat time series). Third, if we assume that the mind is modular, we might feel tempted to discuss recursion in a restrictive and domain-specific fashion, for example, in linguistic terms (Chomsky, 2010; Hornstein & Pietroski, 2009; Roeper, 2011). But it remains an open empirical question whether recursion is domain-general or domain-specific, and attempts to address this issue must therefore start with a definition compatible with both possibilities.

Here we attempt to address these questions systematically, highlighting some crucial distinctions and laying out a grid of empirical hypothesis. We will also provide examples of syntactic and visuospatial recursion, illustrating how a single framework can be applied to different domains.

Defining Recursion

In general, the term recursion has been used to characterize to the process of embedding a constituent inside another constituent of the same kind (Fitch, 2010; Pinker & Jackendoff, 2005; van der Hulst, 2010). Recursive processes can generate hierarchical structures that display similar properties across different levels of embedding. This feature, called self-similarity, is a signature of recursive structures. In language, this process establishes a dependency relationship between two constituents of the same category. An example of a recursive linguistic structure is the compound noun "[[student] committee]", where we find a noun phrase embedded inside another noun phrase. In contrast, a sentence with a noun and a verb together, such as "[[trees] grow]", is hierarchical, but not recursive, because a constituent of a given type is not nested within a constituent of that same type.

Although recursion has been hypothesized as a uniquely human trait and as a necessary capacity for the evolution of language (Hauser et al., 2002), the diversity of definitions in use has prevented the consistent interpretation of empirical results (Fitch, 2010). One of the major problems is the difficulty in establishing clear distinctions between recursion and similar processes such as hierarchical embedding and iteration (van der Hulst, 2010).

According to the framework that we adopt (Fitch, 2010; Martins, 2012), "iteration" refers to the process of repeating an operation a given number of times. Such processes may or may not generate hierarchical structures or create dependency relationships between different elements. For example, adding one marble at a time to a bag is an iterative process, but neither hierarchical nor recursive. On the other hand, "hierarchical" structures always involve the embedding of elements within other elements. This embedding can refer to the grouping of a set of constituents within a higher order element, such as the grouping of individuals within a family; or it can refer to the establishment of dominance-subordination relationships such as in social hierarchies. If the hierarchical embedding occurs between constituents of the same category (e.g. such as a noun phrase inside a noun phrase), we classify it as recursive, otherwise as non-recursive. Iteration, hierarchical embedding and recursion are not mutually exclusive processes: Recursion typically involves both hierarchy and iteration. Nevertheless, it is important to conceptually segregate the cognitive abilities necessary to represent the kind of information that each of these processes encodes (Fig. 2.2). For instance, encoding iteration requires the ability to represent the repetition of constituents/elements. Encoding hierarchical embedding requires the ability to represent dependency or grouping relationships between constituents. Finally, encoding recursive embedding requires the ability to represent successive hierarchical dependencies (hierarchical levels related by parenthood) with the same rules. A specific behaviour trait that this ability enables is the possibility to generate new hierarchical levels beyond those previously experienced or specified, maintaining consistency with existing levels at a higher level of abstraction.

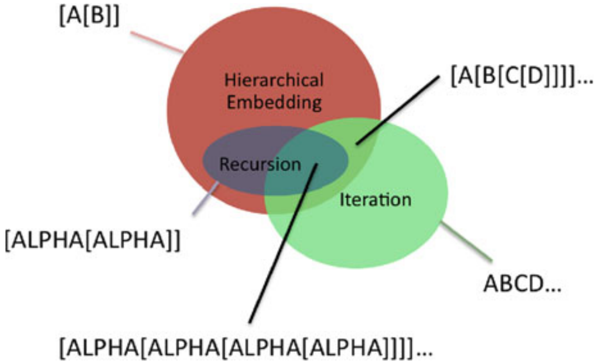


Fig. 2.2 Examples of structures produced by iteration, hierarchical embedding and recursion and by the combination of these processes. A non-iterated hierarchical embedding corresponds to the establishment of a dependency but without repetition. The ability to represent repetition and the ability to represent dependencies may be orthogonal

Iteration, Hierarchy and Recursion: Some Illustrative Examples

Nature provides nice illustrative examples of the distinction between iteration, hierarchy and recursion (Fig. 2.3). Some marine algae grow in a recursive fashion, illustrating self-similarity ("self-embedding"). Multiple, hierarchical growth tips remain undifferentiated and can in principle spawn an endless proliferation of further growth tips. Plants such as grasses can grow by propagating copies along a single extension (a *stolon* or *rhizome*) and illustrate a serial, iterative structure. Iteration can thus exist without hierarchy. Trees and shrubs provide a nice example of hierarchy, because growth occurs in parallel, at multiple growth points, but this becomes non-self-embedding as soon as differentiation into leaves or flowers occurs: Branches bear twigs which bear leaves, but the opposite does not occur. It is interesting to note in this case that more primitive and ancient plants show recursion, while "advanced" plants like angiosperms do not (Niklas, 1997). Another ubiquitous example of non-recursive hierarchy is found in chemistry and atomic physics. Compounds are made of molecules, which are made of atoms made of electrons, neutrons and protons, the latter composed of quarks. As these examples show, many real-world examples of hierarchy occur without recursive self-similarity.

To further exemplify the difference between hierarchical embedding and recursion, consider the following algorithm, which specifies how some letters of the alphabet can be incorporated in a hierarchical structure, in dependency to other letters:

> "Execute one of the following rules: {1) incorporate one or more [B]s in dependency of [A]; 2) incorporate [C]s in dependency of [B]; 3) incorporate [D]s in dependency of [C]}. Repeat as desired".

With such an algorithm, one could generate structures such as the one depicted in Fig. 2.4.

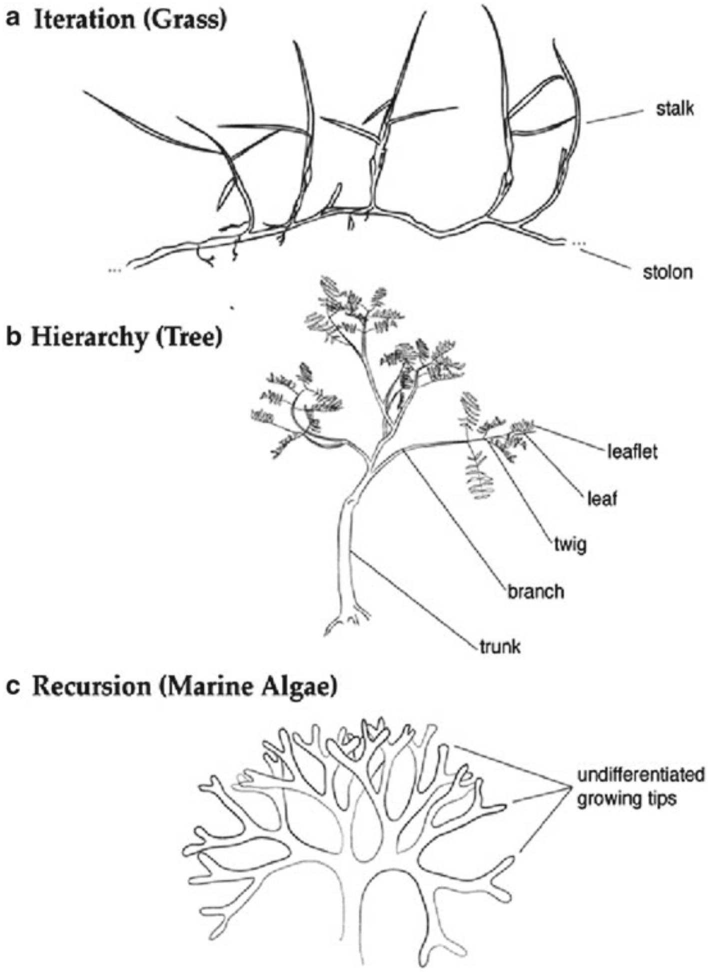


Fig. 2.3 Botanical Examples of Iteration, Hierarchy, and Recursion. (**a**) Grass lateral growth illustrating **Iteration**: Many grasses grow by lateral extension via above ground stolons or below ground rhizomes. Growth along the stolon is iterative: each addition of a new stalk happens singly and independently, with no consequences for other stalks. (**b**) Growth of a tree illustrating non-recursive **Hierarchy**: Growth occurs in parallel, at many different terminals, but is differentiated such that branches bear twigs which bear leaves (which may be compound and made up of leaflets). Leaves cannot bear twigs, and each level is of a different type: this is an example of hierarchy without recursive self-embedding. (**c**) Growth of a marine algae illustrating botanical **Recursion**: As the algae grows, every growth tip can undergo unbounded further subdivision, with no necessary differentiation, into twigs or leaves, and each tip is a potential new plant

This algorithm works fine to handle a predefined number of hierarchical levels. It even allows the incorporation of infinite [C]s within the hierarchy in dependency of a certain [B]. However, if one would like to incorporate an [E] in the structure, this would not be possible without adding a new rule to the algorithm, one that

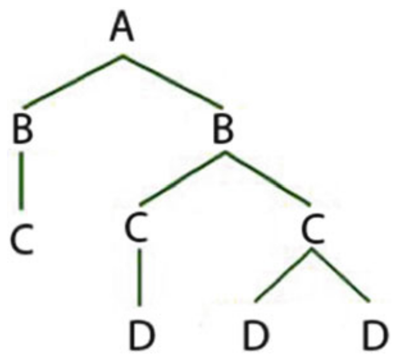


Fig. 2.4 Example of a hierarchical structure generated by a non-recursive algorithm

specifies a priori how [E] can interact with the other letters. Thus, in non-recursive hierarchical embedding, for each hierarchical level that is generated, a specific rule needs to be specified. Recursion overcomes this limitation.

For instance, consider the algorithm defined as "Embed one or more members of the ALPHABET within another member of the ALPHABET". In this example, because all members of the alphabet ([A], [B], [C], [D], [E], etc.) are categorized as having the same properties (regarding the way they interact with the levels above and below), we could incorporate in the hierarchy elements that were not explicitly pre-specified (such as [E] or [F]). Furthermore, we can potentially generate infinite hierarchical levels with one single rule, illustrating the power of recursion to go beyond mere hierarchy.

Clearly, grammars composed only of recursive rules run the risk of being too powerful, allowing the over-generation of structures beyond what would be useful, for example, for transmitting information. The power of recursion, including in language, is only apparent when recursive rules are combined with non-recursive rules (Perfors, Tenenbaum, Gibson, & Regier, 2010). But for this combination to exist, the ability to represent recursion must be present in the first place.

In summary, we propose that the ability to generate novel hierarchical levels beyond those previously specified is a signature trait of recursion, and this trait should be an important object of empirical research aiming to tap into recursive abilities.

Empirical Analysis

Before discussing possible empirical approaches for evaluating recursion, we alert the reader to a widespread misconception: that the formal grammar A^nB^n provides a litmus test for recursion. To our knowledge the first who used the A^nB^n grammar in empirical research was Fitch and Hauser (2004), who proposed it as a test for pattern-perception abilities above the regular (=finite state grammar) level and thus

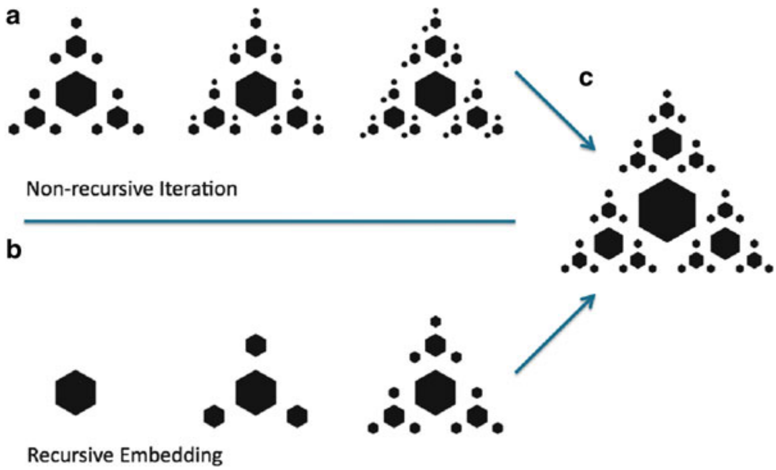


Fig. 2.5 Example of a recursive structure (**c**) generated by non-recursive iteration (**a**) and recursive embedding (**b**)

as a test for George Miller's "supra-regular hypothesis". Miller hypothesized that humans have a propensity to use context-free or other supra-regular grammars when perceiving patterns (Miller, 1967). This is a hypothesis about level in the formal language hierarchy (regular versus context-free), which is quite different from the issue of recursion, which can occur at any level of this hierarchy. Fitch and Hauser (2004) tested Miller's hypothesis for humans and monkeys and did not even mention recursion. Unfortunately, both a commentary on that paper (Premack, 2004) and several subsequent empirical papers have promulgated the misconception that success on the A^nB^n grammar reliably indicates recursive abilities (Abe & Watanabe, 2011; Gentner, Fenn, Margoliash, & Nusbaum, 2006; Marcus, 2006). Despite repeated critiques to the contrary (e.g. Corballis, 2007; Fitch, 2010; Fitch & Friederici, 2012), this seems to be an idea that will not die.

As stated in the introduction, there are 3 levels at which we can investigate the presence of "recursion": (1) the level of the algorithm, (2) the level of structure and (3) the level of representational abilities. Although all levels are interesting, we should not draw inferences about the presence of recursion in one level based on investigations of another. For instance, there are two different processes that can be used to generate Fig. 2.5c. One of these processes is recursive, since it involves self-embedding, and the other is iterative and hierarchical, but not recursive.

At the level of structure, Fig. 2.5c can be plausibly modelled as being recursive, since it contains self-similarity. However, if a certain individual is able to generate Fig. 2.5c, he/she could have used either the recursive procedure (b) or the non-recursive procedure (a). This observation implies that the ability to generate recursive structures such as (c) is neither necessarily informative regarding the ability to use recursive procedures nor the ability to build cognitive representations of recursion. However, there is an empirical way to tackle this issue: If we expose an individual to the first three steps of the recursive procedure depicted in Fig. 2.5b, and he/she is able to generate Fig. 2.5c within that context, then we can assume that he/she was able to (1) represent the underlying rules connecting the previous iterative steps and (2) apply these rules productively to generate one step further. This is compatible with one clear definition of recursion: "an ability to represent multiple levels of hierarchical dependencies (hierarchical levels related by parenthood) following the same rules, entailing the possibility to generalize and produce new levels of embedding beyond those specified a priori (in the algorithm or in the input)".

Modularity and Domain-Specificity Issues

Recently, the development of the human ability to represent recursion has been described as an important step in the evolution of language (Fitch, Hauser, & Chomsky, 2005; Hauser et al., 2002). This ability may have allowed language to take on an unprecedented generative power. For example, recursion allows the representation and generation of sentences where a noun phrase $n+1$ is embedded inside a noun phrase n already embedded in a noun phrase $n-1$ ($[NP_{(n-1)}[NP_{(n)}[NP_{(n+1)}]]]$), as occurs in the sentence "John's sister's house". Besides this gain in generative power, also the speed of acquisition of a certain language may be enhanced by the presence of prior recursive rules in the grammar system (Perfors et al., 2010).

Recursion, as it is used in language, has been hypothesized to be part of a "linguistic computational system […], independent of the other systems with which it interacts and interfaces", potentially restricted to humans (Fitch et al., 2005; Hauser et al., 2002). According to this view, although the usage of recursive rules may be available in nonlinguistic domains such as visual art (Eglash, 1997), music (Jackendoff & Lerdahl, 2006), architecture (Eglash, 1998), humour (Eisenberg, 2008), second-order theory of mind (Miller, 2009), problem solving (Schiemenz, 2002) or action sequencing (Pulvermüller & Fadiga, 2010), these uses may all rely upon a system of abstract arbitrary symbol manipulation dependent on language (Fitch et al., 2005). Alternatively, some authors have proposed that the usage of recursion in some domains, for example, in visual perception, can occur independently of language (Pinker & Jackendoff, 2005).

The idea of recursion as part of a linguistic computational system that is independent of other systems presupposes some concept of modularity. A module, according to Fodor (1983), implies encapsulation and domain-specificity. If recursion is a module, then there are two possibilities: (1) Either recursion is a monolithic operation that does not recruit domain-general computations or (2) recursion results from the interaction of several sub-operations, all domain-specific, and none exchanges information with the external neural milieu during recursion-internal operations.

Although it can be useful to think about the mind as a composite of several encapsulated operations, we think there are some dangers that result from a strict or "massively modular" view. If we entertain the possibility that the ability to represent recursion may result from the interaction of several abilities, some domain-specific and some domain-general, then investigations of the fine-grained structure of recursive operations require broad definitions beyond any one single domain. Currently, there are several plausible hypotheses regarding this fine-grained structure: (H1) Different recursive-type operations in different domains are completely independent; (H2) a single recursive module is recruited by different modalities; and (H3) there is some overlap between recursive operations in different domains, together with some dissociations owing to domain-specific computations and/or interface constraints.

Given that the current empirical evidence does not allow the elimination of any of these hypotheses, we advocate a definition of recursion that is compatible with several domains. The definition of recursion that we proposed above satisfies this desideratum. The concern about assuming domain-specificity is reinforced by recent empirical research in cognitive sciences, where domain-specific activities have been shown to depend more on domain-general operations than previously supposed. This is true, for example, in visual (Kirkham, Slemmer, & Johnson, 2002), verbal (Saffran, Aslin, & Newport, 1996) and motor pattern extraction (Baldwin, Andersson, Saffran, & Meyer, 2008) as well as in social reasoning (McKinnon & Moscovitch, 2007), music perception (Treuhub & Hannon, 2006) and number/quantity representation (Holloway & Ansari, 2008).

It thus seems likely to us that some components of recursive mental operations may be domain-general and other components highly modularized and domain-specific. Domain-specificity may be especially true regarding operations associated with so-called fast thought or expert behaviour.

According to Kahneman (2011) and many others, human cognition uses two separated systems: one fast, intuitive and heuristic and another slow, effortful and abstract. It is possible that the generation of abstract recursive representation rules is domain-general. This would imply a transferability of representational knowledge across domains. However, it is also possible that the representation of recursive structures is achieved not by the generation of abstract rules but by matching the contents of perception with previously acquired (or biologically endowed) templates. Some empirical reports seem to suggest that the processing of familiar structures depends less on domain-general cognitive abilities than the processing of unfamiliar structures. This seems to be true for syntactic processing (Novik, Trueswell, & Thompson-Schill, 2005), visual perception (Sinha & Balas, 2008) and social reasoning (McKinnon & Moscovitch, 2007) and may reflect a progressive automatization and "lexicalization" of structural knowledge, which complements more abstract and flexible representations (Brinton, 2008).

How exactly recursion is represented by different species in different domains and ontogenetic stages is an open question that can be investigated empirically, for example, with dual task paradigms.

Conclusion

In this chapter we outlined some difficulties concerning the assessment of recursion as a cognitive ability and offered a framework that allows specific questions to be assessed separately:

1. Is the ability to represent recursion cognitively distinct from the ability to represent similar operations such as hierarchical embedding and iteration?
2. Are these abilities predicted by general intelligence?
3. Is the ability to represent recursion present in more than one domain?
4. If present in more than one domain, is there a single recursion module that is recruited by different modalities; or is the execution or recursion-type operations achieved by (partially or totally) different cognitive resources in different domains?
5. Are there any cognitive abilities or operations that constitute strict causal precursors of recursion, in the absence of which recursion cannot be represented (e.g. language)?
6. If recursive capacities are present in other species or in multiple domains, are they achieved via a flexible, slow and abstract representational system, or via an automatic and rigid template-matching system?

Our recent work (Martins & Fitch, 2012) begins to answer some of these questions. For instance, our results demonstrate that recursion can be represented in the visuospatial domain and that there is some cognitive dissociation between recursion, general intelligence and hierarchical embedding. Interestingly, both behavioural and imaging data seem to suggest that recursion recruits visuospatial-specific resources less than hierarchical embedding and relies more on domain general resources. Finally, the ability to represent recursion in the visual domain seems to be associated with high reaction times initially. However, with practice, these reaction times decrease. Furthermore, performance in latter trials seems to recruit different cognitive resources in comparison with initial trials, perhaps reflecting the transition from an abstract representational strategy to a more automatic one.

However represented, and whatever its precise role in different cognitive domains, recursion is an important and powerful property of human cognition. The fundamental questions that it raises concerning human cognition and our evolutionary history require a clear framework that both allows and encourages the development of empirical approaches spanning different cognitive domains.

Acknowledgements The authors are grateful to Nadja Kavcik, for her assistance in generating the pictures of this chapter. This work was supported by the FCT grant SFRH/BD/64206/2009 to MM and by ERC Advanced Grant SOMACCA, project number 230604 to WTF.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*(8), 1067–1074.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition, 106*(3), 1382–1407.

Brinton, L. J. (2008). Where grammar and lexis meet. In E. Seoane & M. J. López-Couso (Eds.), *Theoretical and empirical issues in grammaticalization*. Amsterdam: John Benjamins Publishing Co.

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? In R. K. Larson, V. M. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 45–62). New York: Cambridge University Press.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science, 31*(4), 697–704. doi:10.1080/15326900701399947.

Eglash, R. (1997). Bamana sand divination: Recursion in ethnomathematics. *American Anthropologist, 99*(1), 112–122.

Eglash, R. (1998). Fractals in African settlement architecture. *Complexity, 4*(2), 21–29.

Eisenberg, M. (2008). Recursion—or, better computational thinking through laughter. *International Journal of Computers for Mathematical Learning, 13*, 171–174.

Fitch, W. T. (2010). Three meanings of "recursion": Key distinctions for biolinguistics. In R. K. Larson, V. M. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 73–90). New York: Cambridge University Press.

Fitch, W. T., & Friederici, A. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society B, 367*, 1933–1955.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a non-human primate. *Science, 303*, 377–280.

Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition, 97*(2), 179–210. doi:10.1016/j.cognition.2005.02.005.

Fodor, J. A. (1983). *The modularity of the mind*. Cambridge, MA: Bradford.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204–1207. doi:10.1038/nature04675.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hofstadter, D. R. (1980). *Gödel, escher, bach: An eternal golden braid*. Harmondsworth: Penguin.

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science, 11*(5), 644–649.

Hornstein, N., & Pietroski, P. (2009). Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics, 8*, 113–139.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition, 100*(1), 33–72.

Kahneman, D. (2011). *Thinking fast and slow*. New York: Farrar, Straus and Giroux.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition*, B35–B42.

Mandelbrot, B. (1977). *The fractal geometry of nature*. New York: Freeman.

Marcus, G. (2006). Startling starlings. *Nature, 440*, 1204–1207.

Martins, M. D. (2012). Specific signatures of recursion. *Philosophical Transactions of the Royal Society B, 367*, 2055–2064.

Martins, M. J., & Fitch, W. T. (2012). Empirical approaches to recursion. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *The evolution of language – Proceedings of the 9th International conference* (pp. 219–225). Singapore: World Scientific.

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition, 102*, 179–218.
Miller, G. A. (1967). Project grammarama. In G. A. Miller (Ed.), *Psychology of communication*. New York: Basic Books.
Miller, S. A. (2009). Children's understanding of the second-order mental states. *Psychological Bulletin, 135*(5), 749–773.
Niklas, K. J. (1997). *The evolutionary biology of plants*. Chicago: University of Chicago Press.
Novik, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*(3), 263–281.
Penrose, R. (1989). *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford: Oxford University Press.
Perfors, A., Tenenbaum, J., Gibson, E., & Regier, T. (2010). How recursive is language? A Bayesian exploration. In H. van der Hulst (Ed.), *Recursion and human language*. Berlin: De Gruyter Mouton.
Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition, 95*(2), 201–236. doi:10.1016/j.cognition.2004.08.004.
Premack, D. (2004). Is language the key to human intelligence? *Science, 303*, 318–320.
Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*(5), 351–360. doi:10.1038/nrn2811.
Roeper, T. (2011). The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics, 5*(1–2), 57–86.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.
Schiemenz, B. (2002). *Managing complexity with recursion.* Paper presented at the Cybernetics and Systems, Wien.
Sinha, P., & Balas, B. J. (2008). Computational modeling of visual information processing. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 612–634). Cambridge: Cambridge University Press.
Treuhub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*, 73–99.
van der Hulst, H. (2010). Re recursion. In H. van der Hulst (Ed.), *Recursion and human language*. Berlin: De Gruyter Mouton.
Wolfram, S. (2002). *A new kind of science*. Champagne, IL: Wolfram Media.

Chapter 3
Recursive Cognition as a Prelude to Language

Michael C. Corballis

It is commonly held that recursion, the process whereby a computational routine calls itself or calls a similar routine, is the essence of human language and distinguishes language from other forms of animal communication. This has been a continuing theme of Chomsky's work since his 1957 book, *Syntactic Structures*, and reiterated in Hauser, Chomsky, and Fitch (2002). More particularly, these authors distinguish two levels of language, *the faculty of language in the broad sense* (FLB), which they regard as shared between humans and other species, and *the faculty of language in the narrow sense* (FLN), which they regard as unique to humans. They write: "… we suggest that FLN—the computational mechanism of recursion—is recently evolved and unique to our species" (Hauser et al., 2002, p. 1573). In this chapter, I argue that recursion is not restricted to language and indeed may not be universally present in present-day languages, but is rather a property of human thought that probably preceded language.

First, though, I consider the nature of recursion as manifest in language, and Chomsky's view on the nature of language and thought. It is recursion that allows phrases to call other phrases, without formal limit. An often quoted example is the child's story *The House that Jack Built*, in which sentences of increasing length are constructed by adding new phrases. One of them goes like this:

This is the cat that killed the rat that ate the malt that lay in the house that Jack built.

This is an example of recursion where successive phrases are added before or after the sentence.[1] More complex is center recursion, where phrases are inserted

[1] In the child's story, the phrases are progressively added to the beginning. In the example given, the sentence could also be constructed by adding phrases to the end, starting with "This is the cat."

M.C. Corballis (✉)
University of Auckland, Auckland, New Zealand
e-mail: m.corballis@auckland.ac.nz

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_3,

within phrases rather than before or after them, as in the following attempt to shift the focus from the cat to the malt:

The malt that the rat that the cat killed ate lay in the house that Jack built.

(This may take a moment or two to unpack.) The recursive linking of phrases is often used to express the complexities of human thought, as in "Women think that men think that they think that men think that women's orgasm is different" (Premack, 2007).

In Chomsky's view, recursive sentences and recursive thought depend on the same underlying structure. FLN is effectively the same as what he has previously called I-language, the internal language that underlies all human languages and the distinctive character of human thought. I-language is common to all people and is mapped onto the external languages actually spoken or signed in the world. Thus I-language is a symbolic mode of representation and computation, of which language is one manifestation. In Chomsky's (1995) so-called minimalist program, I-language operates according to a recursive principle of "unbounded merge," whereby elements are merged to form larger units, which are then merged to form still larger ones, and so on. Chomsky argues that I-language could not have evolved through natural selection because it consists of abstract symbols that in themselves have no reference to mind-independent objects and therefore cannot have been "selected." He also asserts that there was no language prior to around 100,000 years ago, so that I-language must have emerged in a single step in the recent history of our species. He writes:

*Within some small group from which we are all descended, a rewiring of the brain took place in some individual, call him [*sic*] Prometheus, yielding the operation of unbounded Merge, applying to concepts with intricate (and little understood) properties* (Chomsky, 2010, p. 59).

In similar fashion, the linguist Derek Bickerton once wrote that "… true language, via the emergence of syntax, was a catastrophic event, occurring within the first few generations of *Homo sapiens sapiens*" (Bickerton, 1995, p. 69).

The idea that language and thought are inextricably linked is widely held. It was articulated in Fodor's (1975) book *The Language of Thought*, and authors such as Pylyshyn (1973) argued that even mental imagery, seemingly pictorial rather than symbolic, was based on symbolic, propositional structures. Essentially the same view has been expressed more recently by Penn, Holyoak, and Povinelli (2008), who refer to the idea that human language and thought could have emerged incrementally through natural selection as "Darwin's mistake." Chomsky quotes the archeologist Ian Tattersall (2005), as suggesting that human intelligence is an "emergent quality, the result of a chance combination of factors, rather than a product of Nature's patient and gradual engineering over the eons" (Chomsky, 2010, p. 59). Other archeologists point to the emergence of "modern human behavior," including the discovery of such symbolically motivated artifacts as cave drawings, bodily ornamentation, and burial rituals, as evidence that a dramatic change occurred within the past 100,000 years. Klein (2009), for example, writes that it becomes "at least plausible to tie the basic behavioral shift at 50 ka to a fortuitous mutation that created the fully modern brain" (p. 271). An even more recent date is suggested by Hoffecker (2007):

> *Language is a plausible source for sudden and dramatic change in the archaeological record [after 40 ka] because: (a) it is difficult to conceive of how the system for generating sentences (*i.e.*, syntax) could have evolved gradually, and (b) it must have had far-reaching effects on all aspects of behavior by creating the collective brain (p. 379)."*

Toward a Darwinian Account

The idea that human language and thought emerged in a single step—the "great leap forward," to quote Chomsky (2010)—is clearly counter to Darwinian theory. In 1859, Darwin wrote:

> *If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down. But I can find no such case (p. 158).*

Chomsky has consistently referred to language as an organ—in a recent article, for example, he suggests that the study of language might be "taken to be the study of a real object, a biological organ, comparable to the visual or immune systems …" (Chomsky, 2007, p. 2). If language and thought indeed emerged in a single step, we might therefore consider Darwin's theory of natural selection to be disproven.

The apparent uniqueness of human language and thought need not imply, though, that these faculties emerged only in our species, since a number of different hominin species have arisen since the common ancestor of humans, chimpanzees, and bonobos some 6 or 7 million years ago. The distinctive characteristics of human language and thought could well have arisen incrementally in this time frame. Moreover, research on apes and other primates, and even mammals, suggests a considerable degree of continuity in the evolution of thought, if not of language itself.

At least some aspects of human thought can be linked to what has been termed the *default network* in the brain (Buckner, Andrews-Hanna, & Schacter, 2008). This network consists of interacting neural systems in the frontal, temporal, and parietal lobes and is involved in internally generated activities such as autobiographical memory, imagining the future, and taking the mental perspective of others—activities that we might call *mind wandering* (Corballis, 2012). The same basic neural architecture is present in the monkey brain (Vincent et al., 2007) and even, according to a recent report, in the rat brain (Lu et al., 2012). This suggests a neural template going far back in evolution for the evolution of internal thought processes. Of course, those thought processes may have gained complexity over evolutionary time, but from a Darwinian perspective, this is likely to have occurred in incremental rather than in a single step.

Nevertheless there may be some properties of the default network in humans not shared by other species. Here, I consider two of them, mental time travel and theory of mind.

Mental Time Travel

Tulving (1972, 2005) has long held that episodic memory, which refers to the storage and conscious recollection of individual episodes, is uniquely human. In Tulving's view episodic memory is distinguished from semantic memory, which refers to the knowledge of facts about the world. The distinction is also sometimes expressed as that between remembering and knowing, respectively. The concept of mental time travel is an extrapolation of memory for past experiences into imagined experiences in the future. Just as we can bring a past episode to mind, so we can imagine a possible future one, and indeed it has been proposed that the function of episodic memory is not so much to provide a record of past experience as to form the basis for imagining and perhaps planning possible future ones (Suddendorf & Corballis, 1997, 2007). Subjectively, too, past and future lie on a continuum, as events move from the future through the present to the past. Not surprisingly, perhaps, the reimagining of past episodes and the imagining of future ones activate largely overlapping regions of brain's default network (Addis, Wong, & Schacter, 2007; Addis, Pan, Vu, Laiser, & Schacter, 2009).

Suddendorf and Corballis (1997, 2007) have suggested that mental time travel in general, and not just episodic memory, is a uniquely human capacity. This has proven controversial, with claims that primates (e.g., Martin-Ordas, Berntsen, & Call, 2013; Menzel, 1999; Osvath, 2009) and even birds (e.g., Clayton, Bussey, & Dickinson, 2003; Raby, Alexis, Dickinson, & Clayton, 2007) may be capable of episodic memory and the imagining of future events. Moreover, as we have seen, the default network itself appears to be present in other primate and mammalian species. Although evidence for mental time travel in nonhuman species can, in some instances at least, be explained in terms of simpler mechanisms, such as association (Suddendorf & Corballis, 2008, 2010), recent neurophysiological evidence from hippocampal activity suggests that the seed for mental time may have been sown far back in evolution (Corballis, 2013).

In humans, the hippocampus is activated both during recall of past events and in the imagining of future ones (e.g., Addis et al., 2007), and neurophysiological evidence now suggests that the hippocampus is similarly active in the rat during replay of past events and pre-play of possible future ones. The so-called place cells in the hippocampus respond when the rat is in particular locations in an environment, such as a maze, suggesting that the hippocampus is involved in the construction and activation of cognitive maps of the environment (O'Keefe & Nadel, 1978). This activity also occurs in sharp-wave ripples (SWRs), sometime after the animal has actually been in the maze, either during slow-wave sleep when the animal is awake but immobile. The paths indicated by the SWRs sometimes map out past excursions in the maze but sometimes correspond to paths the animals did not actually take (e.g., Gupta, der Meer, Touretzky, & Redish, 2010) or paths that they are about to take (Pfeiffer & Foster, 2013). These findings suggest at least limited mental time travel into both past and future episodes.

It still seems likely, though, that the human capacity for mental time travel exceeds that yet demonstrated in any other species, in terms of its temporal dimensions, flexibility, and sheer number of episodes recalled, anticipated, or simply invented. Premack (2007) notes, for example, that demonstrations of memory and planning in scrub jays, as reported by Clayton and her colleagues, are centered on the caching of food, whereas human mental time travel is almost unlimited in scope and is typically complex and social:

> *Complex planning differs from simple planning in these respects. It is social: two or more individuals form the plan, and the beneficiary of the plan is likely to be yet another individual, different from those who form the plan; the plan is not one-shot, but a series of plans; the plan extends not for hours but over years. Neither social nor sequential planning, nor planning that extends over long durations, is likely to be found in animals (p. 13863).*

Evidence for hippocampal recording in the rat, described above, is so far limited to spatial events, over restricted time spans. A recent study, though, suggests that chimpanzees and orangutans remember specific events that occurred up to 4 years previously, suggesting that the capacity for mental time travel in great apes may indeed approach that in humans (Martin-Ordas et al., 2013).

Mental time travel is recursive in that past and future experience can be embedded in present experience and even in each other. Thus I might remember that yesterday I had imagined doing something next week. Indeed the embedded nature of mental time travel may well underlie the recursive nature of language—a point I return to below.

Theory of Mind

Theory of mind refers to the ability to understand what is in the minds of others and indeed take the mental perspective of others, and it too involves the default network, especially in the region of the temporoparietal junction (Saxe & Kanwisher, 2003). Thus our minds can wander not only through time but into the lives of others, providing us with the ingredients of biography and fiction. Theory of mind is most often tested in terms of the understanding that another person might have a belief different from one's own. This can be assessed in children with the so-called Sally-Anne test. The child is shown two dolls, one called Sally and the other called Anne. Sally has a basket and Anne has a box. Sally puts a marble in the basket and leaves the scene; Anne then takes the marble out of the basket and puts it in her box. Sally comes back, and the child is asked where she will look for her marble. Children under the age of four typically say that Sally will look in the box, which is where the marble actually is. Older children will understand that Sally did not see the marble being shifted and will correctly say that Sally will look in the basket. They understand that Sally has a false belief (Wimmer & Perner, 1983).

The question of whether it is a uniquely human attribute has also proven controversial. In one recent review of the evidence, Penn and Povinelli (2007) conclude that there is no evidence that nonhuman animals, chimpanzees included, have anything resembling a theory-of-mind system. In contrast to the abundant evidence for theory of mind in young children, they claim the evidence from chimpanzees can be explained in terms of the animals' behavioral observations of other animals, rather than an understanding of what is happening in their minds. In another review, though, Call and Tomasello (2008) conclude that 30 years of research have shown chimpanzees, at least, to have some understanding of the goals, intentions, perceptions, and knowledge of others, but no understanding of the beliefs and desires of others. From a Darwinian perspective, this conclusion suggests a degree of continuity, but with greater complexity in humans.

The difference may lie partially in levels of recursion (Corballis, 2011). To the extent that chimpanzees can take the perspective of others, they may be said to possess first-order recursion. Tomasello (2008), though, has suggested that humans are capable of at least second-order recursion—that is, person A may know not only what person B is thinking but also that B knows what A is thinking. Indeed human social understanding may proceed well beyond the second level; Premack (2007, p. 13865) gives the example "John thinks that Bill thinks that Henry believes that John should put his kids in Sunday school." Just as we humans show greater scope and complexity in mental time travel, so we have greater complexity in the recursive understanding of the mental states of others.

Preludes to Language

Mental time travel and theory of mind may be important preludes of language, and it is perhaps the extended range of mind wandering in humans that led to the emergence of language. As Suddendorf and Corballis (1997, 2007) point out, episodic memory offers adaptive advantages over instinct and semantic or associative memory in that it is highly flexible, allowing for the development of detailed scenarios for the future. Language, moreover, is exquisitely designed for the communication of episodes, whether past, planned, or simply fictional, between individuals, so that the advantages of mental time travel can be shared. The property of language that allows time travels to be shared is displacement, the property that allows us to communicate about the not present. We carry an elaborate system of concepts relating to objects, actions, qualities, and so on and words to label them, along with combinational principles. Syntax may owe its origins to the combining of concepts to convey episodes.

Theory of mind may also be a necessary precursor of language. Its importance can be credited to Grice (1975), who held that language depends on inference rather than explicit decoding. Animal communication, like computer languages, is generally unambiguous, whereas human language, despite its apparent richness, is characteristically ambiguous and imprecise. In order to converse, individuals must understand what is going on in each other's minds, so that each can infer what the other means. This requires at least second-order theory of mind; that is, A must not only know what is in B's mind but must also know that B knows this. This theme has been pursued by Sperber and colleagues (e.g., Sperber & Origgi, 2010), who give the example of an utterance like "It was too slow." This could refer recent economic developments, a movement of a symphony, the speed of a car, a lecture, and many other things. In context, though, speaker and listener usually have no need to spell out the meaning in more detail; each knows what the other is thinking. Viewed in this light, language is itself a mechanism for mind reading.

Mental time travel and theory of mind may well have properties that are distinctively human, but it is likely that they evolved from the default network, a system that probably goes far back in mammalian evolution. The likely existence of the default network in monkeys and even rats suggests that nonhuman animals are capable of internal thought, and as we have seen, rats do seem capable of limited mental time travel. Monkeys appear capable of relatively complex mental operations such as mental rotation and memory scanning (Georgopoulos & Pellizzer, 1995). The capacity to form concepts is not unique to humans, nor is the capacity to attach arbitrary labels to them. Great apes can attach non-iconic labels to hundreds of objects and actions and even combine them to create and understand simple commands, although these fall far short of recursive human language (Gardner & Gardner, 1969; Patterson, 1978; Savage-Rumbaugh, Shanker, & Taylor, 1998). Dogs can also respond to verbal names of objects, although they cannot produce them. Border collies seem especially prolific. In one study, a collie named Rico was reported to know the names of 200 objects (Kaminski, Call, & Fischer, 2004), but this has been surpassed by another collie, Chaser, with a vocabulary of 1,022 objects, including three category names (Pilley & Reid, 2012). These names, moreover, can be used in the absence of the objects they refer to, as when the animal is instructed to go to another room to fetch a named object. This indicates displacement, but perhaps not at the level needed to sustain or communicate mental time travels.

The default network probably gained in complexity in human evolution, perhaps gaining extra recursive structure during the Pleistocene, dating from some 2.6 million years ago. It was during this era that the brain tripled in size, perhaps driven by enhanced social structures necessary for survival in an increasingly challenging dangerous environment, in which forests gave way to more open savanna and the threat of attack from dangerous predators. It may well be the case that something like unbounded Merge, as a device for concatenating concepts into episodic structures, emerged gradually during the Pleistocene, as the pressure to remember complex information, plan futures, and collaborate with others grew more intense. Evolutionary psychologists see the Pleistocene, rather than some miraculous saltation within the past 50,000 years, as the cradle of human thought (e.g., Tooby & DeVore, 1987).

Language, then, may have arisen as a consequence of the emergence of complex thought structures, including mental time travel and theory of mind, that gained recursive complexity during the Pleistocene. Despite the proclamations of Chomsky and others, language itself may not require recursion. Evans and Levinson (2009) give several examples of languages that have no constituent structure, no recursion, or recursion capped at a single level of embedding. One language that appears to lack any recursive structure is that of the Pirahã, but the Pirahã people are nevertheless clearly capable of recursive thought (Everett, 2005, 2009). Recursive thought itself probably emerged from internal thought structures evident also in the primate brain and perhaps even more widely in the mammalian brain.

References

Addis, D. R., Pan, L., Vu, M.-A., Laiser, N., & Schacter, D. L. (2009). Constructing episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imaging and remembering. *Neuropsychologia, 47*, 2222–2238.

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction. *Neuropsychologia, 45*, 1363–1377.

Bickerton, D. (1995). *Language and human behavior*. Seattle, WA: University of Washington Press.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences, 1124*, 1–38.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*, 187–192.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland & M. Gaertner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics* (pp. 1–30). New York: Mouton de Gruyter.

Chomsky, N. (2010). Some simple evo devo theses: How true might they be for language? In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 45–62). Cambridge: Cambridge University Press.

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Trends in Cognitive Sciences, 4*, 685–691.

Corballis, M. C. (2011). *The recursive mind: The origins of human language, thought, and civilization*. Princeton, NJ: Princeton University Press.

Corballis, M. C. (2012). Mind wandering. *American Scientist, 100*, 210–217.

Corballis, M. C. (2013). Mental time travel: A case for evolutionary continuity. *Trends in Cognitive Sciences, 17*, 5–6.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429–492.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology, 46*, 621–646.

Everett, D. L. (2009). Pirahã culture and grammar: A response to some criticisms. *Language, 85*, 405–442.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science, 165*, 664–672.

Georgopoulos, A. P., & Pellizzer, G. (1995). The mental and the neural – psychological and neural studies of mental rotation and memory scanning. *Neuropsychologia, 33*, 1531–1547.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics, 3, Speech acts* (pp. 43–58). New York: Academic.

Gupta, A. S., der Meer, M. A. A., Touretzky, D. S., & Redish, A. D. (2010). Hippocampal replay is not a simple function of experience. *Neuron, 65*, 695–705.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hoffecker, J. F. (2007). Representation and recursion in the archaeological record. *Journal of Archaeological Method & Theory, 14*, 359–387.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in the domestic dog: Evidence for "fast mapping". *Science, 304*, 1682–1683.

Klein, R. G. (2009). Out of Africa and the evolution of human behavior. *Evolutionary Anthropology, 17*, 267–281.

Lu, H., Zou, Q., Gu, H., Raichle, M. E., Stein, E. A., & Yihong Yang, Y. (2012). Rat brains also have a default mode network. *Proceedings of the National Academy of Sciences, 109*, 3939–3984.

Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology, 23*, 1–4.

Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee (Pan troglodytes) after extended delays. *Journal of Comparative Psychology, 13*, 426–434.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford: Clarendon.

Osvath, M. (2009). Spontaneous planning for future stone throwing by a male chimpanzee. *Current Biology, 19*, R190–R191.

Patterson, F. (1978). Conversations with a gorilla. *National Geographic, 154*, 438–465.

Penn, D. C., Holyoak, K. J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 108–178.

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind". *Philosophical Transactions of the Royal Society B, 362*, 731–744.

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature, 497*, 74–79.

Pilley, J. W., & Reid, A. K. (2012). Border collie comprehends object names as verbal referents. *Behavioural Processes, 86*, 184–195.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80*, 1–24.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences USA, 104*, 13861–13867.

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub jays. *Nature, 445*, 919–921.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, language, and the human mind*. Oxford: Oxford University Press.

Sperber, D., & Origgi, G. (2010). A pragmatic perspective on the evolution of language. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language* (pp. 124–131). Cambridge: Cambridge University Press.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind". *NeuroImage, 19*, 1835–1842.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs, 123*, 133–167.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*, 299–351.

Suddendorf, T., & Corballis, M. C. (2008). New evidence for animal foresight? *Animal Behaviour, 75*, e1–e3.

Suddendorf, T., & Corballis, M. C. (2010). Behavioural evidence for mental time travel in nonhuman animals. *Behavioural Brain Research, 215*, 292–298.

Tomasello, M. (2008). *The origins of human communication*. Cambridge, MA: MIT Press.

Tooby, J., & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. G. Kinzey (Ed.), *The evolution of human behavior: Primate models* (pp. 183–237). Stony Brook: SUNY Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic.

Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition*. Oxford: Oxford University Press.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., et al. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature, 447*, 83–86.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103–128.

Part II
Non-verbal Communication Devices, Implicit Learning, Language and Recursion

Chapter 4
Nonverbal Communication Devices, Language, Cerebral Flexibility, and Recursive Exercises

Francis Lowenthal and Laurent Lefebvre

Introduction: Special Representations and Mathematics Education

In 1970, one of us tried to teach mathematics to behaviorally perturbed children with mental retardation. These children, 8- to 13-year-olds, had important difficulties concerning spatial structuration and verbal comprehension. They were unable to focus their attention more than a few minutes on a given task. In mathematics, they were unable to count beyond 50 and to compute beyond 20. Traditional methods of teaching were useless. It was thus decided to adapt the nonverbal approach described by several authors (Frédérique, 1970; Frédérique & Papy, 1968; Papy, 1964) and to use the spiral approach described by Bruner (1960). The use of these representations enabled the subjects to acquire the required knowledge but also to acquire a better knowledge of time, space, and verbal communication (Cordier, Héraux, & Lowenthal, 1975). We believe that these positive results are due to the fact that the approach we used, besides being quasi nonverbal, made a form of "learning without failure" possible. It also confronted the subjects to a nonverbal form of hypothetico-deductive reasoning. In order to better understand what actually happened, we decided to create new pedagogical tools and to use them to introduce problem situations in a quasi nonverbal way. This was the basis of the concept of "Nonverbal Communication Device".

F. Lowenthal (✉) • L. Lefebvre
Cognitive Sciences, University of Mons, Mons, Belgium
e-mail: francis.lowenthal@umons.ac.be

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_4, © Springer Science+Business Media New York 2014

The Nonverbal Communication Device Approach

Definition

In order to study and favor cognitive development, we defined a new approach (Lowenthal, 1983): the concept of Nonverbal Communication Device (NVCD). Each NVCD is the combination of a special intervention and observation method and of a specific device.

The method is based on a quasi nonverbal approach since the experimenter does not speak about the problem to be solved and on manipulations by the subject of specific objects. The exercises are used according to the principles of Bruner's spiral approach (Bruner, 1960). This method enables the observer to create a mediator for representing and communicating.

The specific device is made of little pieces which can easily be assembled in different ways. These objects do not belong to the cognitive background of the subjects. They are provided with technical constraints. These constraints make certain actions possible and other impossible. This in turn suggests a logical structure to the subject. Each problem situation created with the specific device can be solved at each of the three Brunerian levels of representation (Bruner, Olver, & Greenfield, 1966).

We present here two examples and the associated exercises.

First Example: The Pegboard and a Programming Language

The Pegboard is a toy for kids: it is made of a plastic baseboard provided with holes defining a square grid and of little plastic pegs which can be placed by the child in the holes in order to make "a nice drawing" (Fig. 4.1). These pegs are defined by two characteristics: the shape of the head (a square or a quarter of a circle) and the color (blue, yellow, red, green, or orange). For the clarity of this paper, the colors are represented by letters in the figures. The "quarter of a circle" headed pegs are represented by triangles, and the term "triangle" is used to mention these pegs.

We used this device to introduce a programming language based on concrete manipulations. This language, inspired by Logo, is accessible to very young children, handicapped or not (Lowenthal, 1985; Lowenthal & Saerens, 1986). For this type of exercises, we must implicitly divide the baseboard in three columns. The left one is reserved for sequences of square headed pegs assigned to a "triangle" which represents the name of the sequence, thus defining a procedure. The central column is reserved for "triangles" (those used to name the procedures): this is the program. The right column corresponds to the execution of this program (Fig. 4.2a).

In order to introduce an exercise, we show the child a baseboard on which he/she can see two of the columns, and we ask him/her (without any other detail) to "make the missing one with squares (or triangles according to the task)." Some exercises are easy like the plain execution of a program (Fig. 4.2a, b). Other more complex

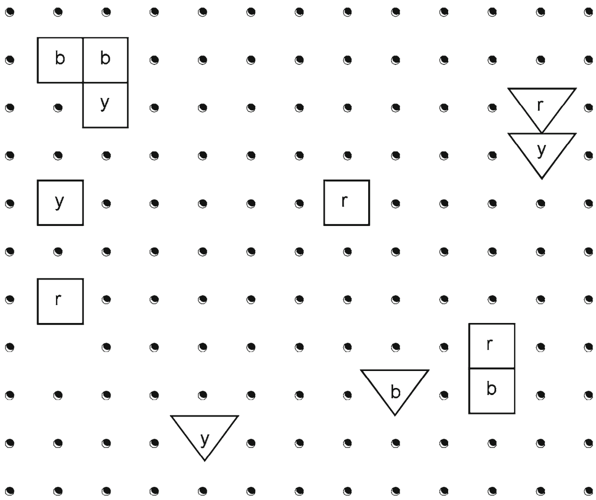


Fig. 4.1 The baseboard and the pegs of the Pegboard

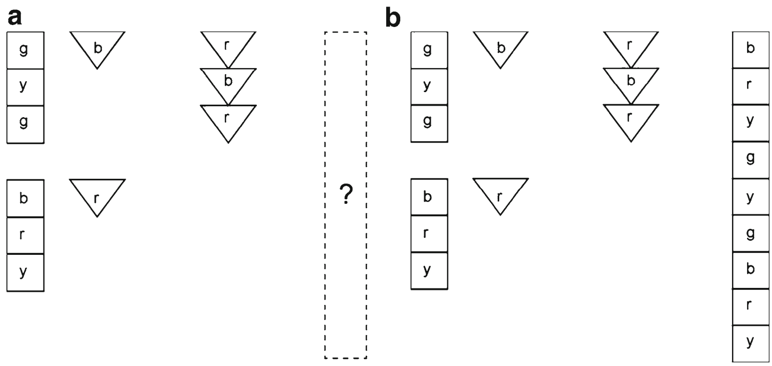


Fig. 4.2 The Pegboard as a programming tool (**a**) The subject is confronted to this (**b**) What the subject has to produce (execution of the program)

exercises require the reconstruction of the program (Fig. 4.3a) or the discovery of the procedures used (Fig. 4.3b).

In all these tasks, the subject is confronted to a nonverbal form of hypothetico-deductive reasoning.

Second Example: The Dynamical Mazes

The Dynamical Mazes have been invented by Cohors-Fresenborg (1978). This device consists of a baseboard provided with holes forming a square grid and several types of pieces which can be placed on this baseboard and combined (Fig. 4.4)

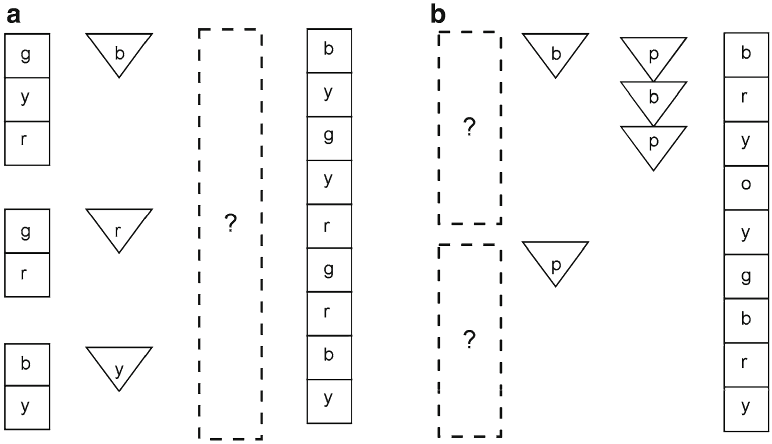


Fig. 4.3 Pegboard exercises. (**a**) Reconstruction of the program (**b**) Discovery of the procedures used

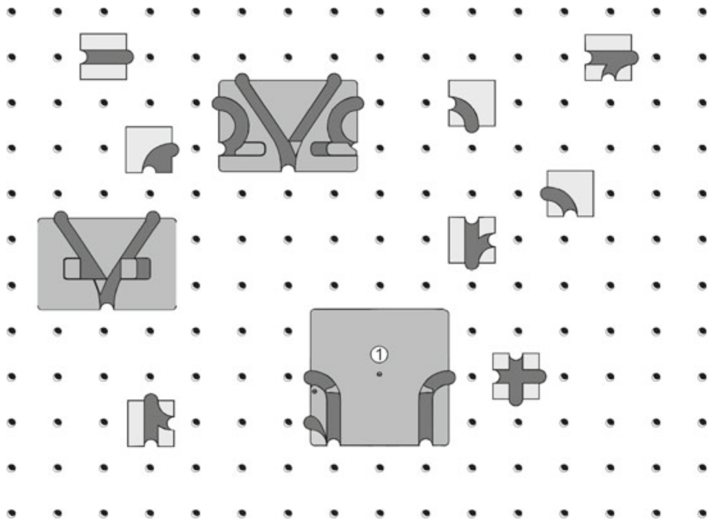


Fig. 4.4 The bricks: simple pieces, switches, and counter created by Cohors-Fresenborg

in order to form a finite automaton. The different types of construction material are "rails" (straight ones, curves to the right or to the left, crossings, dynamical switches, etc.). These pieces enable the user to build a network which is in fact a finite automaton. This material has built-in constraints which restrict the possible actions carried out by the subjects. Cohors-Fresenborg created this device to introduce teenagers to the notion of algorithm: he presented them verbal problems and asked them to construct "the network which is the solution".

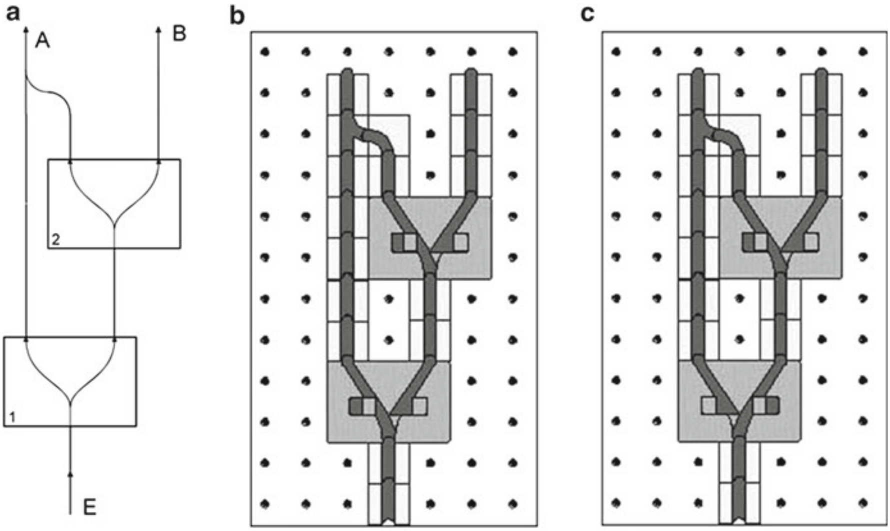


Fig. 4.5 Dynamical Mazes. (**a**) A diagram for a network (**b**) The network before the passage of the first "train" (**c**) The network after the passage of the first "train"

We chose to use this device in a different way: we confront the subject with the diagram representing a network (Fig. 4.5a) and ask him/her to actually build it with the bricks (Fig. 4.5b). Our networks have only one entrance but can have several exits. A little stick (called "train") going through the network will be forced to go to one precise exit by the actual positions of the switches, but while passing through the network, the stick can modify these positions (these positions define the "inner state" of our finite automaton). We then suggest the subject to "explore" the network by letting this "train" go through it and observe what happens. This can lead the subject to the discovery of regularities imbedded in the network (position of the switches, list of exits, etc.). We can then ask him/her "what is this network good for?" (Lowenthal, 1980, 1984).

By convention, at the beginning of an exploration, all the switches are open to the left. The first train will enter the network in E (Fig. 4.5a), use the left path of the first switch (Fig. 4.5b), and thus open the right path for the next "train," changing the inner state of the finite automaton (Fig. 4.5c). The first "train" will exit the network via exit A. The next "trains" will exit, respectively, through exits A, A, B, and A again. At this moment the inner state of the finite automaton is identical to that observed after the passage of the first "train." A possible answer to the question "what is this network good for?" could thus be "counting by 4."

This type of exercises enables the researcher to observe how subjects realize short-term, middle-term, and long-term predictions. Questions can be asked such as the following: which exit will the "train" go through after train 6, after train 11, and after train 101? These questions can best be answered after the realization by the subject of a "synthesis table" such as that shown in Table 4.1.

Table 4.1 An example of a synthesis table "Mechanisms" represent the inner state of the network when "the train" is introduced in the network

	Mechanisms		
Trains	Mechanism 1	Mechanism 2	Exit
1	L	L	A
2	R	L	A
3	L	R	A
4	R	R	B
5	L	L	A
6	R	L	A
7	L	R	A
8	…	…	…

L (open to the) left, *R* (open to the) right

In order to make correct predictions, the subject must discover the regularities of the network he/she is using. As far as short-term predictions are concerned, this can be done simply by looking at the list of exits, but correct long-term predictions require the subject to read the synthesis table simultaneously horizontally and vertically (the fact that line 1 and line 5 of the table contain the same information already gives the solution). It must be noted that here the subject is also confronted to a nonverbal form of hypothetico-deductive reasoning.

Results Observed After Using NVCD-Like Approaches

With Normal Subjects

Manipulations of the Pegboard by 1st and 2nd graders favor the acquisition of more structured verbal productions (Lowenthal, 1990) and of decoding (reading) abilities (Soetaert, 2003). The use of an approach based on manipulations of Dynamical Mazes favors the acquisition of reading abilities in 1st graders (Lowenthal, 1986), a better knowledge of arithmetic facts and a development of the visuospatial analysis in 6-year-old children (Lefebvre, 2002), and better verbal productions (Lowenthal, 1984; Yang, 2005) in 7-year-old children who used this device when they were 6. This type of approach also favors the development of formulation, testing, and adaptation of hypotheses (Lowenthal, 1984, 1992).

With Subjects with Mental Retardation or with Dyslexia

NVCD-like approaches have also been used with subjects with severe mental retardation (Bordignon & Vandeputte, 2005; Bordignon, Vandeputte, & Lowenthal, 2006;

Lowenthal & Vandeputte, 2006), with autistic young adults with Asperger's syndrome (Bordignon, 2004a, 2004b), and with dyslexic children (Bordignon & Soetaert, 2002; Pieters, 1999; Soetaert, 2002, 2003). In all cases, progresses were observed in planning, structuring, attention span, verbal production, and language understanding.

With Psychiatric Patients

The Pegboard and the Dynamical Mazes are presently tested in a clinical research concerning adult schizophrenic patients. The first results show that the patients consider these activities as difficult, but that they are glad to cooperate. They show more attention and are more focused on the task. They work efficiently during longer periods, organize better space and time, and use the verbal language in a more appropriate way to communicate efficiently. It is possible that this approach can serve as cognitive remediation for schizophrenic patients because it is a form of "learning without failure" and because it is a nonverbal approach to hypothetico-deductive reasoning, detached of all forms of affective factors.

With Subjects with Localized Cerebral Lesion

A set of clinical studies has shown that NVCD-type approaches offer the possibility for subjects having lost cognitive functions to reacquire them, at least partially. Some interesting results have been observed in patients with focal brain injuries, particularly for redeveloping some language functions (Ledru & Lowenthal, 1988; Lowenthal, 1991; Lowenthal & Saerens, 1982; Mauro, 1990). Patients reconstruct partially a structured communication and other superior cognitive functions. This is best exemplified by a case study: Saïd (Lowenthal & Saerens, 1986). Saïd's development was normal until the age of 13 months. He was starting to produce one-word sentences and to understand verbal language, when he contracted a meningoencephalitis followed by a right hemiplegia. He recuperated from the hemiplegia but was left with a pseudobulbar syndrome which made him unable to articulate. He also had important bilateral cerebral lesions in the language areas. As a result of all this, Saïd lost all possibilities to produce or understand a structured language. All traditional therapies had failed. When Saïd was 5, Lowenthal and Saerens (1986) started a therapy based on NVCD-like approaches. After 3 years, Saïd was able to understand the verbal language, communicate in a structured fashion, read, write, and compute.

These observations suggested a conjecture (Lowenthal, 1999): "the manipulations implied in NVCD-like approaches can favor a cerebral reorganization." This hypothesis will be examined in the next section.

Language Development by a Nonverbal Training: An fMRI Study

Considerations About Brain and Language

It is now commonly accepted that the brain structures underlying language are well distributed in the brain. Many results tend to prove that language activities imply different parts of the cortex, sometimes not specifically dedicated to language itself. Similar arguments have been proposed by Ullman (2001, 2004; Ullman et al., 1997) who claimed in his Declarative/Procedural Model that some parts of the brain are involved both in language activities and in nonverbal activities. He highlights two principal areas active during a language task: the left superior temporal gyrus, underlying the semantic part, and the left inferior frontal gyrus and the basal ganglia, active when subjects use grammatical structures. Interestingly, these structures are also activated during many nonverbal cognitive tasks, as the implicit procedural learning (Eichenbaum & Cohen, 2001), the probabilistic rule learning (Knowlton, Mangels, & Squire, 1996 ; Poldrack, Prabhakaran, Seger, & Gabrieli, 1999), or the sequence learning (Aldridge & Berridge, 1998; Peigneux et al., 2000).

Interestingly also, a similar hypothesis has been proposed in a completely different context. Notably, in developmental psychology, Saffran, Aslin, and Newport (1996, see also Saffran, 2001) observed that 8-month-old infants are able to discriminate nonwords vs. "words" of an artificial language, on the basis of the probabilistic structures of this "language" only. Seidenberg (1997) interpreted this observation as the proof that the probabilistic constraints in the learning processes can favor the development of cognitive structures. According to these authors, an individual must discover the regularities of his/her environment to acquire the language ability.

The present study goes one step further in that direction: we postulate that these regularities can be provided by nonverbal examples and that it is possible to develop the activation of the cortical structure underlying them (i.e., basal ganglia).

Our fMRI Study

In order to test the above-mentioned hypothesis, we chose to confront the subjects to logical probabilistic problems presented in the context of an NVCD-type approach with the Dynamical Mazes.

On the basis of Ullman's and Saffran's results, we formulated the hypothesis that an intensive nonverbal training based on the detection of probabilistic regularities, in a situation implicating few simple but specific rules, could lead to a functional reorganization of language-related activations.

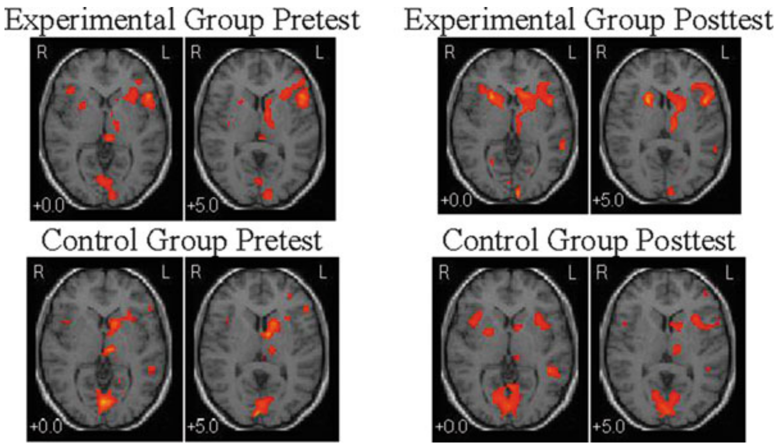


Fig. 4.6 Subcortical activities before and after Dynamical Mazes manipulations

General Study Design and Subjects

Twenty healthy French-speaking volunteers (14 females and 6 males, 18–20 years old) participated in this study. All subjects were right handed by self-report and scored between 75 and 100 on the modified Edinburgh handedness scale (Ransil & Schacter, 1994). Subjects' selection took place about 1 week before the pretest, and selected subjects gave their informed consent. This study was approved by the Ethics Committee of the University of Mons (Belgium). The "Experimental Group" (EG), composed of ten subjects, has been submitted to 4 weekly 1-h training sessions (Lefebvre, 2005; Lefebvre, Baleriaux, Paquier, & Lowenthal, 2006) during which they were confronted individually with logical exercises presented through the Dynamical Mazes. There was no specific cognitive activity planned with the "Control Group" (CG) subjects.

Results

The principal result concerns subcortical activities. In fact, during the pretest, a similar activity was observed for all of our subjects (Experimental (EG) and Control (CG)) when they had to generate verbs (see Fig. 4.6). However, after our treatment, we observe for the EG a great activation increase when for the control group, no subsequent significant activation was highlighted.

More precisely, we observe in the EG a greater activation in the bilateral caudate nucleus and the globus pallidus structures after the training sessions. While the comparison between groups is nonsignificant at the pretest ($Z=0.177$, $\alpha=0.912$), a significant difference is highlighted at the posttest for the experimental group

($Z=2.8$, $\alpha=0.023$). Duffau, Bauchet, Lehericy, and Capelle (2001) have equally shown that this area is implicated in the motor sequence programs, particularly if they are relatively complex. Our sequencing activities training, based on the perception of regularities presented by the manipulation of concrete tools, seems to develop an activation of the premotor cortex and subcortical structures. We make the hypothesis that our results could explain the link, postulated by Saffran, Newport, and Aslin (1996), that probabilistic abilities and language are strongly connected. If we refer to Ullman's researches (2001, 2004), we can observe that basal ganglia are implicated during verbal activities, but equally during nonverbal activities. We think that the specificity of our tools, which allows learning at a probabilistic level, has developed the faculty of our subjects to elaborate language words. It is possible that nonverbal probabilistic exercises, which stimulate basal ganglia, develop some other cognitive activities, such as language, also supported by this structure.

NVCDs and Recursion

The results mentioned in the fMRI observation mentioned above clearly show that it is pertinent to assume that the use of an NVCD-type approach is associated with some sort of cerebral plasticity. A new question can thus be raised: "why is it so?"

In the light of Hauser, Chomsky, and Fitch's paper (2002) on the role of recursion in human communication, it seemed relevant to examine more closely what actually happens in verbal language and whether this could be influenced by the use of special NVCD-like exercises.

We often use short sentences, such as "It rains" or "Could you please give me the salt?" But we also use more complex sentences when we must specify about which specific object we are speaking: "The key is broken," taken out of its context, does not yield much information. We better specify "The door key is broken," or, if our house has two doors, it is better to say "The garden door key is broken." We could also consider the case of the lucky owner of two houses, one on Liverpool street and one on Oxford street: he/she should specify "The garden door of the Liverpool street house is broken," or better "The Liverpool street garden door is broken." The speaker could of course do all this by using short sentences, uttered one after the other: "The key is broken. It is the door key. The door is the garden door. The house is on Liverpool street." One would say in this case that the speaker expresses herself by ***concatenation***, i.e., by a simple juxtaposition of simple expressions. Unluckily, this process is neither elegant nor economic: we often prefer the use of expressions full of subordinates and even full of **nested** subordinates. Apparently, French and English do not behave in the same way. The embedded structures are frequent in English (or German): there are sentences such as "The cheese the rat the cat killed ate lay in the house Jack built." The meaning of such a sentence is obvious for an English-speaking person. A French speaker needs another type of structure based on the use of the *passive voice* and of explicit **relative pronouns** such as the following: "Le fromage **que** le rat, **qui** *a été tué par le chat*, a mangé se trouvait dans la maison **que** Jack a construite." When comparing an English and a French sentence, we must note that the French sentence is longer in number of words, but cognitively easier for our brain: the relatives are not as embedded as in the English equivalent, and the use of the passive voice enables to reduce the cognitive load. Sentences with nested subordinates are rather frequent in English and in German, but not in Latin languages. It must be stressed that sentences with embeddings are more economical, as far as formulation is concerned, but that this economy implies a cognitive overload: even if we do not like them, we are able to understand them. The embedded relatives which we encounter in such sentences represent precisely an example of the use of the principle of recursion. Or better, it is an example of a reduced version of this principle.

We thus believe that it is this "principle of embedding" which is relevant for our verbal language, rather than the full principle of recursion. As a consequence, it appears pertinent to examine whether NVCD-type exercises can be based on such embeddings and whether this fact can explain the results mentioned above.

In order to test this new hypothesis, we constructed new exercises for the Pegboard. Fortemps, Lowenthal, and Wautié (2014) describe how similar exercises can be constructed for the Dynamical Mazes.

"Recursion" and Pegboard Exercises

Exercises Based on Pegboard Manipulations Which Imply Some Recursion

In order to be able to do that, we introduced a new rule: the sequence of pieces defining a procedure can be composed of squares (basic elements) and of triangles which refer to otherwise defined procedures. Using this rule, it is easy to create "pseudo-recursive" exercises, as shown in Fig. 4.7a, b.

This enables the observer to simulate some kind of tail-recursive call (Fig. 4.7a) which can be viewed as similar to the French use of tail relative pronouns in "Peter placed the bow on the table which is in the bedroom." It is also possible, but less interesting, to simulate a pseudo-recursive program (Fig. 4.7b) which will never stop because the Pegboard has no counter. Nevertheless it appeared more fruitful to use embedded sequences as shown in Fig. 4.8a, b.

The exercise shown in Fig. 4.8a is more complex than the previous ones. In this case, the subject must first place the squares corresponding to "green triangle" (i.e., the yellow square followed by the red square). She must then start treating "orange triangle" by placing the orange square, suspend the execution of this procedure, and move one procedure up to place the content of "red triangle" (i.e., the yellow and red squares in that order) and then go back to "orange triangle" and place its last square. The subject can then move to the final procedure of the program and place the content of "red triangle."

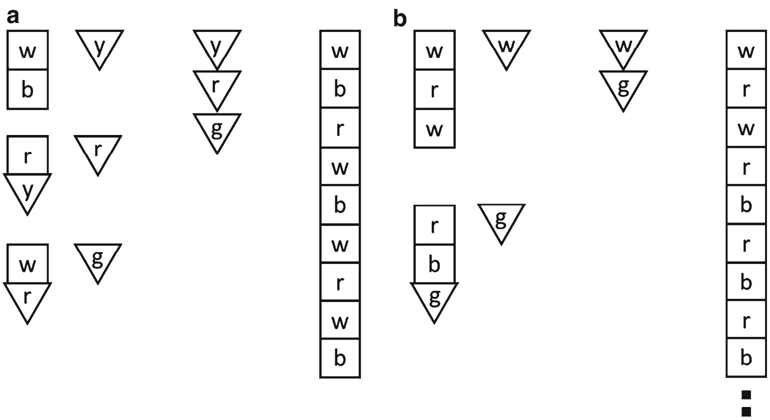


Fig. 4.7 Pseudo recursive call. (**a**) Non-recursive call (**b**) Endless "recursive" call

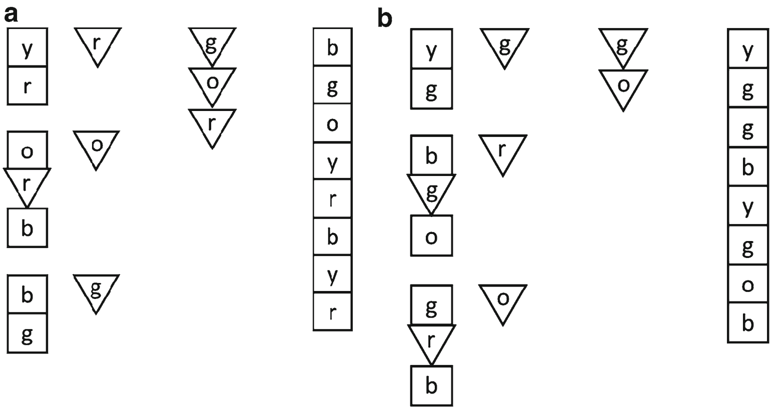


Fig. 4.8 Embeddings. (**a**) A simple embedding (**b**) A double embedding

The exercise presented in Fig. 4.8b is even more complex: in this case, after treating the "green triangle" procedure (by placing the yellow and green squares in that order), one must move to the "orange triangle" procedure. This procedure requires the subject to place a green square, suspend its execution to start the execution of the "red triangle" procedure, suspend this one to execute the "green triangle" procedure once again, and, after this is done, go back to finish "red triangle" and then finally finish the execution of "orange triangle." We clearly have here nested structures of the same type as those we could find in the sentence "The cheese the rat the cat killed ate lay in the house Jack built."

Results Obtained with These Embedded Sequences

We worked with 88 6-year-olds. They were tested using classical tests evaluating morphosyntax, verbal ability, and reading ability in order to enable us to create three equivalent groups. With the first group, we did nothing ("control" group), and the subjects of the second group were trained to use the Pegboard as described in Sect. "First Example: The Pegboard and a Programming Language" (i.e., without embedded sequences), while the subjects of the last group were trained to use the Pegboard as described in Section "Exercises Based on Pegboard Manipulations Which Imply Some Recursion" (and more specifically using exercises similar to those presented in Fig. 4.8a, b).

During the posttest, the subjects were shown a short animated cartoon showing the beginning of the adventures of a little horse, a skunk, and a dog. After this, each subject had to describe verbally to a younger child what he/she had seen and what, according to him/her, "was the end of the story." The results show that subjects trained to use the Pegboard as described in Section "First Example: The Pegboard and a Programming Language" (i.e., without embeddings) use significantly more relative pronouns than the two other groups. They also use significantly more the exact terminology, instead of a metaphor, than the two other groups. There is no significant difference between the "control group" and the "embeddings group." This could be due to the fact that the cognitive overload associated with exercises such as those presented in Fig. 4.8a, b counterbalances the positive action of the standard use of the Pegboard.

These facts lead to a double conclusion. Firstly, an NVCD-like approach as simple as the Pegboard can favor the richness of verbal productions; secondly, embedded sentences and recursion have nothing, or very little, to do with this fact.

"Recursion" and Dynamical Maze Exercises

It is easy to include some recursive elements in the sketches used as starting point for the construction of networks. Either we force the little "train" to go twice through the same switch before reaching an exit (Fig. 4.9a)—this implies a double possibility to change the orientation for the next "train"—or we simply use a special piece functioning as real counter (Fig. 4.9b).

We did not use yet this type of networks to test the subjects' reactions to "recursive" vs. "non-recursive" networks. The recursive networks are delicate to create, and we had to wait until an automatic construction network was available. The principles guiding the creation of such a software by Wautié are described in Fortemps et al. (2014).

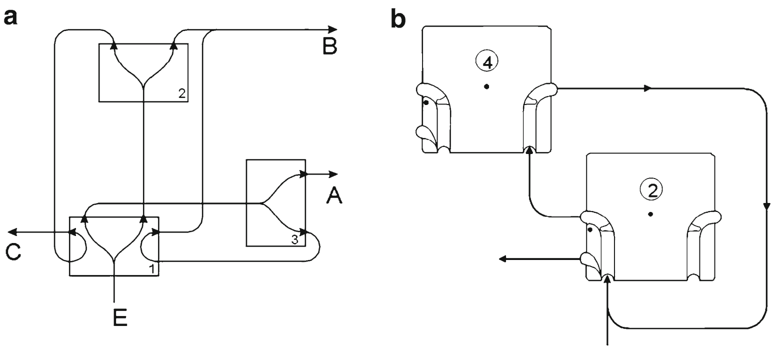


Fig. 4.9 Dynamical Mazes and loops. (**a**) A network with double possibility to modify the first switch (**b**) An "addition" network based on two counters, with forced exit when the right counter reaches 0

Discussion and Conclusion

Recursion and embeddings seem to play an important role in language: either in language use, as mentioned by Hauser, Chomsky, and Fitch (2002), or in language acquisition, as suggested by Seidenberg (1997). NVCD-like approaches seem also to favor language development or language reacquisition in the case of patients with focal cerebral lesions. These approaches can also be used in the context of a cognitive therapy for psychotic patients. Moreover an fMRI study (Lefebvre et al., 2006) based on the use of Dynamical Mazes shows that an NVCD-like approach induces some kind of cerebral reorganization in the context of a language task, and an approach based on the Pegboard can also be used with young normal children: results show that this type of approach favors the richness of verbal productions, more specifically in the field of relative pronouns.

It seems thus reasonable to suggest that our approach has a real influence in the field of language, but that this influence is linked neither to the apparent recursive dimension imbedded in some NVCD-like exercises nor to the embeddings associated with some of the exercises.

For all these reasons, we feel confident in suggesting that this influence is due to two factors. Firstly, NVCD-like standard approaches (without embeddings) enable the subject to perform some kind of "learning without failure" linked to the existence of constraints in the material and to the quasi nonverbal dimension. Secondly, these approaches give him/her the opportunity to be exposed to a nonverbal form of hypothetico-deductive reasoning without excessive cognitive overload. This is also associated with the fact that all the problem situations described here can be solved using any of the three Brunerian levels of representation (enactive, iconic, and symbolic).

In fact, the NVCD-type approaches seem to favor some form of implicit learning based on the discovery of probabilistic regularities.

References

Aldridge, J. W., & Berridge, K. C. (1998). Coding of serial order by neostriatal neurons: A "natural action" approach to movement sequence. *Journal of Neuroscience, 18*(7), 2777–2787.
Bordignon, T. (2004a). Est-il possible de développer des habiletés numériques chez des personnes présentant des troubles autistiques: Etude d'un cas. *APEPA, 89*, 7–11.
Bordignon, T. (2004b). Utilisation "d'outils structuré" pour favoriser le développement cognitif d'un adulte autiste. *Bulletins Scientifiques de l'Arapi, 12*, 38–40.
Bordignon, T., & Soetaert, M. (2002). Can a non verbal device help dyslexic children? *Dyslexia International – Tools and Technologies News letter*, *9*, 6.
Bordignon, T., & Vandeputte, C. (2005). Représentations concrètes de systèmes formels et outils mathématiques abstraits comme outils de remédiation, étude d'un cas: un enfant souffrant du syndrome X fragile. In B. Vivicorsi & R. Collet (Eds.), *Handicap, cognition et prise en charge individuelle: des aspects de la recherche au respect de la personne* (pp. 149–162). Rouen: Publications des universités de Rouen et du Havre.
Bordignon, T., Vandeputte, C., & Lowenthal, F. (2006). Comment favoriser le développement cognitif et la structuration de la pensée chez des enfants présentant un retard mental? In H. Gascon, J.-R. Poulin, J.-J. Detraux, D. Boisvert, & M.-C. Haelewyck (Eds.), *Déficience intellectuelle: Savoirs et perspectives d'action. Tome 2: Formation, interventions, adaptation et soutien social* (pp. 253–262). Québec: Presses Inter Universitaires.
Bruner, J. S. (1960). *The process of education*. Cambridge: Harvard University Press.
Bruner, J. S., Olver, R. R., & Greenfield, P. M. (Eds.). (1966). *Studies in cognitive growth*. New York: Wiley.
Cohors-Fresenborg, E. (1978). Learning problem solving by developing automata networks. *Revue de phonétique appliquée, 46*(47), 93–99.
Cordier, J., Héraux, C., & Lowenthal, F. (1975). Enseignement de la mathématique et exercices de verbalisation chez les enfants caractériels. *Enfance, 1*, 111–114.
Duffau, H., Bauchet, L., Lehericy, S., & Capelle, L. (2001). Functional compensation of the left dominant insula for language. *Neuroreport, 12*, 2159–2163.
Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection: Memory systems of the brain*. New York: Oxford University Press.
Fortemps, P., Lowenthal, F., & Wautié, V. (2014). Computer simulations of recursive exercises for a non-verbal communication device. In F. Lowenthal & L. Lefebvre (Eds.), *Language and recursion*. New York: Springer.
Frédérique. (1970). *Les Enfants et la Mathématique*. Bruxelles-Paris-Montréal: Didier.
Frédérique, & Papy, G. (1968). *L'Enfant et les Graphes*. Didier: Bruxelles-Paris-Montréal.
Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*, 1399–1402.
Ledru, L., & Lowenthal, F. (1988). Utilisation des NVCD dans une optique de diagnostique ou de recherche clinique. *Communication and Cognition, 21*, 17–27.
Lefebvre, L. (2002). Dispositifs concrets, méthode de lecture et compétences verbales. *Revue PArole, 21*, 1–34.
Lefebvre, L. (2005). Discovery of environmental regularities and modification of subcortical activations: A fMRI Study. *Abstracts of the Xth European Workshop on Mental Imagery and Cognition* (*EWIC*).
Lefebvre, L., Baleriaux, D., Paquier, P., & Lowenthal, F. (2006). Basal ganglia: A crossroad between verbal and non-verbal activities? *Acta Neurologica Belgica, 106*(Supplement), 65–66.
Lowenthal, F. (1980). Games, logic and cognitive development a longitudinal study of classroom situations. *Communication and Cognition, 13*(1), 43–63.
Lowenthal, F. (1983). Strategy games, winning strategies, and non verbal communication devices (at the age of 8). In R. Hershkowitz (Ed.), *Proceedings of the Seventh International Conference*

for the Psychology of Mathematics Education (pp. 364–368). Rehovot: Weizmann Institute of Science.

Lowenthal, F. (1984). Production langagière d'enfants manipulant un dispositif non verbal de communication. *Revue de phonétique appliquée, 69*(11), 46.

Lowenthal, F. (1985). Pegboard as basis for programmation in 5 and 6 year olds. In *Proceedings of the 9th International Conference in Psychology of Mathematics Education* (pp. 47–52). Berkeley.

Lowenthal, F. (1986). NVCDs are structuring elements. In L. Burton & C. Hoyles (Eds.), *Proceedings of the Tenth International Conference for the Psychology of Mathematics Education* (pp. 363–368). London

Lowenthal, F. (1990). Pegboard as structuring element for the verbal language. *Revue de Phonétique Appliquée, 95*(97), 255–262.

Lowenthal, F. (1991). Logic and language acquisition. *Behavioral and Brain Sciences, 14*(4), 626–627.

Lowenthal, F. (1992). Formalisms for clinical observations. In F. Poyatos (Ed.), *Advances in nonverbal communication* (pp. 125–143). Amsterdam/Philadelphia: John Benjamin.

Lowenthal, F. (1999). Can handicapped subjects use perceptual symbol systems? *Behavioral and Brain Sciences, 22*(4), 625–626.

Lowenthal, F., & Saerens, J. (1982). Utilisation de formalismes logiques pour l'examen d'enfants aphasiques. *Acta Neurologica Belgica, 82*, 215–223.

Lowenthal, F., & Saerens, J. (1986). Evolution of an aphasic child after the introduction of NVCDs. In F. Lowenthal & F. Vandamme (Eds.), *Pragmatics and education* (pp. 301–330). New York: Plenum.

Lowenthal, F., & Vandeputte, C. (2006). Une remédiation cognitive pour l'acquisition des prérequis de la lecture et du nombre. *Actes du Congrès X fragile – Europe*. Liège, p. 22.

Mauro, A. (1990). *Observation de deux enfants polyhandicapés à l'aide de manipulations concrètes de systèmes formels*. Mémoire de post-graduat. F.S.P.P., Université de Mons-Hainaut.

Papy, G. (1964). *Mathématique Moderne 1*. Bruxelles-Paris-Montréal: Didier.

Peigneux, P., Maquet, P., Meulemans, T., Destrebecqz, A., Laureys, S., Degueldre, C., et al. (2000). Striatum forever, despite sequence learning variability: A random effect analysis of PET data. *Human Brain Mapping, 10*(4), 179–194.

Pieters, V. (1999). *Améliorer les performances en lecture chez des enfants mauvais lecteurs: est-ce possible à l'aide des Labyrinthes Dynamiques ?* Mémoire de licence: Université de Mons-Hainaut.

Poldrack, R. A., Prabhakaran, V., Seger, C. A., & Gabrieli, J. D. (1999). Striatal activation during acquisition of a cognitive skill. *Neuropsychology, 13*(4), 564–574.

Ransil, B. J., & Schacter, S. C. (1994). Test-retest reliability of the Edinburgh handedness inventory and global handedness preference measurements, and their correlations. *Perceptual Motor Skills, 79*, 1355–1372.

Saffran, J. R. (2001). Words in a sea of sounds: The output of statistical learning. *Cognition, 81*, 149–169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

Soetaert, M. (2002). *Apport de la mosaïque auprès d'enfants présentant des troubles de l'apprentissage de la lecture*. Mémoire de licence. Université de Mons-Hainaut.

Soetaert, M. (2003). Développer la conscience segmentale d'enfants présentant un retard dans l'apprentissage du langage écrit en utilisant un dispositif ludique. *Revue PArole, 27*, 159–206.

Ullman, M. T. (2001). The declarative/procedural model of Lexicon and Grammar. *Journal of Psycholinguistic Research, 30*(1).

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., et al. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience, 9*(2), 266–276.

Yang, Y-J. (2005). *L'approche cognitive du langage, Rapport interne*, Service de Sciences Cognitives, Université de Mons, p. 21.

Chapter 5
Computer Simulations of Recursive Exercises for a Nonverbal Communication Device

Philippe Fortemps, Francis Lowenthal, and Vincent Wautié

Optimal and Reference Dynamical Maze Exercises

The problem situations used by Cohors-Fresenborg (1978) for his Dynamical Mazes were basically the following: he described verbally a situation and asked the subject to create a finite automaton which would "give the solution." This author's word problems were similar to this one: "Create a machine which will give a stamp for four coins." This can be viewed as follows: a list of exits was given (e.g., "more coins 1," "more coins 2," "more coins 3," "your stamp is available") and the subject had to create the finite automaton which would yield that list as output. The use of this device in an NVCD-type approach (Lowenthal, 1984; Lowenthal & Lefebvre, 2014) is based on the reverse problem: given a finite automaton, the subject has to find the list of exits produced by this machine. This implies that the researcher must **first** construct an adequate network. This task can be arduous and the result can be biased. The creation of an ad hoc network has been done previously by trial and error. It is not obvious to create in such a way a network with exactly the required complexity, a network which can serve as reference and which is not biased. The purpose of this chapter is to describe an algorithmic method enabling the researcher to create networks in order to test the influence of the possible "recursive" or "embedded" dimensions on the cognitive development and more specifically on language development. The essential aim is to have a method such that each network constructed in this way can be used as reference since it has been constructed according to an algorithm and is thus detached of the experimenter's biases.

P. Fortemps (✉) • V. Wautié
Computer Science, Management and Engineering, University of Mons, Mons, Belgium
e-mail: philippe.fortemps@umons.ac.be

F. Lowenthal
Cognitive Sciences, University of Mons, Mons, Belgium

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_5,

Construction of the Networks

Each Dynamical Maze network is a finite automaton built with a finite number of pieces: this entails that the list of possible outputs constitutes a sequence which has always a finite length and is repeated indefinitely, sometimes after an initial segment (Trakhtenbrot & Barzdin, 1973). In this chapter, the complete sequence of outputs produced by a network is called "pattern." The (possible) initial segment is called "head," and the repeated sequence of exits is called "sequence." We define the period of a pattern as the length of its "sequence".

We present here a methodology to build, in three steps, a set of potential networks for a given pattern and describe how to choose a most economic one among the proposed solutions. Before entering the description of these three steps, we must first give more details about the Flip-Flop (FF) we use and explain why we chose this element as basis for our construction.

The Flip-Flop

In this first study of the automatic creation of Dynamical Maze networks, we consider only the simplest element: the Flip-Flop, also known as automatic switch (Fig. 5.1). The behavior of this kind of element can be modeled by a Petri net (Fig. 5.2) (Peterson, 1981). After passing the upper transition, a primary (black) token (called "train" in the Lowenthal and Lefebvre paper) enters a node (IN) of the Petri net that allows it to move either to the left or to the right. The actual destination of the primary token is determined by the internal state of the Petri net, described by the gray token. If the gray token is in node SL (status-left), the primary token will go to node OL (output-left). Simultaneously the gray token moves to node SR (status-right). If a new primary token (i.e., "train") enters the FF, it will go through it and go to node OR (output-right), while the gray token moves back to SL. We can thus note that the FF alternately directs the token to its left output or to its right output. In the remainder of this chapter, FF nodes will be represented by squares. Since we limit ourselves to something as simple as an FF, we can only reproduce

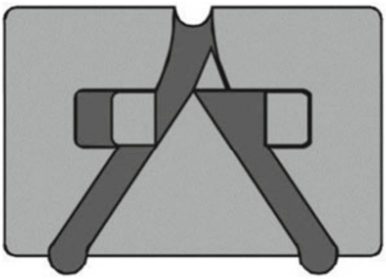


Fig. 5.1 The FF is alternately open to the *left* and to the *right*

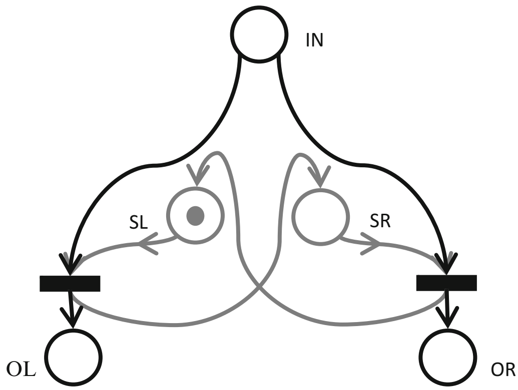


Fig. 5.2 Petri net representation of the basic FF (the *gray dot* represents the internal state)

patterns without a "head." To have the possibility to create a network taking into account an initial segment, we would have to consider other types of nodes: counters, conditional switches… The counter is the only way to introduce a "zero test" in our system.

The FF has a major advantage since it can be treated as a propositional calculus variable: the left exit can be associated to "True" and the right exit to "False." A complete network can thus be viewed as the representation of a formula of some kind of propositional calculus. It is then normal to expect that a network can be built to "produce" any given list of exits (cf. Theorem of Normal Disjunctive Form in propositional calculus).

First Step: Networks Producing a Power-of-Two Period Patterns

We first observe that any sequence which pattern period is a power of two can be produced by a simple structure using only FFs. Indeed, since the FF has two outputs, it is interesting to combine these nodes in a tree structure, more precisely, in a binary tree structure which has then a total number of outputs equal to a power of two: e.g., the tree shown in Fig. 5.3 is adapted to a pattern of period 8 which leads to a tree structure with three levels of FFs (and therefore requires 7 FFs) and a final level with the eight separate outputs. More generally, for an output sequence of length 2^n, the number of FFs required is $\sum_{k=0}^{n-1} 2^k = 2^n - 1$.

If we want to create a network with eight different exits (i.e., a sequence of eight outputs repeated indefinitely), named $\mathbf{E_0}$, $\mathbf{E_1}$, $\mathbf{E_2}$, $\mathbf{E_3}$, $\mathbf{E_4}$, $\mathbf{E_5}$, $\mathbf{E_6}$, $\mathbf{E_7}$, we can limit ourselves to FFs (cf. Fig. 5.4). But we must note that if we let numbered "trains" Y_0, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7 go through this network, they will not reach the eight exits in a left to right numerical order.

In fact, the "trains" will reach the exits in an order determined by the logic imposed by the inner structure of the FFs and the constraints associated to it. To know the

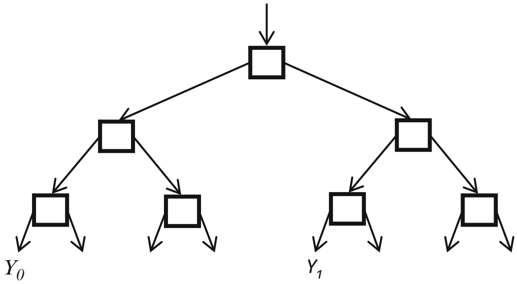


Fig. 5.3 A sketch for a network combining seven FF via "rails" and producing eight exits

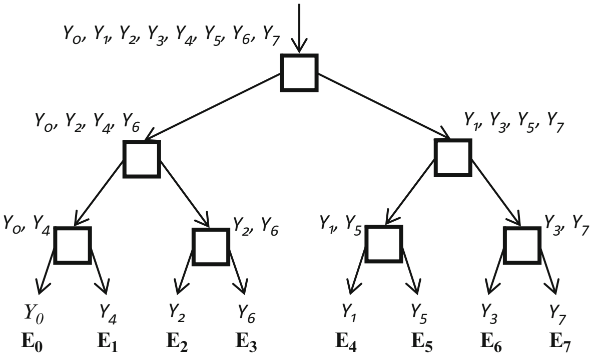


Fig. 5.4 A FF binary tree structure and the order of the associated exits

exact order in which the "trains" will reach the exits, we must note that each FF breaks the cycle of outputs into two sub-cycles, distributing the "trains" alternately to the left and to the right. In the example shown in Fig. 5.4 for this sequence of eight trains, Y_0, …, Y_7, and eight exits, the very first FF breaks this sequence of "trains" into two subsequences: the first one with the even-index labels and the second one with the odd-index labels, the second FF will break the subsequence reaching it into two other subsequences, and so on. More precisely, the behavior of the FF binary tree is such that it sorts the labels according to the binary representation of their index. In this example with labels numbered from 0 to 7, the indices have the following binary representations: 0→000, 1→001, 2→010, 3→011, 4→100, 5→101, 6→110, 7→111. If we sort them lexicographically, reading these binary numbers in reverse order (from the right most digit to the left most one), we obtain 0, 4, 2, 6, 1, 5, 3, 7. So, if we look at the exits of this network, we can see that the "train" Y_0 will reach the left most exit, that the next exit (counting from the left) will be reached by the "train" Y_4, and so on, yielding (from left to right) the apparent sequence of "trains": Y_0, Y_4, Y_2, Y_6, Y_1, Y_5, Y_3, Y_7 (as shown in Fig. 5.4).

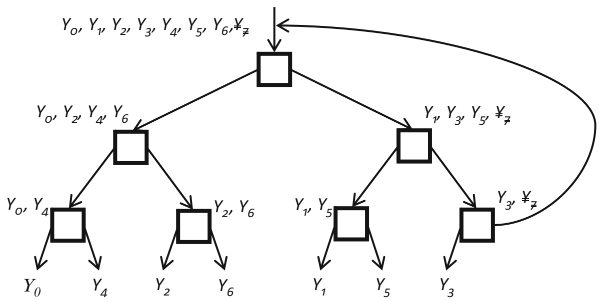


Fig. 5.5 A FF binary tree structure with seven outputs and one back-loop. The eighth output is redirected to the entry as unnecessary

Once the eight outputs of the FF binary tree have been produced, the network is back into its initial inner state. Therefore, the sequence of eight exits will be repeated over the next runs, indefinitely.

Second Step: Networks Producing Any Period Patterns

When the length of the given sequence is not a power of two, we can extend it to an appropriate length using a special technique. We consider an extended pattern of power-of-two period immediately bigger than the actually desired period, in order to compute its FF binary tree easily, but we are then confronted with undesirable outputs. One weak way to deal with them is to leave them void. A more interesting solution is to loop them back to the input of the tree. In this way, when the token or "train" (Lowenthal, 1984) comes out of these additional outputs, it is redirected to the entry of the tree and performs a new run into it (see Fig. 5.5 for a sequence with seven labels).

With this trick, we can avoid unnecessary exits: "train" Y7 will go to the same exit as "train" Y_0. When multiple back loops must be used, we must decide how to organize them. Again, the binary representation of numbers will help. Indeed, when the two outputs of a given node have to be back looped, we choose to back loop the arc entering this node, since it enables us to finally delete this node (as shown in Fig. 5.7). This method can then be applied to several successive levels from bottom to top. For a sequence of length N, let 2^n be the smallest power of 2 such that $N \leq 2^n$. Let $R = 2^n - N$ be the number of unnecessary outputs. The binary representation of R indicates the levels in the tree where a back loop has to be made, counting the levels from the bottom and associating the level 0 to the initial outputs.

Let us consider a pattern of period 9 for which we want to create a network. We first consider a tree capable of producing a pattern of period 16 (Fig. 5.6) and then create back loops for seven individual outputs. But, instead of looping seven separate outputs, we can use a more compact way of looping (Fig. 5.7). Since $7 = 2^2 + 2^1 + 2^0$, a back loop has to be made on each of the three levels 0, 1, and 2. At level 0, a single output is back looped; at level 1, two outputs are back looped and

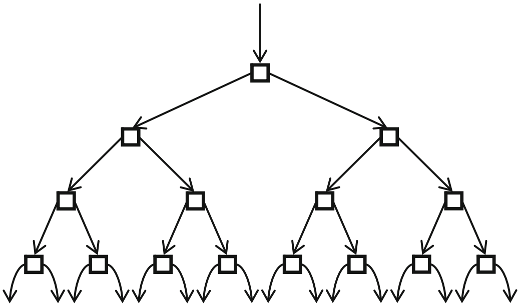


Fig. 5.6 A tree structure with 16 outputs, the basic structure for any number of outputs between 9 and 16

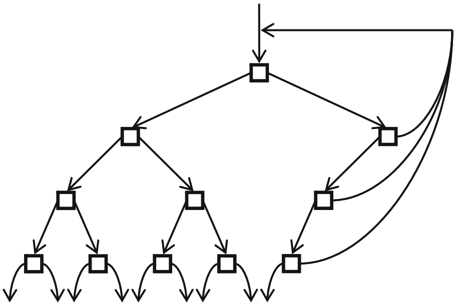


Fig. 5.7 A tree structure with nine outputs and only three back-loops

one node is deleted; at level 2, a group of four outputs are back looped and three nodes are deleted.

In a sense, the previous step consists in extending the initial pattern by a number of dummy outputs. These dummy outputs, once looped back to the network input, are added at the right end of the tree structure. In other words, the tree is built as to reproduce the pattern associated to $Y_0,Y_1,Y_2,*,Y_4,*,Y_6,*,Y_8,*,Y_{10},*,Y_{12},*,Y_{14},*$ which effectively enjoys nine actual outputs and seven dummy ones (denoted by *).

Third Step: Optimizing the Network

The previous section enables us to build a simple tree structure to reproduce patterns of any period. The proposed solution is optimal if the pattern is composed of different labels. In such a case, the only regularity of the pattern consists in its overall period. But a pattern can consist of labels with their own frequency, i.e., labels that appear more than once in the pattern following their own regularity. This internal regularity leads to sub-patterns that must be captured. Two different strategies can be considered, either by selecting the positions of the dummy labels or by factorizing the network into a tree of primitives.

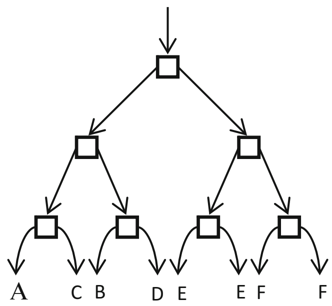


Fig. 5.8 A tree structure for AEBFCEDF without pruning

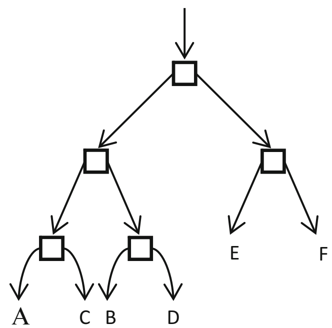


Fig. 5.9 A tree structure for AEBFCEDF after pruning

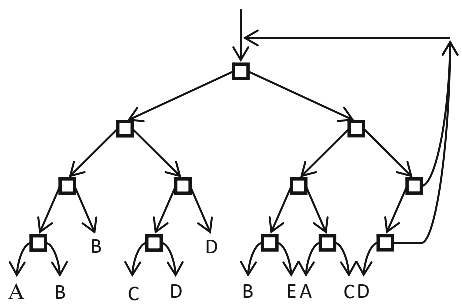


Fig. 5.10 A tree structure with the basic localization of the dummy outputs

Selecting the Positions of Dummy Labels

Let us consider as an example the initial pattern AEBFCEDF (to be repeated) with period 8. As seen in Section "Second Step: Networks Producing Any Period Patterns", we are able to model it by a simple binary tree structure (Fig. 5.8). However, when we build this tree, we observe that for some nodes, both outputs lead to the same label (in our example, E and F). We must then, when we construct the real network, take care to join after the FF tree the two exits leading to E and then do the same for the two exits leading to F. Nevertheless, it is worth noting that

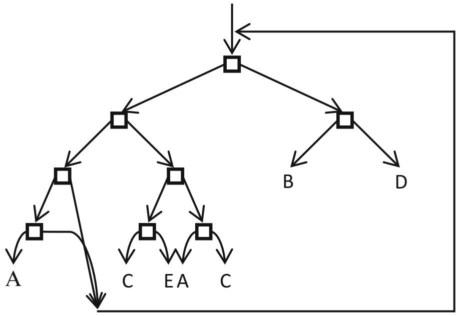


Fig. 5.11 A tree structure with optimized localization of the dummy outputs

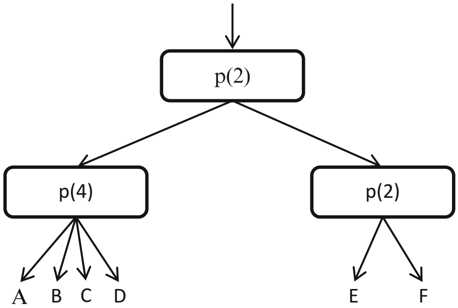


Fig. 5.12 A tree of primitives (cf. Fig. 5.9). In such a tree, the labels of a primitive are delivered from *left* to *right*

in such a case, the tree may be simplified by removing the nodes leading to identical outputs. If needed, this pruning mechanism can be applied several times. On the example mentioned above, it allows for removing two useless nodes (Fig. 5.9).

For more complex examples, the effects of pruning can be more consistent, especially if one inserts or moves intelligently dummy outputs. For example, consider the sequence ABCDBADBEDBCD with length 13. Should we simply follow the method described above, we would complete the sequence by three dummy outputs. Their location in the right of the tree (Fig. 5.10) is similar to inserting some dummy labels (*) in the initial sequence to obtain the sequence: ABCDBAD*BED*BCD*.

However, if we place the dummy labels more subtly, we can exhibit a sub-pattern. Indeed, in ABCD*BAD*BED*BCD, there exists a sub-pattern (–B–D) of length 4. It may be also noted that the length of the sub-pattern (4) is a divisor of the overall extended pattern length (16). The tree for this adapted sequence (Fig. 5.11) is more compact than the previous one. While the initial tree uses 12 FFs, the compact one does not use more than eight FFs.

Factorizing the Whole Network

The appearance of subsequences leads naturally to the attempt to factorize the whole network in order to produce the extended pattern. So instead of considering

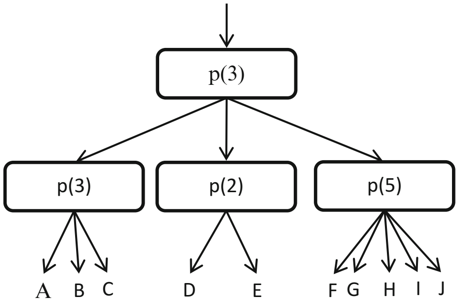


Fig. 5.13 A tree of four primitives able to produce a pattern of length 30

only simple binary tree structures, we consider recursive structures, consisting of primitives. A primitive *p*(*n*) is then a simple network (a binary tree structure as above) producing a pattern of fixed period *n*. Thus, for example, for the AEBFCEDF… pattern, we are led to determine a tree with three primitives (Fig. 5.12). Regardless of its internal structure, a primitive delivers its outputs in the order of its graphical representation from left to right.

The power of factorization is illustrated in Fig. 5.13, where a tree of four primitives (and only 16 FFs) is able to reproduce a sequence of length 30. It must be noted that the sequence of the produced pattern is in fact composed of some subsequences: A–B–C– of length 9, –D–E– of length 6, and –F–G–H–I–J of length 15. The length of each subsequence is the product of the number of exits in the subsequence and the degree of the top primitive.

The addition of dummy outputs and the finding of their best location greatly increase the complexity of the problem but promote the emergence of a recursive structure interesting for the researcher. This research is performed using genetic algorithms. To do this, an objective function is used to guide the solution research in the sense of minimizing the number of elements to be used to build the network. Other criteria could be tested, to promote certain forms of regularity in the adopted solutions.

Finding "the Best" Solution

In order to find easily "the best solution," one of us has created an "intelligent" software (Wautié, 2010) which can produce, for a given sequence of exits, either a complex solution with a minimal number of FF or an apparently simpler solution requiring more FFs.

Conclusion: Advantages of the Method Described Above

The method we described in Section "Construction of the Networks" offers several advantages.

Firstly it enables the researcher to create easily networks adapted to any given sequence of exits. This makes it possible to create adapted exercises for the approach described by Lowenthal (1984) by changing the complexity of the sequence subjects are supposed to discover.

Secondly this method enables the researcher to dispose of reference examples of networks, created without biases. This fact makes it possible to compare the subject's progresses in function of the level of complexity introduced in the exercises.

Thirdly, the method described in Section "Construction of the Networks" enables the researcher to create different exercises for a given sequence, using either an apparently simple presentation with many FFs (thus many inner states to describe) or a more complex presentation using less FFs. Moreover some of the networks produced by the software mentioned above have a recursive structure—and embeddings—which will enable us to test further the influence of these two factors on language acquisition and reacquisition.

References

Cohors-Fresenborg, E. (1978). Learning problem solving by developing automata networks. *Revue de phonétique appliquée, 46*(47), 93–99.

Lowenthal, F. (1984). Production langagière d'enfants manipulant un dispositif non verbal de communication. *Revue de phonétique appliquée, 69*, 11–46.

Lowenthal, F., & Lefebvre, L. (2014). Non verbal communication devices, language, cerebral flexibility and recursive exercises. In F. Lowenthal & L. Lefebvre (Eds.), *Language and recursion*. New-York: Springer.

Peterson, J. L. (1981). *Petri net theory and the modeling of systems*. Upper Saddle River: Prentice Hall.

Trakhtenbrot, B. A., & Barzdin, Y. M. (1973). *Finite automata. Behavior and synthesis*. Amsterdam, NY: North-Holland Publishing Company.

Wautié, V. (2010). *Modélisation et simulation de briques de langage en vue de reproduire des chaînes d'information – circuits séquentiels. Mémoire de Master Ingénieur civil*. Mons: Université de Mons.

Chapter 6
Implicit Learning and Recursion

M. Rohrmeier, Z. Dienes, X. Guo, and Q. Fu

Introduction

Implicit learning constitutes a remarkable human ability to acquire intuitive or unconscious knowledge of environmental structures (Dienes, 2012; Reber, 1989; on the knowledge being unconscious, contrast Shanks, 2005; for a somewhat intermediate position, Cleeremans, 2006). Implicit knowledge governs a large variety of everyday activities ranging from sports to music to language. For example, in the realm of perceptual motor skills, people appear to be able to learn to calibrate golf swings using information that is not conscious (Masters, Maxwell, & Eves, 2009) and can catch balls while consciously holding false theories about how they do this (Reed, McLeod, & Dienes, 2010). Native speakers are competent in the use of their language, yet cannot make explicit all the regularities they are actively following. The same holds for music: most people are competent listeners and perceive musical structure adequately (Bigand & Poulin-Charronnat, 2006; Deliège, Mélen, Stammers, & Cross, 1996; Honing, 2011; Koelsch et al., 2000), but cannot articulate underlying musical rules without education. Implicit knowledge is assumed to constitute the foundation of music or language cognition, being acquired incidentally from mere exposure and little explicit teaching (Rebuschat, 2008; Rohrmeier

M. Rohrmeier
Department of Linguistics & Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA

Z. Dienes (✉)
University of Sussex, Brighton, United Kingdom
e-mail: dienes@sussex.ac.uk

X. Guo
Shanghai Key Laboratory of Magnetic Resonance and School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

Q. Fu
Institute of Psychology, Chinese Academy of Sciences, Beijing, China

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_6,

& Rebuschat, 2012; Williams, 2009). Accordingly (first or second) language and music acquisition are prime instances of implicit learning (Loui, 2011; Rebuschat & Williams, 2011; Rohrmeier, 2010).

Most implicit or statistical learning studies, however, rely on established methodologies which in turn have implications for the complexity of the structures they use: they employ structures which are based on regular (finite-state) grammars or syllable or tone "words" (i.e. triplets) because they are well understood and have controlled properties suitable for experimental studies. However, in the ecological cases of language and music, the complexity of the underlying principles of structure building scarcely involves such artificial structures (e.g. "tone words") and exceeds Markov or finite-state models (cf. Rohrmeier & Rebuschat, 2012; Rohrmeier, Fu & Dienes, 2012). Furthermore, linguistic as well as musical syntax is governed by tree structures and underlying recursive structures or grammars as part of their principles of structure building (e.g. Chomsky, 1956, 1995; Hauser, Chomsky, & Fitch, 2002; Hofstadter, 1979; Jackendoff, 2011; Jackendoff & Lerdahl, 2006; Lerdahl & Jackendoff, 1983; Rohrmeier, 2007, 2011; Steedman, 1984). Hence there is a theoretical gap between the ecological structure of the domains and the structures by which the implicit acquisition is studied in the domain. Accordingly one core aspect of this gap concerns the concept of *recursion* as a core principle of structure building in language and music. Thus given that language and music employ recursive structures, the underlying learning and processing mechanisms have to be capable of dealing with structure that can be recursively generated. Before considering its empirical implications, we discuss the concept of recursion in the context of music and language.

The Concept of Recursion in Artificial Grammars

The relationship between the concepts of recursion and related debates concerning formal languages, the Chomsky hierarchy (and, particularly, non-finite-stateness), is not straightforward (cf. Fitch & Friederici, 2012; Lobina, 2011; Martins, 2012; Tomalin, 2007; for an account of implicit learning and recursion, see Rohrmeier, Dienes, Guo, & Fu, 2014). In this section we review the concept of recursion in formal languages—grammars as well as processes dealing with them—in order to provide a basic, somewhat less formal introduction to readers with different backgrounds and unfamiliar with the concepts and subsequently to relate these concepts to the empirical instantiations in music and language as well as the field of implicit learning.

Recursion serves as a construction principle for infinite sets. Grammars are sets of formal rules to describe a finite or infinite set of sequences (a *language*) through construction principles. While a finite number of finite sequences could be simply listed in a set, an infinite number of sequences (or an infinite sequence) requires an indirect definition to describe them as a set. The (mathematical) definition of an infinite set by a set of production rules is grounded in recursion (by virtue of the multiple applications of some production rules). Generally, an account of recursion is relatively straightforward:

1. A definition of a structure is recursive, if it includes itself as part of the definition.

A common mathematical example employing this definition is the Fibonacci sequence (0, 1, 1, 2, 3, 5, 8, 13, etc.) in which each term is defined as being the sum of the previous two terms (given 0,1 starts the sequence), i.e. $\mathrm{fib}(n)=\mathrm{fib}(n-1)+\mathrm{fib}(n-2)$ with $\mathrm{fib}(1)=0$, $\mathrm{fib}(2)=1$; note that the definition of the starting condition is crucial for the recursion to terminate. Definition (1) does not directly relate to formal grammars and further does not directly translate into a difference between finite-state and context-free grammars. When transferred to formal languages, the definition may state:

2. A formal language is recursive, if a sequence of productions leads to the production of a non-terminal symbol (as part of the string) that was part of its earlier production.

Terminal symbols are the elements that a string (i.e. sequence) is actually composed of, for example, words in a sentence or letters in the letter strings often used in artificial grammar learning experiments. Non-terminal symbols are variables that are used as part of a rewrite process in order to produce a final sequence of terminal symbols. For example, consider the following production (rewrite) rules[1]:

1. [0] → M[1]
2. [1] → T[1]
3. [1] → Q[2]
4. [2] → B[1]
5. [1] → Q

The letters M, T, Q and B are the terminals; the states [0], [1] and [2] are non-terminals. The first rule uses the starting symbol [0] to rewrite it according to the right-hand side as M and the nonterminal [1]. In other words, M can start a string. So [0] becomes M[1]. We still need to rewrite further (because the generation cannot terminate before there are no non-terminals in the string), which we can do by rewriting it as T and itself (rule 2). So M[1] becomes MT[1]. Rule 2 is thus recursive, by definition (2). Rules 3 and 4 together are also recursive (though not one directly visible in either rule alone) because the further production of the nonterminal [1] ultimately yields another instance of itself. Rule 5 produces the single symbol Q, which ends the production sequence.

Rewrite rules generate a set of strings (in our case, e.g. MQ, MTQ, MTTQ, MQBQ). The processing of such strings (in order to decide whether a string is part of the grammar or not) requires an *automaton*. Figure 6.1 constitutes a representation of the *finite-state automaton* (or *finite-state machine*) corresponding to the grammar above; it is the formal device to recognise strings that are generated by the rules above. Hence the rules are indirectly represented in this graph and it also

[1] Here we use the convention in the artificial grammar learning literature in experimental psychology, with numbers referring to non-terminals and capital letters as terminals; another convention is to use capital letters for non-terminals and lower-case letters for terminals. The formal definition of a formal grammar is the 4-tuple $G=(N, \Sigma, P, S)$, for which N represents the set of non-terminal symbols, Σ the set of terminal symbols (surface symbols), P the set of generative production rules and S the starting symbol (from N).

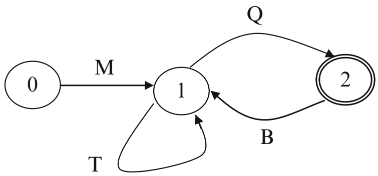


Fig. 6.1 The finite-state automaton that corresponds to the regular grammar defined above. The states are represented as circles and transitions between states as arrows. The starting state is [0]. The final (accepting) state [2] is denoted by a double-circle

expresses all production sequences that the rules can possibly generate by different paths. Because regular grammars as well as finite-state automata provide a formally equivalent characterisation of regular languages, finite-state automata are frequently used as representations of the underlying grammar and sequence structure in the artificial grammar learning literature (i.e. rewrite rules 1–5 are represented by the finite-state automaton; note that non-terminals map to states and terminals to arrows).

Regular languages (finite-state grammars) are defined by using rewrite rules where terminals on the right-hand side of the production are rewritten to only one side of a non-terminal (as in the above rules where terminals were always written to the left of the non-terminal, defining a right-linear regular grammar). As the example illustrates, regular languages may involve recursive rules, specifically *tail recursion*, i.e. a form of recursion that produces a sequence that recursively expands to one side of the sequence (left or right; note, however, that the combination of both right and left regular production sequences instantiates a grammar that is no longer regular). Such instances of tail-recursion correspond to cycles in the finite-state automaton.

A context-free grammar allows for rules rewriting nonterminals to any sequence of terminals and nonterminals on the right-hand side. This leads to a whole different form of expressive power including *centre-embedding recursion* which results in a sequence with self-embedding within a sequence. A famous, most simple example of this is the language A^nB^n (i.e. after any number of As, have the same number of Bs) created from the rules:

6. [S] → A [S] B
7. [S] → ε

(Or the rules [1] → A [2], [2] → [1] B, [1] → ε, if rules are restricted to binary productions)

Applying rule 6 once produces A [S] B, twice produces AA [S] BB and three times AAA [S] BBB, and then applying rule 7 produces the string AAA [S] BBB. Obviously rule 7 can be applied after any number of applications of rule 6 (note that this leads to a tree representation in which the symbols produced by each rule application are the children of the node corresponding to the left-hand side variable (in the present example [S]) that is rewritten by the rule; in the context of our example it results in a ternary tree in which each new subtree is embedded in between an A and B symbol). As rewriting occurs either side of the non-terminal, these rules cannot be represented by a finite-state diagram. For example, the following (recursive) finite-state diagrams (Fig. 6.2) do not produce all A^nB^n strings and also produce

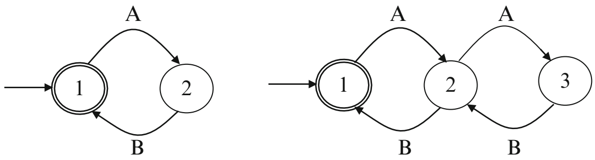


Fig. 6.2 Finite-state automata for different languages over {A,B}

strings which are not A^nB^n (see Fig. 6.3 for a better matching representation of A^nB^n for finite *n*).

Another very simple example of a recursive context-free language would be the *Dyck language*, which describes expressions of all correct nested bracketings (such as [ ], [[ ] [ ]] or [[[][[][]][]][[]]]), which is empirically relevant, e.g. with respect to grouping structure in music (Bod, 2002; Lerdahl & Jackendoff, 1983). It is important to note here that the A^nB^n or Dyck language are two of the *most simple* examples of context-free languages that can be constructed, which feature a structure of recursively nested brackets that is typical for context-free languages. However, they are not representative of the potential expressiveness and complexity of context-free languages (the grammar of a programming language or (simplified) English may be expressed with context-free languages).

Supraregular (context-free and higher) recursive grammars also produce various forms of mirror symmetries (Hopcroft, Motwani, & Ullman, 2000). While mirror symmetries exhibit properties that are easily modelled by variants of the A^nB^n grammar, translational symmetry such as $A_1A_2A_3B_1B_2B_3$ requires (mildly) context-sensitive grammars. Because of these interesting properties, sometimes debates restrict the use of "recursion" to languages above regular languages.

As the construction rules of such cross-serial dependencies or translational symmetries are rarely made explicit, we would like to show them:

8. $S \rightarrow a_i\ S\ T_i$
9. $S \rightarrow \varepsilon$
10. $a_i\ T_j \rightarrow a_i\ b_j$
11. $b_j\ T_i \rightarrow T_i\ b_j$

The grammar first constructs (a T) pairs and then employs the context-sensitive rules 10 and 11 to create b symbols from T and swap as and bs until they are in the right order. (Note that a generation by this grammar cannot described by a tree; further note again that the cross-serial dependency grammar is a very simple and not representative case of the expressive power underlying context-sensitive grammars.)

The Parsimony Argument

Structures generated by recursive grammars (be they regular, context-free or of higher complexity) may exhibit particular features that cannot be expressed by simpler models. The self-embedding/centre-embedding nature of context-free grammars entails nonlocal dependencies. While regular grammars can express some forms of potentially infinite long-distance dependencies (see Chomsky, 1956; Cleeremans,

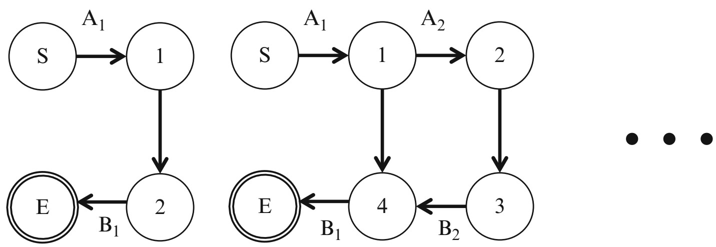


Fig. 6.3 Finite-state automaton approximations of the A^nB^n language for finite numbers of n. The displayed automata can be extended to any number n by adding 2n additional states

1993; Rohrmeier, Fu, & Dienes, 2012), the structures linking both ends of several dependencies have to be represented multiply—and nested nonlocal dependencies hence require a factorially growing form of representation. Particularly with respect to the latter, context-free grammars constitute a considerably simpler and more parsimonious form of representation (and model informing production/perception). We refer to this as the *parsimony argument*. Consider, for example, representing A^nB^n with a finite-state grammar. It can be done for a finite n. The grammar on the left works for $n=1$; the (non-recursive) grammar on the right works for $n=2$ (Fig. 6.3).

But for each increase in n, another set of states have to be added to the representation. If people were trained to detect A^nB^n for $n=1$, 2 and generalised to higher n's, it may be more parsimonious to postulate the representation of a context-free grammar that contains only two rules (or a corresponding processing mechanism) than of a finite-state machine where the subject constructed more, potentially a factorially growing number of, states or rules than needed to account for what they were exposed to. The regular grammar representation cannot make elegant use of recursion to express the repeating regularity underlying the structure. In addition, if such a catch-all finite-state model postulates learning states or chunks not required for learning the training material to predict participants' generalisation performance, the explanation is similarly less parsimonious. Altogether, that is, postulating the mental representation of a grammar or corresponding processing device (automaton) lower down the Chomsky hierarchy, just because one can, is not necessarily to give the simplest theory in particular when it comes to context-free structures in the case of ecological language acquisition that are far more complex than the A^nB^n language. Postulating a representation of higher Chomsky hierarchy complexity buys less description length and better representation for the price of what may be little more than the the use of a specific form of memory buffer (such as a stack or a queue).

Recursive context-free structures are frequently represented with hierarchical tree-based representations. These hierarchical representations involve recursion in a way such that a subtree may recur as a substructure at another level of the tree (even in the case of tail recursion, i.e. regular tail recursion constitutes a tree that is degenerated to a list). For instance, relative clauses in language and modulation in music are examples of such recursive structures consisting of linguistic or musical phrases embedded within phrases. However, hierarchical tree-based representation does not in itself imply that the structure is recursive. There are meaningful hierarchical representations of music (e.g. piece, section, phrase, motif) or, for instance, some kinds of phonological trees that are hierarchical but may not be recursive.

Recursion in Language and Music

There are many examples of linguistic and musical structures that exhibit tail recursion and centre-embedding recursion as their construction principles. For instance, the generative syntax model of tonal harmony (GSM; Rohrmeier, 2011) provides a concise specification which features of musical harmonic sequences exhibit recursion: applied secondary or diatonic dominants are instances of tail recursion, while centre-embedding recursion is created by modulation (remarkably, first pointed out by Hofstadter, 1979; cf. Giblin, 2008) and functional region expansion, both creating the potential for nested nonlocal dependencies (see Rohrmeier, 2011, for details). Further, musical grouping structure involves recursive nesting (Bod, 2002; Lerdahl & Jackendoff, 1983). Similarly it is easy to come up with two examples of tail-recursion and recursive centre-embedding in language (see 1a, 1b and 2 below). Comparably, Steedman (2002) and Jackendoff (2009, 2011) argued similar cases for embedded complex action or movement sequences.

1a. This sentence continues and continues and continues and continues.
1b. The green, red, yellow, purple, … balls are in the box.
2. Peter who gave Anne who helped Frank (who …) paint the fence the ball laughed.

Implications Regarding Mental Representation, Learning, Processing and Methodologies

Based on this theoretical background, the empirical investigation of recursion in learning and processing involves a number of methodological and philosophical issues. The core point revolves around the issue of finiteness: empirical research deals with finite materials, yet as outlined above, while recursive rules or mechanisms are a means for the description or production of (potentially) infinite sets of sequences, finite sets *potentially* can be acquired, processed and represented by lists of examples or chunks, simpler forms of grammars, etc. Hence though it is theoretically difficult to disprove simpler forms of representation, a number of methodological steps may assure to render alternative explanations much less compelling or plausible.

As remarked by Dulany, Carlson, and Dewey (1984), Reber (1993) and others, the structure according to which the materials are constructed need not be equivalent to the structures that are mentally represented. The reflection of this insight has shaped the prevalent notion in artificial grammar learning research that rarely the regular grammars themselves are assumed to be acquired or represented (as rules or finite-state automata; see, e.g. Pothos, 2007). With respect to more complex recursive structures, this issue is more intricate. First, the view that chunk or n-gram learning suffices to account for the structures in natural language is theoretically inadequate (see, e.g. Chomsky, 1956). Second, a positive finding of acquired recursive structures has implications for the underlying cognitive architecture: the mental representation may require not only a representation that is isomorphic to one dealing with the structure (such as a tree) but also an additional parsing process dealing with these structures (including potential mechanisms of revision and backtracking to solve ambiguities and garden path sequences). For instance, the implicit learning of the *Dyck language* (of which A^nB^n is the simplest subset and example) will require the instantiation of a parsing process that keeps track of each open bracket and "ticks off" the innermost open bracket (and potentially triggers another related action) once a closing bracket is encountered. A sequence is perceived as ungrammatical when one closing bracket on an empty stack or an open bracket at the end of the sequence is encountered. This example illustrates that this process does not necessarily involve the mental formation of a representation of a tree structure from the encountered sequence: merely the parsing process itself suffices to deal with the structures. It further illustrates that the involved parsing mechanism may but need not be recursive itself: the parsing process of the context-free, recursive bracketing structure merely involves memory (e.g. a stack) and an *iterative* (potentially nonrecursive or not explicitly recursive) process[2] that adds or ticks off brackets. Generally, processing (and dealing with) recursive structures processing (and dealing with) recursive structures is potentially achieved without recursive representations or recursive mechanisms. With respect to the learning process, it is an implausible assumption that the parsing mechanism is acquired together with its examples. But it may be the case that an established cognitive mechanism to deal with embedded structures is triggered as the most efficient to deal with the perceived sequences.

Furthermore, a separation of training sequences and testing sequences such that testing sequences are novel and force generalisation beyond exemplar and chunk memorisation (cf. Rohrmeier & Cross, 2010; Rohrmeier, Rebuschat, & Cross, 2011) warrants that a more general and abstract representation is acquired. A second issue concerns the examined structure from which materials are taken. Some grammars that may be expressed by formal rules like context-free rules may not exhibit recursion (such as Saffran, 2001, 2002, and the *Brocanto* language, as e.g. in Opitz & Friederici, 2003).

There are also two issues concerning the interpretation of the findings and their generalisation: Does the finding of learning of the well-studied A^nB^n language indeed warrant the conclusion that *generally* (all or most) context-free grammars can be learned (implicitly)? Whether such a finding expands to general cases of

[2] Note that the distinction between recursive and iterative algorithms involves a notion of recursion that avoids encompassing all kinds of iteration (cf. Tomalin, 2007).

context-free languages (closer to ecological sequences) or languages of the complexity of formal languages like PASCAL or LISP remains a barely addressed issue. Second, despite its theoretical importance, the Chomsky hierarchy (Chomsky, 1959; Chomsky & Schützenberger, 1963; Jäger & Rogers, 2012) does not constitute the only way to carve out the infinite space of sequential structures. It may well be that other distinctions deriving from processing issues (the types of sequences that can be processed by a sequential neural network, like the SRN, or a graphical model [like an HMM]) may provide implementation-motivated ways (cf. Marr, 1982) to distinguish between types of structures based on the way in which brain mechanisms are able to deal with them. These distinctions may not coincide with the Chomsky hierarchy.[3]

Implications for Mental Representation or Processing of Recursive Structures

As argued above, mechanisms of processing or parsing of recursive structures may not require (but could employ) recursion itself. In fact, the ability to deal with recursive structures may not require a representation of the recursive rules or the full recursive structure: e.g. instead of a full tree, a partial mental representation may potentially suffice, for instance, for music, without having the full structure ever represented at once (see also Rodriguez, Wiles, & Elman, 1999, for an account of how an SRN model could learn A^nB^n structure). However, when materials are designed in ways such that chunking or similar accounts are insufficient (see, for instance, the modelling of Dienes & Longuet-Higgins, 2004; in Rohrmeier, 2010; or Koelsch, Rohrmeier, Jentschke, & Torrecuso, 2013), findings reveal that at least one of the potential mechanisms to deal with non-regular structures is at work. While finite-state machines offer ways to deal with finite numbers of embeddings, the parsimony argument may act in favour of the learning/representation of non-finite-state machines: mechanisms isomorphic with recursive representation or processing are less complex than all-encompassing catch-all finite-state machines. In this context, computer simulations with exemplar or chunking learning mechanisms provide the most powerful way to establish that certain structures cannot be acquired by a plausible chunk learning process. Further it is important to note that the frequent counter argument, that the simplicity of recursive rules is traded for remarkable complexity involving recursion rather than no recursion in the processor, does not hold here since processing mechanisms powerful enough to deal with recursion (e.g. empirically occurring levels of centre embedding) may not involve recursion themselves. Finally, however, such accounts will have to be based on Bayesian

[3] Note also that the distinction between context-free and context-sensitive structures is, in practice, separated by worlds of complexity which are not adequately reflected by the simple distinction of centre-embedded and (mildly context-sensitive) cross-serial structures, e.g. $A_1A_2A_3B_3B_2B_1$ and $A_1A_2A_3B_1B_2B_3$.

model comparison formally taking into account the model complexity involved (compare, e.g. Perfors, Tenenbaum & Regier, 2011). The *parsimony argument*, however, suggests that recursive models can be favoured despite the apparent "simplicity" of using processors lower down the Chomsky hierarchy.

Evidence from Empirical Research

AGL in Language and Music

When Reber (1967) coined the term "implicit learning" by looking at how people learn artificial grammars (following George Miller explorations with artificial grammar learning; Miller, 1967), he decided to start at the bottom of the Chomsky hierarchy, with regular grammars. The choice was a sensible starting point. Regular grammars are still sufficiently complex that people do not consciously notice all of the structure, while providing structure that can be unconsciously learned. Regular grammars have provided decades of work on clarifying some simple things people can learn (see Pothos, 2007, for a review). For example, when exposed to a regular grammar, people can learn chunks of especially two or three successive elements (sometimes four) (e.g. Perruchet & Pacteau, 1991; Rohrmeier et al., 2011; Rohrmeier & Cross, 2010, 2013; Servan-Schreiber & Anderson, 1990), people can learn specific training exemplars and use these as analogies (e.g. Brooks, 1978; Jamieson & Mewhort, 2009) and people can learn specific repetition structures (Brooks & Vokey, 1991). A repetition structure is a way of learning a long-distance relationship through memorisation (and thus is important to control in work on learning non-regular grammars). The strings of letters MTTVT and ABBCB have the same repetition structure, which could be represented as 12232, meaning the first letter is unique, the second letter is different from the first, but the third letter is the same as the second, the fourth letter is unique and the final letter is the same as the second. Chunks, exemplars and repetition patterns can be learned implicitly (e.g. Scott & Dienes, 2008). While these structures fall short of learning all structure in any regular grammar, the question is still left open of just what structure in non-regular grammars can be implicitly learned. Given the apparent qualitative difference between implicit and explicit learning (Dienes, 2012), our main interest is in what structures can be implicitly learned (note that explicit learning of context-free grammars such as a programming language comes as no surprise).

Learning of Context-Free Grammars in Language and Music

As outlined above, natural languages and musical structures exhibit some structural principles that require (at least) context-free grammars to be expressed. There are a number of studies that explore learning of A^nB^n and related structures (e.g. Bahlmann, Schubotz, & Friederici, 2008; De Vries, Monaghan, Knecht, & Zwitserlood, 2008; Fitch & Hauser, 2004; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Hochmann, Azadpour, & Mehler, 2008; Lai & Poletiek, 2010; Perruchet & Rey, 2005; Poletiek, 2011; Uddén, Ingvar, Hagoort, & Petersson, 2012). There have been several reviews discussing these studies in the context of recursive processing and learning (e.g. Fitch & Friederici, 2012; Rohrmeier et al., 2012). Accordingly here we focus on recent work that goes beyond the scope of these studies and explored implicit learning of more complex context-free grammars that were modelled to reflect ecological features of music and language more closely.

Rohrmeier et al. (2012) constructed materials that modelled abstract word order in linguistic relative clauses, such as “the man, who met the woman, sold the chair”. To construct a simplified model language, the artificial grammar accordingly features three categories *N*, *V and R* representing categories nouns, transitive or intransitive verbs and a relative pronoun. Each of the categories *N* and *V* featured four possible words, whereas the *R* class consisted only of one word. For modelling the embedded structure, two core distinctions were taken into account: the embedding structure could be either left or right branching, and structures could be either centre embedded or tail recursive. All four possible combinations were explored each with one group in the study. Starting from the main clause “N V” or “N V N” (modelling a simple sequence such as “boy kisses girl”), relative clauses of the form “R N V” or “V N R” (centre embedding and right/left branching) or “R V N” or “N V R” (tail embedding and right/left branching) could be attached to each noun with up to two levels of embedding. Further relative clauses without noun classes were used: “R V” or “V R” (right and left branching). Example sequences are “NV(VNR)N” and “N(RN(RNV)V)VN” for left-branching and right-branching embedding of the first and second order. Stimuli were rendered as spoken sequences. The experiment used four experimental and control groups which all first completed a learning phase being exposed to 168 examples from the grammar featuring up 0 up to 2 levels of embedding (the control group was trained on random sequences). The experimental phase employed the Process Dissociation Procedure (Jacoby, 1991) presenting pairs of grammatical stimuli that were presented in the training (old grammatical) or only in testing (new grammatical) and ungrammatical stimuli that featured either systematic violations of one embedded subsequence or a random sequence. Finally participants completed a category test in which participants had to choose two out of three words that would belong to the same category.

Results showed that people could classify both new and old grammatical structures above chance with little difference between them. A comparison between responses to random and layer-violating ungrammatical sequences further indicated that participants had acquired detailed knowledge about subtle difference in the grammatical structure. The category test revealed that participants acquired implicit knowledge of word classes.

In greater detail, analyses revealed that participants were performing better for sequences with violations in the first or second layer than the third layer, suggesting that higher-order embeddings turned out more demanding. Finally participants outperformed controls on local as well as nonlocal violations. A subsequent logistic regression analysis was carried out to examine the extent to which responses could be explained by local n-gram learning or repetition structure. It was found that nonlocal dependencies formed the greatest predictor of responses although n-grams and local repetition structure had a small yet significant impact. Crucially, subjects were sensitive to long-distance relations above and beyond their sensitivity to n-grams or repetition patterns, at the level of either terminals or non-terminals. Further, people were sensitive to the long-distance dependencies after partially out sensitivity to fixed distance dependencies, consistent with the knowledge of long-distance dependencies being structure sensitive. Accordingly the findings show evidence for implicit learning of complex (nonlocal and local) dependencies as constructed by a recursive context-free grammar. Finally, the results suggested that tail-recursive structures were better learned than centre-embedding structures and that left branching was better learned than right branching when the grammar was tail embedding. This performance difference correlates with Hawkins's (2004) performance and correspondence hypothesis for natural languages—supporting the claim that the more complex type of grammar may impede performance and therefore be less frequently established across language varieties.

To investigate implicit learning of musical harmonic structure (which is recursive and context-free; see Rohrmeier, 2011, and Steedman, 1996), Rohrmeier and Cross (2009) as well as Rohrmeier (2010, ch. 4) used the same experimental and structural design as Rohrmeier, Fu and Dienes, transferring the entire setting to the musical domain. Instead of three word classes, three chord classes were constructed using octatonic scales, two of which contained four chords and one contained a single dissonant chord. Octatonic scales were used to make it possible to employ familiar categorisable materials (major and minor chords) in a novel, unfamiliar context (not a major or minor scale). Since musical structure is organised predominantly with left branching (i.e. supporting *goal-directed* sequences), only tail- and centre-embedding left-branching structures were employed. In analogy to the linguistic findings, results indicated that Western musicians acquired generalised implicit knowledge of the grammatical sequences, in particular local as well as nonlocal structures. According to the logistic regression analysis, *n*-gram learning was only a weak to non-existent predictor (exp. 2), whereas grammatical structure turned out to be the strongest predictor. Again, centre-embedding materials were less well learned than the sequences featuring tail embedding. This suggests that there may be a domain-general processing/performance advantage for simpler tail-recursive sequences to centre-embedding structures in line with Hawkins's (2004) theory.

In further experiments, the same musical paradigm was employed to test implicit learning in Chinese musician and nonmusician subjects. The results showed that all groups acquired both grammars above chance; however, Western musicians outperformed Chinese musicians significantly. Further the difference between Chinese musicians and nonmusicians was comparably small (only with respect to random and layer 1 violations). One potential reason why the structures were harder for Chinese participants may be an effect of acquired implicit musical knowledge (cf. Rohrmeier & Cross, 2013), given that traditional Chinese music features a stronger emphasis on melodic structure and pentatonic scales (even in pop music) and less on harmonic structure.

Learning Context-Sensitive Grammars in Music, Poetry and Movement

Symmetry is a structure that can be generated recursively and requires at least a context-free grammar to parse. Symmetry occurs not only in language (e.g. centre embedding of the form $A_1A_2A_3B_3B_2B_1$ and cross-serial dependencies of the form $A_1A_2A_3B_1B_2B_3$) but also in art, including music. For example, serialist (twelve-tone) music makes use of the symmetries of inversion (produced by changing upward to downward motion and vice-versa) and retrograde (produced by placing a mirror vertically beside a line of music score, i.e. generating a reversed "retrograde" version of a melodic line). Dienes and Longuet-Higgins (2004) provided preliminary evidence that experts in serialist music could implicitly learn to become sensitive to a particular musical symmetry [inverse, retrograde, inverse retrograde and transpose (i.e. copy)] after brief exposure to it. Kuhn and Dienes (2005) simplified the materials and showed that musically unselected participants after exposure to musical inversions, increased their liking of inversions compared to non-inversions compared to an untrained control group. Further, people were unable to classify which strings were rule following and which were not, demonstrating the implicit nature of the knowledge that expressed itself in liking. Kuhn and Dienes controlled chunk strength in the materials to rule out this simple explanation. However, for the actual materials used, which were of fixed length, Kuhn and Dienes (2008) showed an SRN could learn a set of long-distance associations between a tone in one position and a tone a fixed distance later. That is, subjects need not have learned the symmetry per se in order to show learning on the test set.

The music inversion results have been further explored using Chinese Tang Poetry. Chinese words are spoken in one of four tones, which are traditionally categorised into two groups, ping and ze. Treating ping and ze as opposites, Jiang et al. (2012) constructed artificial poetry in which successive lines of each poem were ping-ze inverses of each other and hence constitute instantiations of cross-serial dependencies. Jiang et al. strictly controlled both chunk strength and repetition patterns (Vokey & Brooks, 1994). People asked to memorise such poems later classified new poems (of the same line length) as better formed when they showed the inverse pattern. Further, on a trial-by-trial basis, they almost exclusively claimed to classify based on pure guessing or on intuitive feelings they could not explicate further. That is, the learning was implicit. Li et al. (2013) showed that poems expressing a retrograde symmetry could also be classified as well as formed following training on retrogrades (once again controlling chunk strength and repetition pattern). Retrogrades can be parsed using a mechanism that can deal with a context-free grammar, but inverses require something more than context free. So which is easier—retrogrades or inverses? Li et al. showed that implicitly learning retrogrades was harder than implicitly learning inverses. That is, grammatical complexity as expressed by Chomsky hierarchy predicts the order of difficulty the opposite way (given it is actually retrogrades and inverses that subjects are learning in the poetry experiments). Both inverses and retrogrades require a buffer memory mechanism to be parsed; inverses could be most simply parsed using a first-in-first-out buffer (a queue) and retrogrades using a first-in-last-out buffer (a stack). Thus, the results are consistent with implicit learning using a memory functionally like a first-in-first-out buffer for non-regular grammars. Jiang (2012) showed an SRN could also learn the poetry inversion materials. She also showed that people only learned the inversion when ping and ze categories were used; when a linguistically arbitrary binary category was used for the tones, people did not learn (thus people were learning a relation over tone classes and not tones). Similarly the SRN only learns when given the ping and ze categories as input. Guo et al. (in preparation) also showed the SRN could learn the retrograde materials but less well than the inversion materials, just like people. The match between the characteristic behaviour of people and SRN is remarkable. But has either actually learned the symmetry per se?

Guo et al. performed a crucial experiment looking at transfer to different lengths. If people have simply learned a long-distance association, there is no reason for the knowledge to transfer to different length test items. However, when people were trained on inversions where the length of each line was five words, they could generalise to poems of both four words and six words. There was a notable performance decrement from trained to untrained lengths. Thus, the simplest explanation may be that people form a graded fixed length association (for the ability of people to learn fixed length associations, see Remillard, 2008). That is, it is not yet clear people could treat length itself as a variable. If when trained on different lengths people could generalise to yet new (longer) lengths with facility, it would motivate the claim that people could learn to map a vector of unspecified length to its negation, i.e. learned the inverse rule per se. In the meantime, the simplest explanation is that people learned fixed long-distance associations in a graded way. However, people may also have a partial grasp of a relevant context-sensitive grammar (though presumably they have not explicitly or symbolically represented the context-sensitive rules we gave above, but might instantiate a process and memory representation to capture such cross-serial dependencies)—people's natural, conscious, yet automatic, appreciation of symmetry in nature, art, faces and scientific theories (Dienes, Kuhn, Guo, & Jones, 2011; Westphal-Fitch, Huber, Gómez, & Fitch, 2012) motivates our conjecture that people can detect mirror symmetries implicitly as well as explicitly. This is an area for further exploration. Guo et al. (in preparation) also found that the SRN characteristically learns to generalise from length five to lengths four and six for these materials, in a similar way as people. As for people, there was a notable decrement in performance from the trained length to the neighbouring lengths. Thus, so far it remains fair to characterise the SRN as a graded finite-state machine for our materials (as Cleeremans (1993) did). But we seek to test this claim and see if the SRN will become a graded recursive structure processor (when trained on different lengths) (compare Rodriguez et al. (1999) who found the SRN could learn A^nB^n, which gives us optimism).

We have implemented the same inversion structure as used by Kuhn and Dienes (2005) in movement sequences (see Dienes et al., 2011). Kuhn and Dienes used the scale of C major for all their "melodies". Thus, each melody could be construed as a movement around a circle (clock arithmetic with 8 points on the clock face). Dienes et al. (2011) asked subjects to trace a circle with their finger, so in effect the subjects traced out inversions isomorphic to musical ones. The materials were controlled for chunk strength and repetition structure. Subjects came to like the inversions more than non-inversions to a greater extent than untrained controls, even when just considering subjects who claimed no awareness of the symmetry at the end of the experiment. We hope that the movement paradigm will be one that other researchers can take up (more easily than music or Chinese tonal paradigms) to explore in more detail how the implicit system comes to parse recursively generated structures, an under-researched topic in implicit learning.

Conclusion

Although implicit learning is a well-established, long-standing research tradition, there is a gap between the types of structures used implicit learning studies and the ecological complexity and type of structure (such as in music and language). In particular, language and music exhibit a variety of recursive features which cannot be modelled by regular languages, but require context-free or (mildly) context-sensitive structures. We discuss empirical issues concerning the exploration and the conclusion drawn from the study of implicit learning of recursive structures. Although finite cases of recursive structures could (theoretically) always be explained by catch-all exemplar-based or finite-state representations, such accounts are challenged by their inefficient representation (the *parsimony argument*) facing recent findings concerning learning nonlocal dependencies and the generalisation of acquired knowledge. We report empirical studies which suggest that musical and linguistic recursive context-free structures can be acquired implicitly in ways which are not explained by n-gram, repetition structure or fixed length association learning. Another series of studies suggests that simple context-sensitive cross-serial dependencies are acquired implicitly, above and beyond the learning of n-grams and repetition structures (but the case for the knowledge extending beyond fixed long-distance associations is not yet made). Taken altogether there is a growing body of evidence that supraregular structures can be acquired implicitly—findings which considerably extend the present knowledge about the limits of implicit learning and challenge the differences in complexity predicted by the Chomsky hierarchy.

References

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical sequencing engages Broca's area. *NeuroImage, 42*, 525–534.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*, 100–130.

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research, 17*, 289–308.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. *Cognition and Categorization, 3*, 170–211.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology. General, 120*(3), 316–323.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, IT-2*, 113–124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control, 2*(2), 137–167.

Chomsky, N. (1995). *The minimalist program* (Vol. 28). Cambridge, MA: MIT press.

Chomsky, N., & Schützenberger, M. P. (1963). The algebraic theory of context free languages. In P. Braffort & D. Hirschberg (Eds.), *Computer programming and formal languages* (pp. 118–161). Amsterdam: North Holland.

Cleeremans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Cleeremans, A. (2006). Conscious and unconscious cognition: A graded, dynamic, perspective. In Q. Jing, M. R. Rosenzweig, G. D'Ydewalle, H. Zhang, H.-C. Chen, & K. Zhang (Eds.), *Progress in psychological science around the world* (Neural, Cognitive and Developmental Issues, Vol. 1, pp. 401–418). Hove: Psychology Press.

De Vries, M., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition, 107*, 763–774.

Deliège, I., Mélen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real time listening to a piece of music. *Music Perception, 14*(2), 117–160.

Dienes, Z., Kuhn, G., Guo, X. Y., & Jones, C. (2011). Communicating structure, affect and movement: Commentary on Bharucha, Curtis & Paroo. In P. Rebuschat, M. Rohrmeier, I. Cross, & J. Hawkins (Eds.), *Language and music as cognitive systems* (pp. 156–168). Oxford: Oxford University Press.

Dienes, Z. (2012). Conscious versus unconscious learning of structure. In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 337–364). The Hague: Mouton de Gruyter Publishers.

Dienes, Z., & Longuet-Higgins, H. C. (2004). Can musical transformations be implicitly learned? *Cognitive Science, 28*, 531–558.

Dulany, D., Carlson, R., & Dewey, G. (1984). A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology, 113*, 541–555.

Fitch, W. T., & Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377–380.

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences, 367*(1598), 1933–1955. doi:10.1098/rstb.2012.0103.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*(7), 2458–2463.

Giblin, I. (2008). *Music and the generative enterprise. Situating a generative theory of tonal music in the cognitive sciences*. University of New South Wales

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hochmann, J. R., Azadpour, M., & Mehler, J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science, 32*, 1021–1036.

Hofstadter, D. R. (1979). *Godel, Escher, Bach: An eternal golden braid*. London: Penguin.

Honing, H. (2011). *The illiterate listener. On music cognition, musicality and methodology*. Amsterdam: Amsterdam University Press.

Hopcroft, J. E., Motwani, R., & Ullman, J. D. (2000). *Introduction to automata theory, languages and computation* (2nd ed.). Reading, MA: Addison-Wesley.

Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language, 87*(3), 586–624.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception, 26*(3), 195–204. doi:10.1525/mp.2009.26.3.195.

Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and What's special about it? *Cognition, 100*, 33–72.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jäger, G., & Rogers, J. (2012). Formal language theory: Refining the Chomsky hierarchy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1598), 1956–1970.

Jamieson, R. K., & Mewhort, D. J. K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *The Quarterly Journal of Experimental Psychology, 62*, 550–575.

Jiang, S. (2012). *Implicit learning of Chinese tonal regularity* (Unpublished dissertation). East China Normal University, China.

Jiang, S., Zhu, L., Guo, X., Ma, W., Yang, Z., & Dienes, Z. (2012). Unconscious structural knowledge of tonal symmetry: Tang poetry redefines limits of implicit learning. *Consciousness & Cognition, 21*, 476–486.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520–541.

Koelsch, S., Rohrmeier, M., Torrecuso, R., & Jentschke, S. (2013). *Processing of hierarchical syntactic structure in music*. Proceedings of the National Academy of Sciences. Published ahead of print September 3, 2013, doi:10.1073/pnas.1300272110.

Kuhn, G., & Dienes, Z. (2008). Learning non-local dependencies. *Cognition, 106*, 184–206.

Kuhn, G., & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 31*, 1417–1432.

Lai, J., & Poletiek, F. H. (2010). The impact of adjacent-dependencies and staged-input on the learnability of center-embedded hierarchical structures. *Cognition, 118*, 265–273.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Li, F., Jiang, S., Guo, X., Yang, Z., & Dienes, Z. (2013). The nature of the memory buffer in implicit learning: Learning Chinese tonal symmetries. *Consciousness & Cognition, 22*(3), 920–930.

Lobina, D. J. (2011). "A running back" and forth: A review of recursion and human language. *Biolinguistics, 5*(1–2), 151–169.

Loui, P. (2011). Statistical learning – What can music tell us? In P. Rebuschat & J. Williams (Eds.), *Statistical learning and language acquisition* (pp. 433–462). The Hague: Mouton de Gruyter.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Martins, M. D. (2012). Distinctive signatures of recursion. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1598), 2055–2064.

Masters, R. S. W., Maxwell, J. P., & Eves, F. F. (2009). Marginally perceptible outcome feedback, motor learning and implicit processes. *Consciousness and Cognition, 18*, 639–645.

Miller, G. A. (1967). *The psychology of communication*. New York: Basic Books.

Müller, S. (2010). *Grammatiktheorie*. Tübingen: Stauffenburg.

Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage, 19*, 1730–1737.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*(3), 306–338.

Perruchet, P., & Pacteau, C. (1991). Implicit acquisition of abstract knowledge about artificial grammar: Some methodological and conceptual issues. *Journal of Experimental Psychology. General, 120*(1), 112–116.

Perruchet, P., & Rey, A. (2005). Does the mastery of hierarchical structures distinguishes humans from non-human primates? *Psychonomic Bulletin & Review, 12*(2), 307–313.

Poletiek, F. H. (2011). What in the world makes recursion so easy to learn? A statistical account of the staged input effect on learning a centre-embedded hierarchical structure in AGL. *Biolinguistics, 5*(1–2), 36–42.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin, 133*, 227–244.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology. General, 118*, 219–235.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Rebuschat, P. (2008). *Implicit learning of natural language syntax* (PhD Dissertation). Research Centre for English and Applied Linguistics, University of Cambridge, Cambridge.

Rebuschat, P., & Williams, J. N. (Eds.). (2011). *Statistical learning and language acquisition* (Vol. 1). New York: De Gruyter Mouton.

Reed, N., McLeod, P., & Dienes, Z. (2010). Implicit knowledge and motor skill: What people who know how to catch don't know. *Consciousness & Cognition, 19*, 63–76.

Remillard, G. (2008). Implicit learning of second-, third-, and fourth-order adjacent and nonadjacent sequential dependencies. *The Quarterly Journal of Experimental Psychology, 61*, 400–424.

Rodriguez, P., Wiles, J., & Elman, J. (1999). A recurrent neural network that learns to count. *Connection Science, 11*(1), 5–40. doi:10.1080/095400999116340.

Rohrmeier, M. (2007). A generative grammar approach to diatonic harmonic structure. In G. Spyridis, & A. Kouroupetroglou (Eds.). *Proceedings of the 4th Sound and Music Computing Conference* (pp. 97–100).

Rohrmeier, M. (2010). *Implicit learning of musical structure. Experimental and computational modelling approaches* (Ph.D Dissertation). Centre for Music & Science, Faculty of Music, University of Cambridge, Cambrdige, UK.

Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music, 5*(1), 35–53.

Rohrmeier, M., & Cross, I. (2009). Tacit tonality: Implicit learning of context-free harmonic structure. In J. Louhivuori et al. (Eds.). *Proceedings of the 7th Triennial Conference of the European Society for the Cognitive Sciences of Music* (*ESCOM 2009*) (pp. 443–452).

Rohrmeier, M., & Cross, I. (2010). Narmour's principles affect implicit learning of melody. In Demorest, et al. (eds.). *Proceedings of the 11th International Conference on Music Perception and Cognition (ICMPC 2010)*.

Rohrmeier, M., & Cross, I. (2013). Artificial grammar learning of melody is constrained by melodic inconsistency. Narmour's principles affect melodic learning. *Plos ONE 8*(7), e66174.

Rohrmeier, M., Dienes, Z., Guo, X., & Fu, Q. (2014). Implicit learning and recursion. In F. Lowenthal & L. Lefebvre (Eds.), *Language and recursion*. New York: Springer.

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive, hierarchical grammatical structures. *PLoS ONE, 7*(10), e45885. doi:10.1371/journal.pone.0045885.

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science, 4*(4), 525–53. doi:10.1111/j.1756-8765.2012.01223.

Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness & Cognition, 20*(2), 214–222.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493–515.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47*, 172–196.

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 1264–1288.

Servan-Schreiber, D., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16*, 592–608.

Shanks, D. R. (2005). Implicit learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 202–220). London: Sage.

Steedman, M. (2002). Plans, affordances, and combinatory grammar. *Linguistics and Philosophy, 25*, 723–753.

Steedman, M. (1996). The Blues and the abstract truth: Music and mental models. In J. Oakhill & A. Garnham (Eds.), *Mental models in cognitive science: Essays in honour of Phil Johnson-Laird* (pp. 305–318). Hove: Psychology Press.

Steedman, M. J. (1984). A generative grammar for Jazz Chord sequences. *Music Perception, 2*(1), 52–77.

Tomalin, M. (2007). Reconsidering recursion in syntactic theory. *Lingua, 117*, 1784–1800.

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2012). Implicit acquisition of grammars with crossed and nested non-adjacent dependencies: Investigating the push-down stack model. *Cognitive Science, 36*, 1078–101. doi:10.1111/j.1551-6709.2012.01235.x.

Vokey, J. R., & Brooks, L. R. (1994). Fragmentary knowledge and the processing-specific control of structural sensitivity. *Journal of Experimental Psychology, 20*, 1504–1510.

Westphal-Fitch, G., Huber, L., Gómez, J. C., & Fitch, W. T. (2012). Production and perception rules underlying visual patterns: Effects of symmetry and hierarchy. *Philosophical Transactions of the Royal Society B, 367*, 2007–2022.

Williams, J. N. (2009). Implicit learning. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (pp. 319–353). Bingley, UK: Emerald Group Publishing Limited.

Part III
Emergence of Grammar in Human and Non-human Animal Communication

Chapter 7
Early Rule-Learning Ability and Language Acquisition

Judit Gervain

Rule Learning: The Basic Framework

Whether human language is a rule-based mental knowledge system has been debated since Chomsky's original proposal more than half a century ago (Chomsky, 1959; Elman et al., 1997; Marcus, Vijayan, Rao, & Vishton, 1999; Parisse, 2005; Pinker, 1991; Tomasello, 2000). While several theories concur that adult native speakers possess a rule-based representation of their native language's grammar, it is heatedly debated whether this rule system is in place already in the initial state, i.e., innate, or whether it emerges during development, e.g., through some abstraction-processed extracting regularities from initially item-based knowledge (Tomasello, 2000 and related work). Consequently, research for evidence demonstrating the existence of rule-based, abstract knowledge early in development has been intense. The purpose of this chapter is to present a series of recent experiments that provide some of the earliest evidence in favor of the existence of such rule-based knowledge.

These studies follow the rule-learning paradigm introduced by Marcus et al. (1999), examining infants' ability to learn abstract structural regularities based on the identity relation. In the original study, Marcus et al. (1999) familiarized 7-month-old infants with artificial grammars generating three-word sequences in which two words were identical, e.g., ABB: "wo fe fe," ABA: "wo fe wo," or AAB: "wo wo fe." In one experiment, half of the infants were familiarized with different sentences generated by the ABB grammar, the other half with sentences from the ABA grammar for 2 min. In the subsequent test phase, both groups were presented with novel, previously unheard sentences from both grammars, and their looking times to the two types of sentences were measured using the headturn preference

J. Gervain (✉)
Laboratoire Psychologie de la Perception, CNRS and Université Paris Descartes, Sorbonne Paris Cité, France
e-mail: judit.gervain@parisdescartes.fr

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_7,

procedure, a classical method in infant speech perception and language acquisition research (Kemler Nelson et al., 1995). Infants looked longer to the sentences that were inconsistent with their grammar of familiarization, i.e., infants familiarized to the ABB grammar looked longer to ABA sequences and vice versa. This indicates that they have learned the structure of the grammar to which they were familiarized and could generalize it to novel instances. This latter aspect of the study is crucial: the test items were all new, made up of syllables that never occurred during familiarization, so infants could not simply rely on transition probabilities or rote memory to recognize the familiarization grammar. They necessarily generalized the abstract structure, which is the only common feature of the familiarized items and the test items. To further confirm this conclusion, in a second study Marcus et al. (1999) used the same procedure to test infants' ability to discriminate two grammars in which the repetition was always immediate, i.e., AAB vs. ABB. They found the same discrimination behavior, infants looking longer to the test items that were inconsistent with their familiarization grammar.

These seminal findings, showing rule learning at 7 months of age, gave rise to an interesting debate concerning the exact nature of this mechanism. One line of research (e.g., Frank, Slemmer, Marcus, & Johnson, 2009; Johnson et al., 2009; Marcus, Fernandes, & Johnson, 2007; Rabagliati, Senghas, Johnson, & Marcus, 2012; Wagner, Fox, Tager-Flusberg, & Nelson, 2011) pursuing the original claim that this is an abstract, symbolic learning mechanism has further explored rule learning, especially its specificity to speech and language. Marcus et al. (2007) reported that 7-month-old infants were able to perform the ABB vs. ABA discrimination on sequences implemented as animal sounds, musical tones, or timbre only if they were first trained on speech sequences, i.e., only then were they able to transfer the extracted structure. The authors interpreted this result as suggesting that language, i.e., a cognitive domain with symbolic structure, is the privileged input for rule learning. Subsequently, Frank et al. (2009) showed that infants as young as 5 months of age were able to learn the ABB and ABA sequences when they were presented multimodally, i.e., as a combination of visual and speech stimuli, but not when only one modality was used. Testing the visual modality alone, Johnson et al. (2009) found that 8- and 11-month-old infants were unable to learn the ABA pattern when shown sequentially presented visual stimuli but succeeded with ABB and AAB structures, suggesting that there may be an asymmetry between structures with adjacent and nonadjacent repetitions. However, this might depend on task-specific factors such as the mode of presentation. In fact, Saffran, Pollak, Seibel, and Shkolnik (2007) found that when the visual stimuli used were presented simultaneously and were natural (breeds of dogs) as opposed to geometric figures, even 7-month-olds succeeded (at least in the usual adjacent vs. nonadjacent repetitions comparison). More recently, Rabagliati et al. (2012) tested rule learning with communicative visual stimuli, namely, hand gestures similar to sign language. They found that 7.5-month-old infants could learn the ABB, but not the AAB rule with signs, i.e., their performance with these stimuli was less robust than for speech. The authors remain agnostic as to whether this is due to a hardwired preference for the auditory modality or to several months of experience with spoken language (and none with sign), but argue that their findings do show that it is not the communicative nature of speech that induces better performance with this stimulus type than with any other nonlinguistic auditory and visual stimulus that had been tested, since sign language gestures are also communicative.

Another line of research (for a review, see Endress, Nespor, & Mehler, 2009) has raised concerns about whether what infants learn in these studies is really abstract and symbolic in nature or whether lower-level, perceptually based mechanisms might be sufficient to explain the findings. In particular, it has been suggested that two properties of the stimuli used, adjacent repetitions and their edge (initial/final) position, might represent privileged configurations that our sensory system rapidly and automatically detects. Indeed, Endress, Scholl, and Mehler (2005) have found that adults were good at detecting repetition at the edges of 7-syllable-long sequence (e.g., ABCDEFF), but performed poorly when repetitions were inside these sequences (e.g., ABCDDEF). The privileged status of sequence onsets and ends is well known in the sequence learning and memory literature (Ng & Maybery, 2002): material in these two edge positions is remembered better than material in sequence middles (known as the primacy and the recency effects, respectively). Repetition, identity, or "sameness" itself also seems to be a special configuration. Endress, Dehaene-Lambertz, and Mehler (2007) found that adult learners were much better at learning the identity relation (ABB) between tones in a 3-tone sequence than the higher than/lower than relation (e.g., ABC, where tone A is always higher than tone B). For a purely abstract, symbolic learner, the two relations, i.e., identity and greater than/less than, should be equivalent, as formally they are both two-place relations between categories of items. Yet adults are much better at learning the former, at least when implemented in tone sequences, implying that repetition/identity is a special perceptual primitive, which even nonlinguistic animals such as honeybees are sensitive to (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001). On the basis of these results, Endress et al. (2009) have proposed that learning structures based on immediate repetitions in edge positions (e.g., ABB or AAB) might not constitute a real case of abstract rule learning. Rather, simpler perceptual shortcuts might do the job.

The exact nature of the mechanisms underlying the detection and learning of identity-based regularities is, therefore, still not fully understood. One particular aspect of the previous findings that has not been addressed in any systematic way in the literature is the potential asymmetry between the processing of adjacent and nonadjacent repetitions. Some studies in the symbolic rule-learning framework have found better performance in infants for adjacent than for nonadjacent repetitions, although only under certain conditions, suggesting that the asymmetry might be the result of factors external to the rule-learning system proper, e.g., cognitive load and task demands. The perceptual primitive account predicts a clear and systematic asymmetry, since adjacent repetitions constitute a well-attested perceptual primitive found even in nonlinguistic animals, but there is no evidence of nonadjacent repetitions being a perceptual primitive in the same sense. Despite its theoretical importance, the question of whether adjacent and nonadjacent repetitions are processed by the same mechanisms has not been systematically addressed.

The developmental trajectory of the processing of these two patterns has not been explored either. The aim of this chapter, therefore, is to offer a synthesis of a series of recent brain imaging and behavioral studies from our laboratory seeking to elucidate these questions.

Adjacent and Nonadjacent Repetitions at Birth

If rule-learning ability is hardwired into our cognitive system or language acquisition faculty, we might find evidence of it very early in life, possibly at birth. We ran two brain imaging studies with newborns using near-infrared spectroscopy (NIRS) to address this question. Using this technique also allowed us to investigate the neural basis of rule learning.

In the first study (Gervain, Macagno, Cogoi, Pena, & Mehler, 2008), we tested whether newborns are able to distinguish adjacent (ABB: "mubaba," "penana," etc.) and nonadjacent (ABA: "bamuba," "napena," etc.) repetitions from random controls (ABC: "mubage," "penaku," etc.). This is an important departure from the previous studies, which all compared different repetition-based structures between them. With a random control, we could explore the processing and learning of each repetition structure separately, independently of the other. In one experiment, we used NIRS to monitor 3-day-old infants' responses to 14 blocks of ABB sequences intermixed with 14 blocks of ABC sequences (Fig. 7.1a). Blocks consisted of 10 different sequences. All sequences occurred only once throughout the whole experiment, ensuring that the response we observe is not evoked by the recognition or learning of any particular items, but rather to the repetition structure all ABB items share. The ABB and ABC sequences were synthesized to have monotonous pitch and equal syllable durations. The optical sensors were placed over the bilateral temporal and frontal areas (Fig. 7.1b), as these areas are known to be involved in auditory processing and higher order sequence learning, respectively, in adults and infants (Dehaene-Lambertz, Hertz-Pannier, Dubois, & Dehaene, 2008). NIRS measures the changes in the concentration of oxygenated and deoxygenated hemoglobin (oxyHb and deoxyHb, respectively) accompanying focal brain activity. The results are shown in Fig. 7.1c. We obtained a significantly greater increase in oxyHb for ABB than for ABC in both the left and the right temporal areas, with a stronger response in the left. Further, we also observed the same differential response in the left frontal area, involving inferior frontal regions, possibly Broca's area.

Interestingly, in the left frontal area, the differential response increased over the course of the experiment: the difference between the two grammars was already significant over the first 4 blocks but further increased over subsequent blocks (Fig. 7.2).

These results are compatible with a double mechanism account: the adjacent repetition is immediately detected and distinguished from the random control sequence by an automatic, low-level mechanism, and over the course of the experiment, a higher order representation is built of the structural commonality of the items by

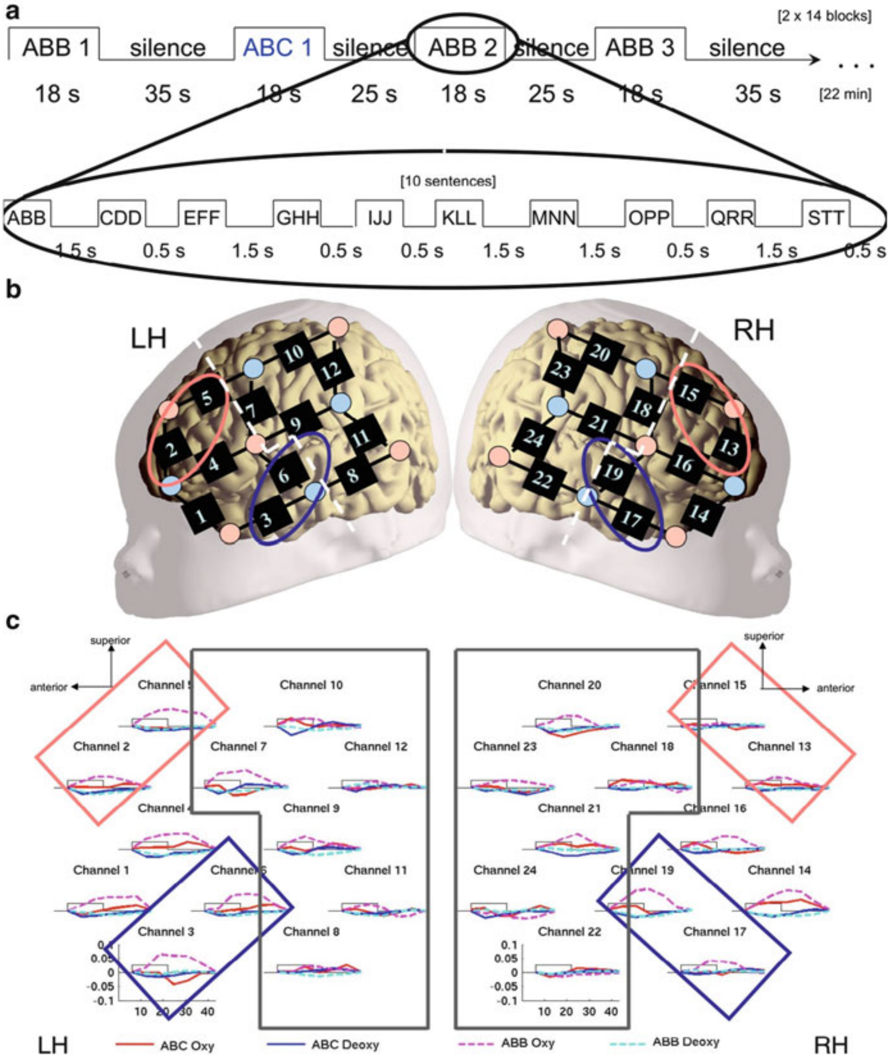


Fig. 7.1 The design, the optode placement, and the results of Experiment 1 in Gervain et al. (2008)

a symbolic mechanism which thus extracts the ABB structure. It is not implausible that the two mechanisms are linked, the output of the automatic repetition detector feeding into the more abstract symbolic mechanism, which assigns a sequential position to the detected immediate repetition. The localization of the responses confirms this interpretation: the immediate differential response suggesting rapid, automatic, perceptual-level detection is found in both the bilateral temporal and the left frontal areas, whereas the increase over time, corresponding to the extraction of the abstract regularity, is restricted to the left frontal area, known to be involved in

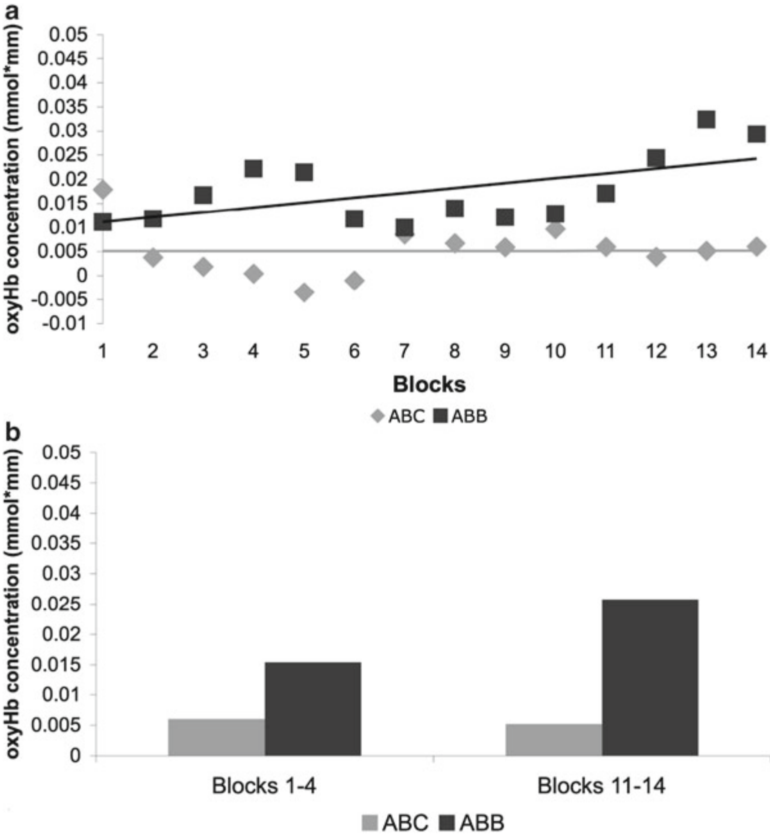


Fig. 7.2 The time course of the oxyHb response in the left frontal area in Experiment 1 of Gervain et al. (2008)

higher order learning, sequence learning, and structural organization in language and other domains.

In a second experiment, we tested whether nonadjacent repetitions were also detected. We used the same design and procedure as in the previous experiment except that we used ABA vs. ABC sequences. Interestingly, in this case, we did not find a differential response. We observed canonical hemodynamic responses in the bilateral temporal areas as in the first experiment, confirming that the stimuli were processed and the absence of a difference is not due to methodological errors, but the amplitude of the response was similar for both grammars. It seems, therefore, that newborns cannot detect nonadjacent repetitions. Several not mutually exclusive explanations exist. Nonadjacent repetitions, unlike adjacent ones, do not constitute a special configuration; hence, they are not detected automatically. Without the input of such an automatic detection, the symbolic representation might not be built. It is also possible that the processing window in which stimuli are compared is very small in newborns, only allowing local comparisons. This can be seen as a memory

Fig. 7.3 The alternating/non-alternating design of Experiment 2 in Gervain et al. (2012)

limitation or the interfering effect of the intermediate B item. Further research will be needed to fully explore this question. However, some of our subsequent work, discussed in Part IV below, already sheds some light on this issue.

In sum, we found that newborns can detect adjacent, but not nonadjacent, repetitions.[1] The pattern of results observed suggests that both an automatic perceptual response and a more abstract symbolic response might be present. However, more empirical evidence is necessary to confirm the existence of the latter.

In a second study (Gervain, Berent, & Werker, 2012), we therefore sought to corroborate the existence of this abstract mechanism. As discussed before, the abstract mechanism might be responsible for building a symbolic representation of the grammars by combining the information about the relation between items, i.e., repetition, whether it is detected by an automatic mechanism or not, with information about the sequential position of items and their relations. In other words, even if the detection of immediate repetition is automatic and perceptual, a more abstract mechanism is needed to account for the successful discrimination of grammars in which the adjacent repetition holds between items in different positions. We, therefore, directly compared newborns' discrimination of sequence-initial (AAB) vs. sequence-final (ABB) adjacent repetition sequences—an ability 7-month-olds have been demonstrated to possess (Marcus et al., 1999). We ran three NIRS experiments using the same procedure and optode placement as above. In the first experiment, we simply demonstrated that AAB sequences are discriminated from random ABC controls in exactly the same fashion as ABBs are: the differential response, i.e., greater oxyHb and smaller deoxyHb concentration changes for the AAB than for the ABC grammar were localized in the bilateral temporal and left frontal areas. In the second and most relevant experiment, we compared ABB vs. AAB sequences in a block design known in behavioral research as the alternating/non-alternating paradigm (Fig. 7.3). In half of the blocks, we only presented one regularity (either only

[1] As NIRS is a relatively new technique in developmental cognitive neuroscience, we ran a behavioral validation study (Gervain et al., 2011) using the classical high-amplitude sucking procedure to make sure that the absence of a differential response for ABA vs. ABC was not simply due to the low sensitivity of the NIRS measure. The behavioral results confirmed the NIRS findings: newborns can discriminate between ABB vs. ABC, but not between ABA vs. ABC grammars.

AAB or only ABB); in the other half, we presented both regularities in a mixed fashion (AAB and ABB items alternated in the block). If babies can discriminate the two grammars, the alternating blocks will sound different for them than the non-alternating ones; if they cannot, all blocks will sound the same. In particular, we expect a different response to the alternating and non-alternating blocks in the left frontal area, which is responsible for high-order structural learning.

The findings confirmed our predictions: we found a canonical hemodynamic response in the bilateral temporal areas with no difference between the two block types, but a greater response to the non-alternating blocks in the left frontal areas. This suggests that newborns could discriminate the two blocks and that this discrimination took place, as expected in the brain regions responsible for pattern extraction and higher order processing of structure, confirming the existence of an abstract mechanism. The non-alternating blocks, implementing a single regularity, gave rise to a greater response, because it was straightforward to extract the rule from these stimuli. The alternating blocks with both regularities present did not allow for a single generalization to be extracted, and learning both was, not surprisingly, too challenging for newborns. In a third experiment, we directly compared AAB vs. ABB blocks in a simple block design used in Gervain et al. (2008) and in the first experiment of this study to test for any preference between initial and final repetitions. The hemodynamic response measured by NIRS is an indicator of the metabolic effort related to neural processing. It could be the case that one structure is easier to process than the other, since certain linguistic phenomena preferentially appear at one sequence edge or the other. In particular, repetitions are not allowed in sequence-initial position in Semitic languages (Berent & Shimron, 2003). However, we obtained similar responses to the two grammars: canonical hemodynamic responses in the bilateral temporal and left frontal areas, suggesting that neither edge position is privileged over the other. Similar results were observed when the ABB and AAB non-alternating blocks of the second experiment were compared.

Taken together, these two studies suggest that newborns have the ability to detect adjacent repetitions using an automatic mechanism, but they are also able to build abstract, symbolic representations of complex structures based on these repetitions. However, neither ability extends to nonadjacent repetitions at this age.

Adjacent and Nonadjacent Repetitions During Development

In the light of the previous newborn studies, it was crucial to determine at what age infants can first detect nonadjacent repetitions. The previous studies, e.g., Marcus et al. (1999), always compared adjacent and nonadjacent repetition structures directly. In such tasks, the recognition of the adjacent repetition alone could suffice to show successful performance by identifying the nonadjacent structure as simply being different by exclusion. It was therefore important to compare the two repetition structures to the same random ABC control as in the newborn studies.

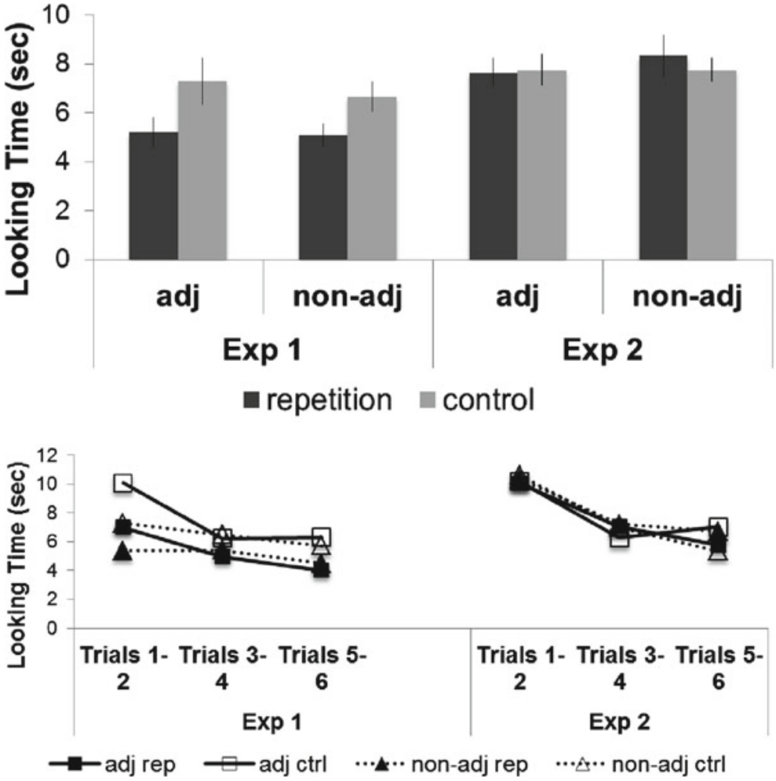


Fig. 7.4 The results of the Gervain and Werker (2012) study. The *upper panel* shows average looking times per condition. The *lower panel* represents looking times to subsequent trial pairs per condition

In a behavioral experiment (Gervain & Werker, 2012) very similar to Marcus et al.'s (1999) original design, we thus familiarized one group of 7-month-old infants with the ABB grammar and another group with the ABA grammar. In the subsequent test phase, we presented them with novel tokens of their grammar of familiarization as well as with ABC tokens. Both groups showed significantly longer looks to the unfamiliar ABC tokens. We found no difference between the two groups either in their overall looking times to the repetition vs. control items or in the time course of their looking times over the 12 test trials (Fig. 7.4). This suggests that by 7 months, infants can represent both adjacent and nonadjacent repetitions.

However, it might be the case that infants paid more attention to the ABC controls not because they were novel with respect to the familiarization pattern, but because they were more variable than the repetition patterns, as they had three rather than two different syllables. To exclude this possibility and to test whether there is any spontaneous preference for adjacent and/or nonadjacent repetitions, we ran a second experiment, which was identical to the first except that infants received no familiarization. Thus we simply tested them on their presence between ABB vs. ABC and ABA vs. ABC. If the detection of (at least adjacent) repetitions is automatic, it might induce a preference for this easily detectable structure. However, we found no evidence of any difference: infants looked equally long to adjacent repetitions, nonadjacent repetitions, and control sequences (Fig. 7.4).

These results show that at 7 months, infants can represent nonadjacent repetitions, and there is no indication of any difference between the processing of adjacent and nonadjacent repetitions. As nonadjacent repetitions are not considered to be a perceptual primitive in the same way as adjacent repetitions are, 7-month-olds' success in this task suggests, although does not constitute definitive evidence, that abstract symbolic representations extend to both adjacent and nonadjacent identity relations at this age. This is not to deny that adjacent repetitions might be easier to process under some circumstances, e.g., as in the already cited Johnson et al. (2009) study, where even 11-month-old infants were unable to learn ABA patterns, when shown sequentially presented visual stimuli, but succeeded with ABB and AAB structures.

Discussion and Conclusion

This chapter has summarized two recent lines of work exploring the learning of simple structural regularities based on adjacent and nonadjacent repetitions in young infants. These studies have found that newborns are able to detect, learn, and discriminate adjacent repetition-based structures from random controls and from one another when the repetition is at different sequence edges, but they cannot do the same with nonadjacent repetitions. However, by 7 months of age, both are readily learned and discriminated from random controls.

More research will be needed to explore the change that takes place between birth and 7 months, allowing infants to start tracking nonadjacent regularities. The scenario that emerges from these results is that two different mechanisms exist with different developmental trajectories. One candidate mechanism that has been proposed to account for adjacent repetitions is an automatic, low-level repetition or identity detector (Endress et al., 2009). As this mechanism is also present in nonlinguistic animals, e.g., honeybees (Giurfa et al., 2001), it is not implausible to assume that it might be present in humans as early as birth. The output of this mechanism might then feed into a more abstract, symbolic rule-learning mechanism to represent adjacent repetitions. By contrast, nonadjacent repetitions might only be processed by a more abstract, symbolic mechanism, which requires time to develop sufficiently to allow the representation of distant relations between items.

Elucidating the nature of the mechanisms involved in rule learning will take us closer to the understanding of how language, this uniquely human trait, is acquired over ontogenetic and phylogenetic development.

References

Berent, I., & Shimron, J. (2003). Co-occurrence restrictions on identical consonants in the Hebrew lexicon: Are they due to similarity? *Journal of Linguistics, 39*(01), 31–55.

Chomsky, N. (1959). A review of B. F. Skinner's verbal behavior. *Language, 35*(1), 26–58.

Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., & Dehaene, S. (2008). How does early brain organization promote language acquisition in humans? *European Review, 16*(4), 399–411.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1997). *Rethinking innateness: A connectionist perspective on development (Neural networks and connectionist modeling)*. Boston, MA: The MIT Press.

Endress, A. D., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition, 105*(3), 577–614.

Endress, A. D., Nespor, M., & Mehler, J. (2009). Perceptual and memory constraints on language acquisition. *Trends in Cognitive Sciences, 13*(8), 348–353.

Endress, A. D., Scholl, B. J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology. General, 134*(3), 406–419.

Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps five-month-olds learn abstract rules. *Developmental Science, 12*(4), 504–509.

Gervain, J., Berent, I., & Werker, J. F. (2012). Binding at birth: The newborn brain detects identity relations and sequential position in speech. *Journal of Cognitive Neuroscience, 24*(3), 564–574.

Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America, 105*(37), 14222–14227.

Gervain, J., Mehler, J., Werker, J. F., Nelson, C. A., Csibra, G., Lloyd-Fox, S., et al. (2011). Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience, 1*(1), 22–46.

Gervain, J., & Werker, J. F. (2012). Learning non-adjacent regularities at age 0 ; 7. *Journal of Child Language*, *FirstView*, 1–13.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of "sameness" and "difference" in an insect. *Nature, 410*(6831), 930–933.

Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., et al. (2009). Abstract rule learning for visual sequences in 8- and 11-month-olds. *Infancy, 14*(1), 2–18.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A. E., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*, 111–116.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18*(5), 387.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77–80.

Ng, H. L., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *The Quarterly Journal of Experimental Psychology Section A, 55*(2), 391–424.

Parisse, C. (2005). New perspectives on language development and the innateness of grammatical knowledge. *Language Sciences, 27*(4), 383–401.

Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.

Rabagliati, H., Senghas, A., Johnson, S., & Marcus, G. F. (2012). Infant rule learning: Advantage language, or advantage speech? *PLoS ONE, 7*(7), e40517.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition, 105*(3), 669–680.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*(3), 209–253.

Wagner, J. B., Fox, S. E., Tager-Flusberg, H., & Nelson, C. A. (2011). Neural processing of repetition and non-repetition grammars in 7- and 9-month-old infants. *Frontiers in Psychology, 2*.

Chapter 8
Is There a Brain Basis of Recursion?

Angela D. Friederici

Introduction

It is not possible to give a simple answer to the question "Where in the brain is recursion?" Nevertheless, considering the fact that human beings process recursive structures, or hierarchical structures, in language, it is obvious that these processes must be supported by certain neuroanatomical structures in the brain. It is thus pertinent to look at those structures in the brain that support the human language capacity which has been claimed to be characterized by recursion (Hauser, Chomsky, & Fitch, 2002).

Phylogenetic Differences Between Human and Nonhuman Animal Brain Structures

When considering Broca's area in the context of the human ability to process language, a look across primates comparing those with language abilities and those without might be at order (Fig. 8.1).

At a microstructural level Petrides and Pandya (2009) have shown similar cytoarchitectonic structures that are present in the human and the macaque brain with respect to certain areas in the prefrontal cortex which, however, differ in relative size. This appears to be the case in particular for Broca's area, consisting of BA 44, BA 45a, and BA 45b (Brodmann, 1909), which are associated with language in humans. In macaques, BA 44 exists but is not as visible at the lateral

A.D. Friederici (✉)
Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
e-mail: angelafr@cbs.mpg.de

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_8,

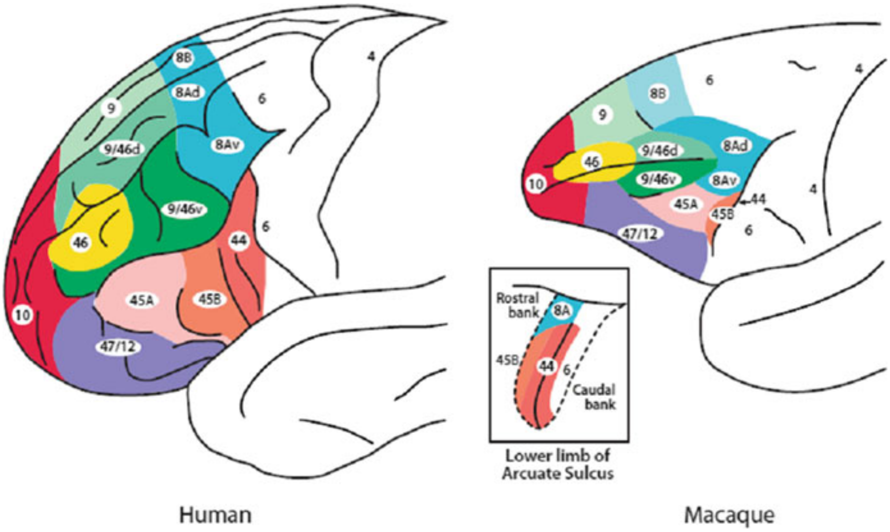


Fig. 8.1 Comparison between human and macaque brain (*Figure adapted from Petrides & Pandya, PLoS Biology,* 2009)

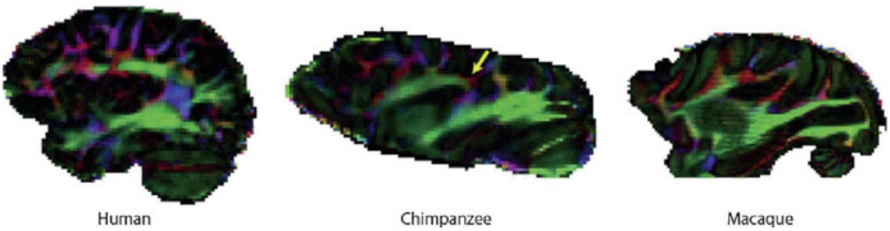


Fig. 8.2 Comparison between fiber tracts connecting the frontal and temporal lobes in humans, chimpanzees, and macaques (*Figure adapted from Rilling et al., Nature Neuroscience,* 2008)

surface as it is located in the sulcus. A closer observation shows that BA 44 is much smaller than BA 45 in the macaque compared to humans. Thus, the relative size of these areas appears to be different, but so far nobody has provided a quantification of these differences. In the temporal cortex there are obvious differences: in the human brain there are three gyri, i.e., superior, middle, and inferior temporal gyrus, whereas in the macaque brain there are only two, i.e., superior and inferior temporal gyrus. However, the superior temporal gyrus, where the auditory processes are taking place, is very similar in the macaque and in the human brain. It seems thus that both brains have a similar structure at least for the processing auditory information (Fig. 8.2).

These frontal and temporal areas are connected by fiber bundles. Rilling, Glasser, Preuss, Ma, Zhao, Hu and Behrens (2008) have compared these fiber bundles in the human brain, the macaque brain, and the chimpanzee brain in order to see how

Broca's area in the inferior frontal gyrus and Wernicke's area in the temporal gyrus are connected in the different species. It is well known that there are two streams of fibers in the brain connecting frontal and the temporal cortex: the first one is the dorsal pathway; the other one is the ventral pathway. Rilling et al. (2008) data show that the dorsal pathway is very prominent in the human brain and less prominent in the macaque. In the chimpanzee, the strength of this pathway lies in between. On the contrary, the ventral pathway connecting the inferior frontal gyrus to the temporal cortex is less prominent in the human brain but very dominant in the macaque brain. Thus there is a phylogenetic difference between macaques and humans in connecting those areas implied in language processing. The question arises to what extend can this be related to the human language faculty.

Behavioral and Neurological Differences in Processing Artificial Grammars

We do know that nonhuman primates cannot deal with sentences of a natural language. However, it has been discussed whether nonhuman primates are able to process hierarchical structures in an artificial grammar (Fig. 8.3).

Fitch and Hauser (2004) have used the grammars described in Fig. 8.3 to compare cotton-top tamarins' and humans' ability to process grammars. Their results show that both tamarins and humans could learn the Finite State Grammar (FSG), but that the nonhuman animals were not able to process the Phrase Structure Grammar (PSG), while humans were able to do it. These results were at the basis of a whole set of studies in human and nonhuman animals. One study in nonhuman animals, Gentner et al. (2006), showed that starlings, after thousands of trials, were able to process both types of grammars. But it has been discussed whether those animals had succeeded to process PSG only by counting or because they had succeeded to create an internal representation the concerned hierarchical structures (Beckers et al., 2012).

This critique, however, also holds for one of our studies with humans (Friederici et al., 2006). We used the same structures as those used by Fitch and Hauser (2004) in order to observe whether these two grammars which differentiated humans from

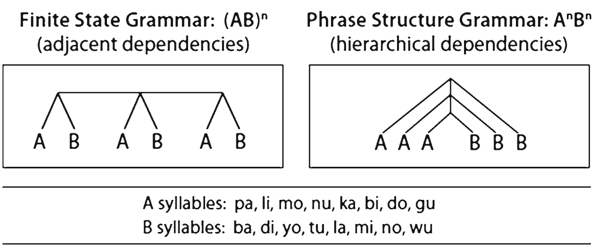


Fig. 8.3 Grammars used by Fitch and Hauser for their comparison (*Figure adapted from Fitch & Hauser, Science,* 2004)

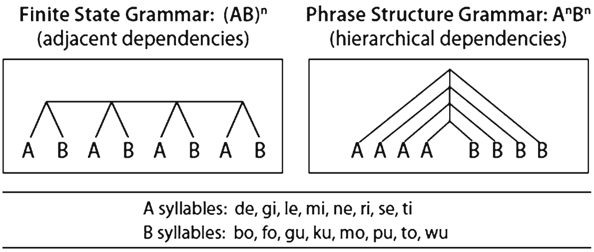


Fig. 8.4 Grammars and syllables used to observe brain activations (Friederici et al, 2006)

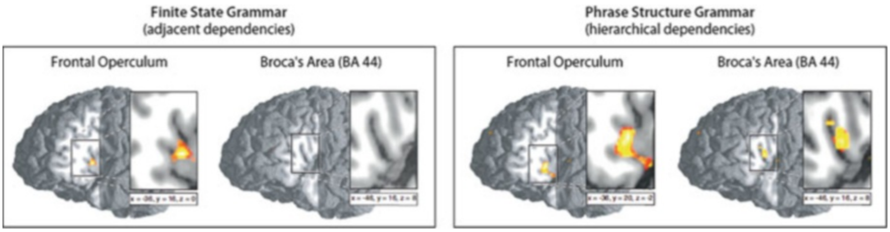


Fig. 8.5 Activation in Broca and FOP for FSG and PSG (*Figure adapted from Friederici et al., PNAS,* 2006)

animals are processed by different parts of the human brain. We hypothesized that the FSG, also processed in nonhuman primates, might be processed by areas that are phylogenetically older than those used to process the PSG, the grammar only processed by humans. Sequences of both grammars were presented visually to our subjects who had to learn, by trial and error, the rules used underlying the sequences (Fig. 8.4). Two days later, in the scanner, the subjects were presented with correct and incorrect sequences of both grammars (Fig. 8.5).

The results show that for the FSG, the frontal operculum is the only active area: this area is located ventral to Broca's area and is phylogenetically older than Broca's area. Sanides (1962) has made an argument of why this area is an older part of the cortex; because through the evolutionary principle of gradiation, Broca's region originates among others from opercular cortex (see also Amunts & Zilles, 2012).

On the contrary, for the more complex sequences based on a PSG, we observed activation in both the frontal operculum and in Broca's area. We controlled the fact that this was not just a consequence of the difficulty of the sentences: the subjects were equally good in learning these two grammars. We thus concluded that it was a function of the particular syntactic structure they had learned.

There are nevertheless two possible caveats. First again, it must be noted that the subjects could reach the correct solution simply by counting, without having constructed an inner structural representation. Secondly, since this study was based on a between subject design, one could argue that subjects in the FSG group were just much better in learning languages than those in the PSG group. These two issues will be taken up below.

About Fiber Tracts and Artificial Grammars

Before discussing these issues additional evidence for a neuroanatomical differentiation between Broca's area and FOP will have to be provided. With this goal we examined whether these areas, which are located very adjacently in the inferior frontal cortex, can be differentiated by their projections to the temporal cortex by means of different fiber tracts (Friederici et al., 2006).

This research is based on 40 subjects. Figure 8.6 displays 4 representative subjects. The data provide structural information showing that from the point of maximum activation for the simple grammar with adjacent dependencies (FSP), the fiber tract projects ventrally from FOP to the temporal cortex, for all the subjects we have been testing. On the other hand, when taking the point of maximum activation for the complex grammar with hierarchical dependencies (PSG), we start the fiber tracting from Broca's, and we observe a projection via the dorsal pathway.

These findings indicate that FOP and Broca's area cannot only be functionally segregated, but that they also are part of different structural networks. This seems to be a strong argument in favor of the hypothesis that those two areas are different.

As preliminary conclusion, we can thus state that the processing of FSG sequences is associated to FOP which is a phylogenetically older area, while the processing of PSG sequences is associated to Broca's area, which is a phylogenetically younger area. These two areas are involved in different networks.

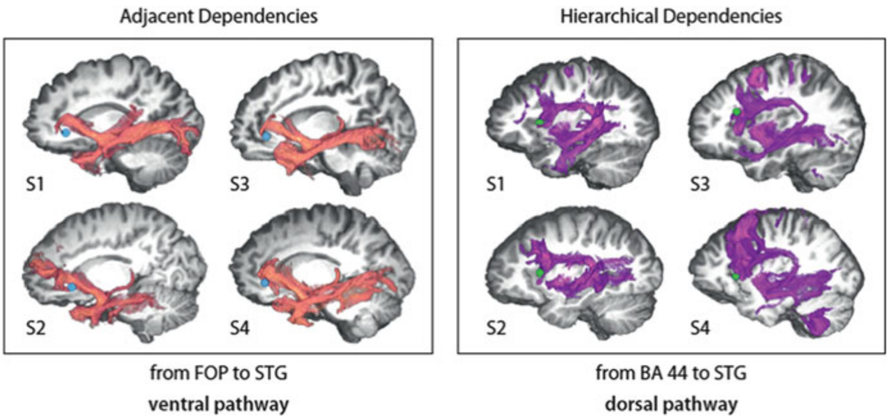


Fig. 8.6 Projections of Broca and FOP to temporal lobe (*Figure adapted from Friederici et al., PNAS,* 2006)

Checking the Construction of the Sequences' Hierarchies

There were two caveats to our preliminary conclusion. One was that the PSG could have been processed without constructing the hierarchies between the respective members in the tree. The other was that the two types of grammar were tested in different subjects. We tried to counter these caveats by another experiment.

We wanted to make sure that the dependency between the elements constituting a sequence was actually such that the subject had to build up a hierarchical structure. In order to do so we establish a relation between the elements that related to each other on the basis of voiced and unvoiced consonants (Bahlmann et al., 2008) (Fig. 8.7).

We presented both sequence types to the same subjects, two days before the scanning took place. They learned the rules by trial and error. During the scanning processes, the subjects were shown correct sequences and incorrect sequences. For each sequence the subjects had to decide whether it was "grammatical according to the rule." This enabled us to compare directly those two grammar types.

In Fig. 8.8a we collapse over correct and incorrect elements. The data show that there is more activation for the hierarchical structures than for the adjacent structure in the Broca's area. It would be tempting to say that this is only due to the detection of a syntactic error in the incorrect sequence. The results shown in Fig. 8.8b, however, prove that this is not the case: the results are exactly the same when we take the correct sequences only. This leads us to the following (and second) interim conclusion: syntactic hierarchy processing occurs in Broca's area.

Two further issues should be considered before a final conclusion can be drawn. First, does our interim conclusion also hold for similar structures in a natural language? Second, to what extend is the activation in Broca's area driven by aspects of working memory which may come into play when processing embedded structures?

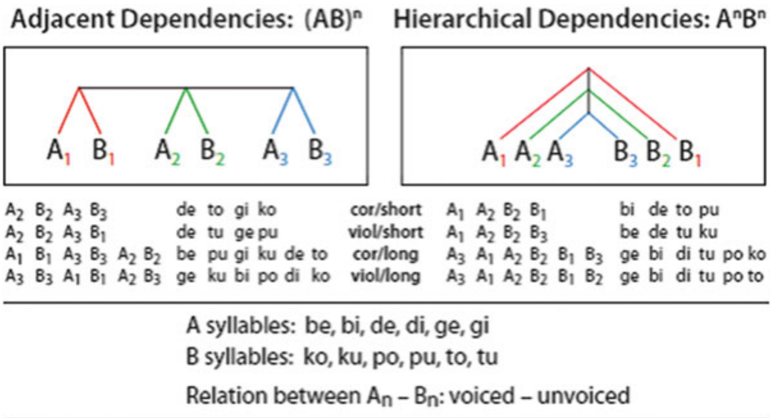


Fig. 8.7 New experiment based on voiced–unvoiced relation (*Figure adapted from Bahlmann et al., NeuroImage,* 2008)

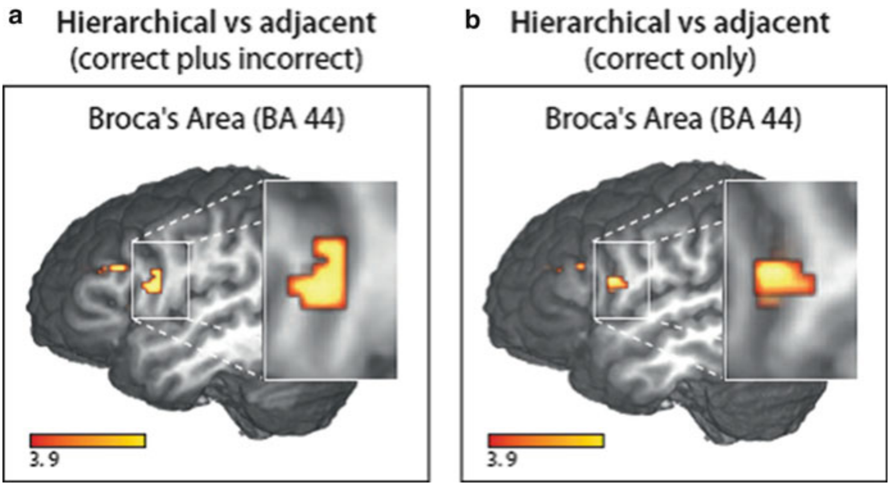


Fig. 8.8 Activation in Broca's area (BA 44). (**a**) Correct plus incorrect. (**b**) Correct only. Adapted from Bahlmann et al., NeuroImage, 2008

Dissociating the Roles of Syntactic Hierarchies and of Working Memory

Since Broca's area (BA 44/45) has also been discussed to serve the function of working memory (Jonides et al., 1998; Smith and Jonides, 1998), the next experiment investigated whether syntactic hierarchy (SH) and working memory (WM) can be dissociated neuroanatomically.

This experiment was based on a 2-by-2 factorial design: hierarchical syntactic processing on the one hand and working memory on the other. The purpose of this experiment was to show a possible dissociation of these two aspects neuroanatomically and to do so in an experiment using a natural language (Makuuchi et al., 2009) (Fig. 8.9).

We used German sentences: the nature of German allows us to introduce a distance variation between the dependent elements (subject–verb distance) and sentences with or without nested structures, including adverbial phrases in the middle. In Fig. 8.10 exemplifies two examples of good German sentences which is the mother tongue of our subjects. The first sentence clearly has an embedded structure: "Peter wusste, dass Maria, die Hans, der gut aussah, liebte, Johann geküsst hatte"/"Peter knew that Maria who Hans who is good looking loved John kissed." The subject "Maria" goes with the verb "geküsst," while the subject relative pronoun "die" goes with the verb "liebte," and finally the subject relative pronoun "der" goes with the verb "aussah." Using the structure of the German language, it is possible to create sentences with or without embeddings. The structure of the German language enabled us to independently vary the effect of the distance of subject–verb dependency (working memory) and the number of embeddings (syntactic hierarchy).

We worked with 18 right-handed German natives. The following stimuli were used: 46 linear structure, dependency short; 46 linear structure, dependency long;

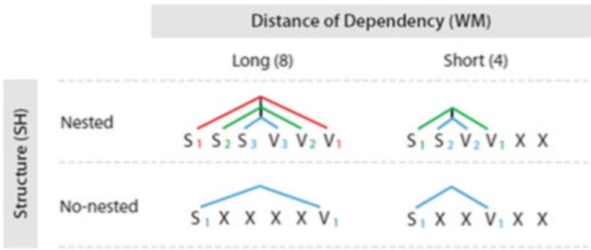


Fig. 8.9 2×2 Design with the factors working memory (WM) and syntactic hierarchy (SH) (*Figure adapted from Makuuchi et al., PNAS, 2009*)

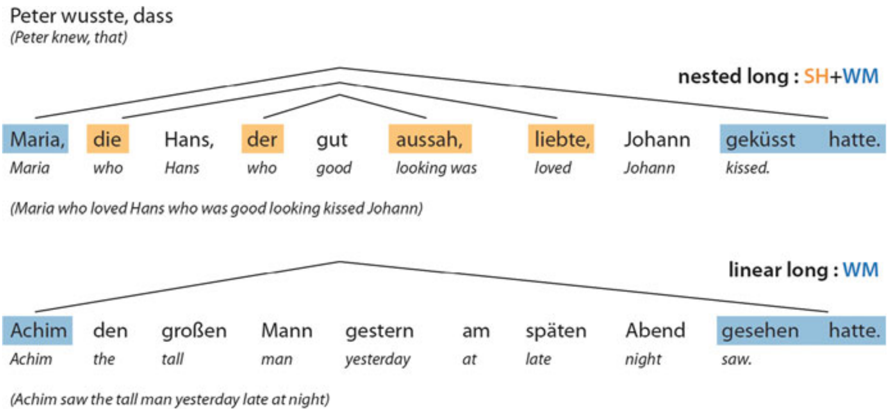


Fig. 8.10 Natural language: German center-embedded sentences (*Figure adapted from Makuuchi et al., PNAS, 2009*)

46 nested (embedded structures), dependency short; and 46 nested (embedded structures), dependency long. The stimuli were presented visually. The subject also had to undergo a question-answering task in 20 % of the trials. The brain activation results are shown in Fig. 8.11.

The green area in Fig. 8.11 is the cytoarchitectonically defined area BA 44. The yellow one is BA 45. Together they form Broca's area. The main effect of syntactic hierarchy (as shown by the orange activation) is exactly located in BA 44, the left pars opercularis (LPO). The main effect of distance, color coded in light blue, is located in the left inferior frontal sulcus (LIFS). The violet zone represents the cluster used for the functional connectivity analysis. The lines drawn between the different areas represent the functional connectivity between those areas, which indicate that BA 44 works together with the inferior frontal sulcus, when processing syntactically complex sentences.

We can thus conclude: there is a functional segregation in the left IFG between the syntax and working memory. Hierarchical structure processing occurs in the BA 44 while distance processing occurs in the left inferior frontal sulcus. Although both areas have their separate roles, they have to work in concert in order to

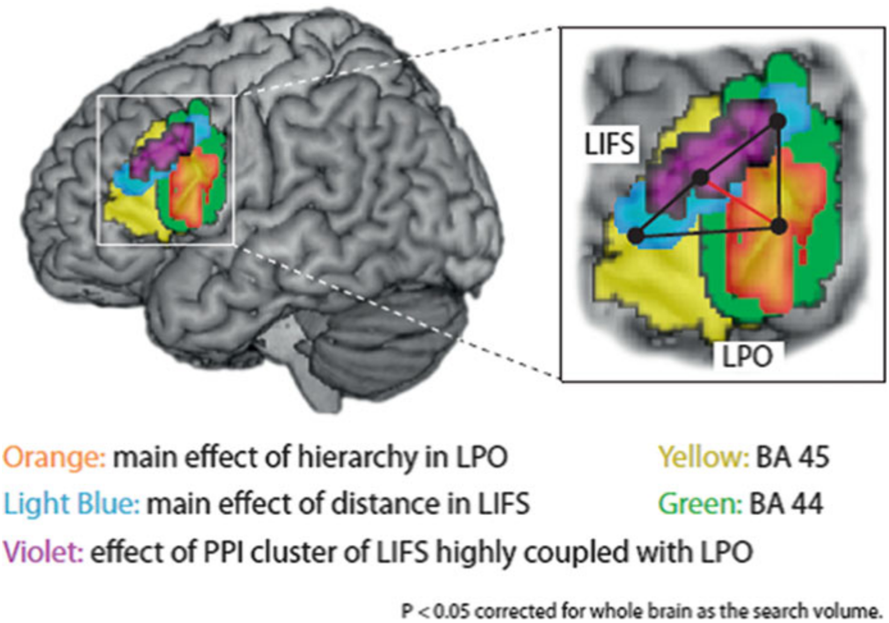


Fig. 8.11 Functional activation and connectivity in inferior frontal gyrus (*Figure adapted from Makuuchi et al., PNAS,* 2009)

correctly process hierarchically structured sentences. In order to check whether the subjects really treated these sentences, we asked each of them questions concerned the role assignment in these embedded and very complex sentences. All subjects performed well on this comprehension task.

About the Connections Between Broca's Area and the Temporal Lobe

We have seen that Broca's area is involved in the processing of complex hierarchical structures. We have seen previously the fiber tract showing the links between Broca's area (BA 44) and the temporal cortex via a dorsal fiber tract (Fig. 8.6). The posterior temporal cortex is often activated in natural language sentence processing (Bornkessel et al., 2005; Sanit, & Grodzinsky, 2010). This area is not activated when subjects are confronted to artificial grammars, but there is activation when subjects have to process complex sentences in a natural language.

The data shown in Fig. 8.12 show that, as far as natural language sentences are concerned, in the case of embedded sentences, Broca's area (BA 44) is active and also the posterior superior temporal gyrus (STG) (Friederici, Makuuchi, & Bahlmann, 2009).

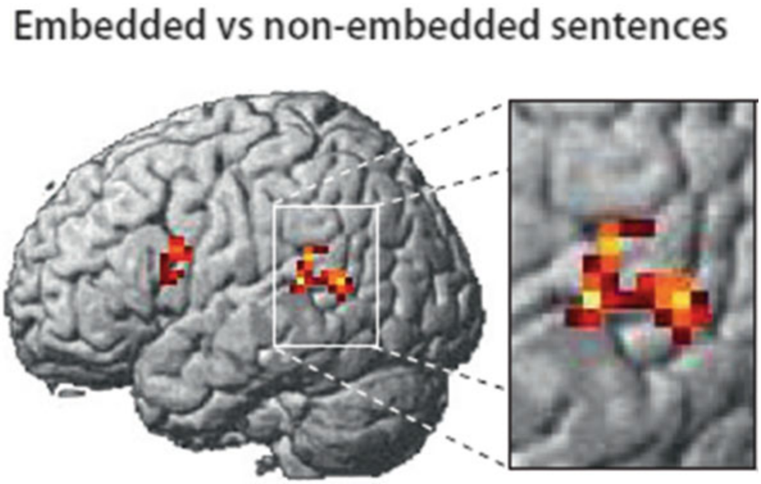


Fig. 8.12 Activation in the posterior superior temporal cortex (*Figure adapted from Friederici, Makuuchi & Bahlmann, NeuroReport,* 2009)

This suggests the hypothesis that BA 44 and the posterior STG are crucially involved in the processing of complex syntax and that the fiber tract connecting those two areas, the dorsal pathway as seen above, is of particular importance. There are several arguments in favor of this hypothesis.

We first have phylogenetic arguments. Going back to the phylogenetic data, we know now that the dorsal pathway, the fiber track which seems important for meaningful complex syntactic processes, is strong in humans but weak in the nonhuman primate (Rilling et al., 2008). We also have results showing that nonhuman primates do not process embedded sentences whereas humans do (Fitch & Hauser, 2004).

We also have onto-genetic arguments. Here the idea is if this dorsal pathway is not fully developed in children, then this less mature pathway in children should be associated to poor performances in processing complex syntactic structures.

The question is thus the following: "at what age does this behavior change?" Dittmar et al. (2008) have looked into children's comprehension of complex, but very short sentences. This implies that there was no heavy memory load implied. The authors used either canonical, subject-first, structure or noncanonical, object-first, structure. This can be done easily in German using case marking of the noun-phrase to place the object in front of the verb. Furthermore, in order to not bias the child by the semantics of the verb, they used pseudoverbs, i.e., verbs that do not exist in German, but carry an inflectional ending. These authors have conducted a couple of experiments: sentence-picture matching and acting out in order to get the children involved in the task. Their results from the picture-matching tasks are shown in Fig. 8.13. The scores of correct processing of object-first structures by children aged 2 years and 7 months are at chance level. The same is true for children at the age of 4 years and 10 months. These scores of processing these object-first

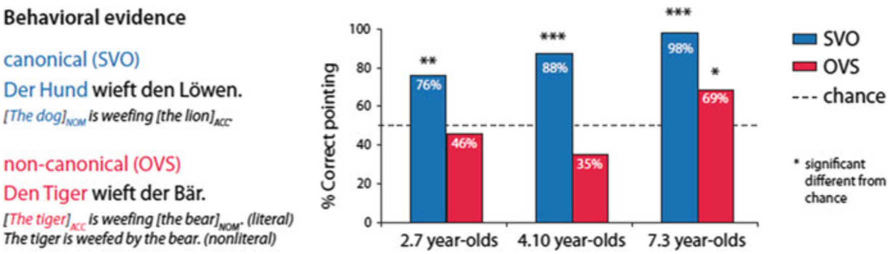


Fig. 8.13 Dorsal pathway and ontogeny: behavioral evidence (*Figure adapted from Dittmar et al., Child Development,* 2008)

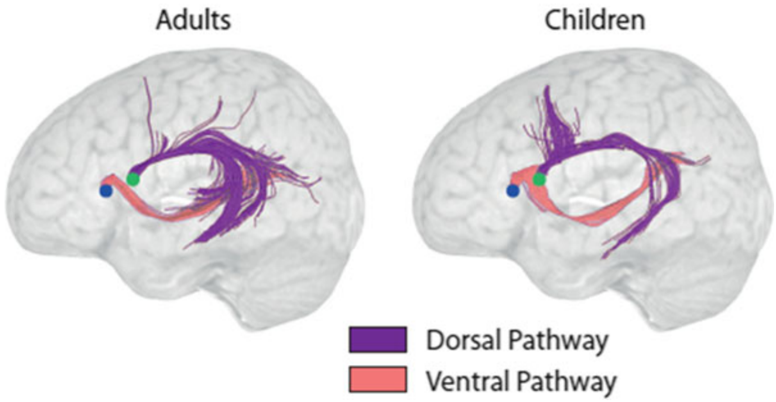


Fig. 8.14 Dorsal pathway and ontogeny: tractographic evidence (*Figure adapted from Brauer et al., Cerebral Cortex,* 2011)

structures are just above chance by the age of 7 years. This suggested to us that it should be interesting to have a closer look at the seven-year-olds with respect to their brain structure. The hypothesis was that if the dorsal fiber tract as seen in the adults is crucial for the processing of complex sentences, then this fiber tract may not have reached an adult stage of maturation in 7-year-olds (Fig. 8.14).

For this experiment, we worked with adults and children. In order to tests the children, we first gave them a training in a mock scanner before they went to the real scanner. We first did a functional study (Brauer and Friederici, 2007) and then the structural analysis (Brauer, Anwander, & Friederici, 2011).

The functional study in processing sentences showed that adults activate mainly BA 44 and a fiber tracking computation from the most activated point shows an involvement of the dorsal pathway in these subjects. For 7-year-old children the main activation was found in BA 45, anterior to BA 44. A fiber tracking computation from their most activated point shows that children involve the ventral pathway. A comparison between adults' and children's ventral pathways shows that these ventral pathways are very similar in terms of their strength whereas the dorsal pathways are not. Strength means here myelination of the fiber tracts, with the myelin being essential from the transmission of electrical impulses and thus the transmission of information. A close examination of the dorsal pathway in children shows that this fiber tract connecting the posterior portion of Broca's area (BA 44) to the temporal lobe is much less mature than in the adults. Children, in contrast to adults, also some connections to the motor cortex. This might be a reminiscence of the fact that children at this age, when dealing with a difficult sentence, still involve subvocal articulation. At least the structural connection needed to do so is still present.

Conclusion

We can thus conclude that those processes related to recursion that we have been discussing here recruit Broca's area. Although it appears to be inappropriate to say "recursion is in Broca's area," it is obvious from the data presented above that recursion in language recruits Broca's area (in particular BA 44). Hierarchical processing in natural grammar and artificial grammar mainly recruits BA 44 (for a recent review see Friederici, 2011). The fiber tract that connects the posterior portion of Broca's area to the temporal cortex dorsally plays a crucial role in processing complex hierarchical structures in language.

An open question is what happens to recursion in the nonlanguage domain. Additional studies in the nonlanguage domain (e.g., the visual domain, the mathematics domain) indicate that an involvement of Broca's area (Friederici et al., 2011), however, with Broca's area each time being part of a different neural network. Thus it is not Broca's area alone, but it is Broca's area plus some other brain regions that supports the process necessary to treat recursive and hierarchical structures in different domains. For language it is the temporal cortex, and, whereas for example, for visual space it would be rather the parietal cortex. Therefore, the main conclusion of this paper is to urge the reader not to think about "a particular process is in a particular area," but rather it is a particular network of different areas that supports a particular process. We now have to find out what are the overlaps and what are the non-overlaps between the networks in the different cognitive domains.

Acknowledgment The present article is based on a talk presented at the Conference on Language and Recursion held on March 14, 2011, at the University of Mons, Belgium. I would like to thank Prof. Dr. Lowenthal for his invaluable help in transcribing the talk and transforming the transcript into a readable text which laid the ground for the present text.

References

Amunts, K., & Zilles, K. (2012). Architecture and organizational principles of Broca's regions. *Trends in Cognitive Sciences, 16*, 418–426.

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage, 41*, 525–534.

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *NeuroReport, 23*, 139–145.

Bornkessel, I., Zyssett, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage, 26*, 221–233.

Brauer, J., Anwander, A., & Friederici, A. D. (2011). Neuroanatomical prerequisites for language functions in the maturing brain. *Cerebral Cortex, 21*, 459–466.

Brauer, J., & Friederici, A. D. (2007). Functional neural networks of semantic and syntactic processes in the developing brain. *Journal of Cognitive Neuroscience, 19*, 1609–1623.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde*. Leipzig: Johann Ambrosius Barth.

Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child Development, 79*, 1152–1167.

Fitch, W. T., & Hauser, M. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 303*, 377–380.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*, 1357–1392.

Friederici, A. D., Bahlmann, J., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion. *Biolinguistics, 5*, 87–104.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458–2463.

Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport, 20*, 563–568.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204–1207.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A., & Reuter-Lorenz, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 8410–8413.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences, 106*, 8362–8367.

Petrides, M., & Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biology, 7*(8), e1000170.

Rilling, K. J., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience, 11*, 426–428.

Sanides, F. (1962). The architecture of the human frontal lobe and the relation to its functional differentiation. *International Journal of Neurology, 5*, 247–261.

Sanit, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage, 51*, 1285–1293.

Smith, E. E., & Johnides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences of the United States of America, 95*, 12061–12068.

Chapter 9
Primate Communication: Meaning from Strings of Calls

Klaus Zuberbühler and Alban Lemasson

Introduction

Understanding the evolution of human language is a major challenge for science. With no relevant fossil record, one empirical approach has been to investigate primate communication. An early method has been to teach human-raised apes an artificial language, although results have been somewhat disappointing because it has been very difficult to teach primates to vocalise on command, or to modify the acoustic structure of their vocalisations (Fitch & Zuberbühler, 2013). Teaching apes communication systems based on visual signals has been somewhat more successful, but a major weakness has always been that subjects are required to interact with human caretakers who have used operant conditioning techniques to elicit communication. In the typical situation, subjects communicate to persuade a caretaker to release food under his or her control. However, natural primate communication is rarely used to persuade others to release food. For this reason, studies of natural communication in ecologically relevant situations are likely to provide evolutionarily more relevant insights into how human language has evolved from earlier forms of communication. Our approach in this chapter is to survey the literature on natural primate communication for signs of acoustic flexibility that may have functioned as preadaptations for the evolution of speech.

K. Zuberbühler (✉)
School of Psychology and Neuroscience, University of St Andrews, Fife, Scotland, UK

Cognitive Science Centre, University of Neuchâtel, Neuchâtel, Switzerland
e-mail: Klaus.zuberbuehler@unine.ch

A. Lemasson
Animal and Human Ethology, Université de Rennes 1 – CNRS, Rennes, France

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_9,

Acoustic Flexibility

The older literature has sought to explain speech evolution in terms of anatomical specialisations that enable humans to generate constrictions in the vocal tract to produce speech sounds (Lieberman, 1984). More recent studies have suggested that mammalian vocal tracts are in principle capable of producing speech sounds, suggesting that limitations are more likely at the neural and not the anatomical level (e.g. Fitch & Zuberbühler, 2013; Riede, Bronson, Hatzikirou, & Zuberbühler, 2005) although little is known about the actual dynamics of primate vocal tracts during vocalisation (de Boer & Fitch, 2010; Fitch, 2000).

An unresolved problem is what genetic changes during human evolution have led to greater motor control of the vocal tract in humans. One hypothesis is that the main cause is recent mutations in the human lineage in the FOXP2 gene, a structure that plays a role during brain development and orofacial control (Enard et al., 2002). Prior to this event, the argument goes, early humans would have possessed vocal behaviour not fundamentally different from modern primates. Modern humans still use a considerable range of nonlinguistic vocalisations, which often play integral roles during speech acts (Clark, 1996). One hypothesis thus is that speech evolution has taken place within a vocal apparatus in early humans that was already capable of fully functional vocal behaviour, as still found in modern primates.

When studying the natural vocal behaviour of non-human primates, areas of enhanced vocal control are thus of particular relevance, as these may have been important as preadaptations in human speech evolution. Two main functions of acoustic variation have been identified in primate calls. One is to express social affinity and identity and another to refer to external events, such as food or predators. In the following, we will review evidence for both functions. Acoustic flexibility, finally, can be increased in at least two ways, either by morphological changes of vocal units or by combining vocal units into sequences. The two mechanisms grade into each other, but for conceptual reasons, we will continue to discuss them separately.

Call Variants

The relation between modifications of acoustic structure of calls and social relations has been investigated in Japanese macaques. These primates produce acoustically distinct 'coo' calls to maintain social contact. Calls given by one individual typically trigger a response by another individual, usually an animal that is socially close. The 'coo' calls vary in their acoustic structure and show distinct individual vocal signatures (Sugiura, 1993; Sugiura & Masataka, 1994). In one experiment, recordings of coo calls with different acoustic structures were played back to examine whether the responding monkeys matched the acoustic features of their own calls to the calls they have heard. The study reported some effects in terms of the frequency range between playback and response calls (Sugiura, 1998).

More recent research on Campbell's monkeys has reported similar findings. In one study, an analysis of the acoustic structure of contact calls has revealed several call variants, some of which are shared between individuals, often in relation to social affinity (Lemasson, Gautier, & Hausberger, 2003). The call variants change throughout adult life, with greater changes occurring after social disturbances (Lemasson & Hausberger, 2004). Playback of females' currently produced variants triggers vocal responses from other group members, whereas former and no longer used variants never trigger responses, indicating that acoustic variants are part of long-term social memories (Lemasson, Hausberger, & Zuberbühler, 2005).

For closely related Diana monkeys, Candiotti, Zuberbühler, and Lemasson (2012a) have reported patterns of acoustic similarity and dissimilarity in the contact calls. Vocal dissimilarity was enhanced during poor visibility and in the presence of neighbours, while vocal similarity was enhanced during vocal interactions, possibly because females matched the frequency contour of their own call with another female's preceding call. It is possible that this effect was due to closely related individuals being morphologically more similar and more likely to respond to each other's calls. For Campbell's monkeys, this does not seem to be a valid explanation since the acoustic similarity of contact calls was related to social bond strength but not to genetic relatedness (Lemasson et al., 2011). Whether this is also true for Diana monkeys or other primate species is currently unknown.

In some non-human primate species, adult males and females have the same or similar vocal repertoires, but in some other species, vocal behaviour is very sex-specific. Overall, females appear to be acoustically more variable than males, possibly the result of higher rates of social interactions compared to males (Bouchet, Pellier, Blois-Heulin, & Lemasson, 2010). A general pattern seems to be that the degree of acoustic variability of different calls is directly linked to their social functions, with particularly high variabilities in affiliative calls (Bouchet, Blois-Heulin, Pellier, Zuberbühler, & Lemasson, 2012; Bouchet, Blois-Heulin, & Lemasson, 2013; Lemasson & Hausberger, 2011). In species with strong sex-specific calling behaviour, males appear to undergo a transition during puberty, suggesting that this development is hormonally governed. However, even in species with strong sexual dimorphisms in vocal behaviour, such as forest guenons or gibbons, adult females occasionally produce male call variants (Bouchet, Blois-Heulin, Lemasson, 2012; Chen, Kamolnorranath, & Kaplan, 2008; Geissmann, 1983; Ouattara, Zuberbühler, N'goran, Gombert, & Lemasson, 2009), notably when males have failed to respond to events that normally trigger their vocalisations. Overall, the relationship between male and female vocal repertoire is not so well understood, and individuals may be more flexible than what is generally assumed.

Another important context of call production is agonistic interactions, in which acoustically variable vocal behaviour has been observed (Gouzoules, Gouzoules, & Marler, 1984). In chimpanzees, acoustically distinct screams depend on the role that individuals play in a conflict, victim or aggressor, and on the severity of the attack (Slocombe & Zuberbühler, 2005a). In playback experiments, subjects consistently discriminated between victim screams to mild and severe aggression, although the calls formed part of a graded continuum, suggesting that listeners attended to the acoustic structure of calls as indicators of different types of social interactions (Slocombe, Townsend, & Zuberbühler, 2009). Victims have also been observed to alter the acoustic structure of their screams depending on the severity of aggression experienced. In addition, victims of severe attacks produce screams that appear to exaggerate the true level of aggression experienced but only if there is at least one listener in the audience who matches or surpasses the aggressor in rank. This type of vocal behaviour may be an expression of understanding third-party relationships (Slocombe & Zuberbühler, 2007).

A second main function of primate calls and call variants is to refer to external events. An early study with Japanese macaques showed that individuals responded to different foods with acoustically distinct food call variants (Green, 1975). In subsequent playback experiments, subjects discriminated between the acoustic variants even if individual signature features were artificially removed (May, Moody, & Stebbins, 1988). Similarly, chimpanzees produced acoustically distinct 'rough grunts' when encountering food, and calls varied acoustically with the perceived quality of the food (Slocombe & Zuberbühler, 2006). In playback experiments, chimpanzees behaved as if call variants were meaningful to them, as indicators of how others assessed the quality of the encountered food (Slocombe & Zuberbühler, 2005b).

Call variants have also been reported in primates responding to predators. Vervet monkey alarm calling has long been the paradigmatic example of how primates use vocalisations in response to predators. In this species, there is a close and direct relationship between acoustically distinct alarm vocalisations and the presence of distinct predator types. In free-ranging Campbell's monkeys, females also produce three basic alarm call types, but one call type can be discriminated into acoustic variants depending on the type of disturbance. Females raised in captivity do not produce these alarm calls, although one acoustic variant, given in response to humans, was absent in the repertoire of free-ranging individuals (Ouattara, Zuberbühler, N'goran, Gombert, & Lemasson, 2009).

Finally, call variants can be the result of an affixation process. In male Campbell's monkeys, two acoustically distinct alarm calls have been described (Zuberbühler, 2002), which are acoustically modified by a process analogous to affixation. Male alarm calls consist of an acoustically variable stem, which can be followed by an acoustically invariable suffix. Observations and field experiments showed that suffixation functioned to broaden the calls' meaning by transforming predator-specific alarms to general disturbance and alert calls (Ouattara, Lemasson, & Zuberbühler, 2009a). Other examples of affixation are found in red-capped mangabeys (Bouchet et al., 2010) and Diana monkeys (Candiotti, Zuberbühler, & Lemasson, 2012b). In mangabeys, an acoustically invariable 'uh' unit can be affixed to four different call types given in feeding, aggressive and affiliative contexts. Affixation is determined by whether the individual is calling alone or participating in call exchanges with other group members. Similarly, in Diana monkeys, the suffix is a frequency-modulated arched unit, which can be added to three different call types. Two suffix variants have been discriminated, depending on whether the arch is complete or broken, which is determined by the general visibility.

Sequences of Calls

Another way to study the origins of acoustic flexibility is by investigating sequential production of calls. As with acoustic call variants, this can be done in relation to external events and in the context of maintaining social relationships. Variation can be generated by callers varying the frequency and rate of calls, or by varying the composition of calls in a sequence.

Frequencies and Rates

When confronted with leopards or crowned eagles, black-and-white colobus monkeys across Africa produce two basic alarm call structures, snorts and acoustically variable roaring sequences. Leopards typically trigger utterances containing many roaring sequences with a few calls each, while eagles trigger a small number of sequences with many calls each (Schel, Tranquilli, & Zuberbühler, 2009). Playbacks of the two types of call sequences did not cause differences in locomotor responses as movement was always towards the simulated caller, but gaze direction was highly predator-specific. When hearing leopard-related stimuli, monkeys were significantly more likely to scan the area beneath them than when hearing eagle-related stimuli, which caused more scanning above. If monkeys responded with their own calls, then always with the matching sequences, suggesting that the monkeys have attended to the compositional aspects of the utterances (Schel, Candiotti, & Zuberbühler, 2010).

Another source of variability is the difference in call rates. Apart from producing roars to predators, colobus males often also roar in the predawn hours and without any obvious external trigger. There are no conspicuous structural differences between predawn roars and eagle roars, although detailed acoustic analyses have shown differences in the duration of the first roaring phrase and in the initial call rate between the two contexts (Schel & Zuberbühler, 2012a, 2012b). Other evidence for systematic differences in call rates is from Campbell's monkeys in relation to predation than non-predation events, to visual and auditory evidence of predators, to differences in antipredator behaviour, and as part of individual differences. Although the relevant playbacks have yet to be carried out, differences in call rate could provide listeners with cues about the caller's antipredator behaviour, event type experienced, and his identity (Lemasson, Ouattara, Bouchet, & Zuberbühler, 2010). Another species that has been investigated is the red-capped mangabeys that produce most of their calls in sequences. Here, the sequence length and complexity appears to be dependent on the number of other individuals vocalising at the same time (Bouchet et al., 2010).

Call Combinations

Evidence for combinatorial properties in primate vocal behaviour comes from Campbell's monkeys, putty-nosed monkeys, lar gibbons, and titi monkeys. All four species have been studied in the context of predation avoidance with the following key findings.

First, a study on Campbell's monkeys has revealed that, as a result of the suffixation process described before, males generate at least six distinct call types, which they combine into various sequences in context-specific ways. Relations between acoustic structure and context were found in relation to travel, falling trees, neighbouring groups, non-predatory animals, unspecific threats, and different predators. Some combinations of calls altered the meaning of calls, other combinations changed the meaning completely (Ouattara, Lemasson, & Zuberbühler, 2009b). Within the predator responses, different sequences were a reflection of how the caller discovered the predator, i.e. by acoustic or visual means. Wild Diana monkeys, who frequently form polyspecific associations with Campbell's monkeys, comprehend the semantic changes caused by some Campbell's monkey call combinations: in response to Campbell's males alarm calls to leopards and crowned eagles, Diana monkeys respond with their own corresponding alarm calls (Zuberbühler, 2000). However, in less dangerous situations, Campbell's males emit sequences composed of a pair of low, resounding 'boom' calls before their alarm calls. Playbacks of boom-introduced Campbell's alarm call sequences did not elicited alarm call responses in Diana monkeys, indicating that the booms have affected the semantic specificity of the subsequent alarm calls (Zuberbühler, 2002).

Second, a series of studies carried out with free-ranging putty-nosed monkeys in Gashaka Gumti National Park, Nigeria, has provided further evidence that primate call sequences act as carriers of meaning. In response to leopards and crowned eagles, males produce two call types, hacks and pyows, but unlike previous studies of alarm calling in guenon monkeys, there were no clear predator-specific regularities at the level of individual calls. There were, however, striking regularities between call combinations and predator types (Arnold & Zuberbühler, 2006a). Playback experiments further revealed that monkeys combined the two calls into at least three meaningful sequences, to signal the presence of (1) a crowned eagle (series of hacks) and (2) a leopard (series of pyows) or to initiate (3) group movement (pyows followed by hacks; Arnold & Zuberbühler, 2006b). 'Pyow-hack' sequences were not compositional in that the individual calls contributed to their overall meaning. Instead, the monkeys appeared to perceive sequences as idiomatic expressions (Arnold & Zuberbühler, 2008, 2012).

Third, related work carried out with lar gibbons at Khao Yai National Park, Thailand, showed that these primates produced duet songs assembled from a repertoire of call notes to repel conspecific intruders, advertise pair bonds, and attract mates. In field experiments, gibbons also used songs to predators. Predator-induced songs were identical to normal songs in the basic notes, but there were differences in how the notes were assembled into songs, thus providing evidence of referential signalling based on combinatorial rules in a free-ranging ape species (Clarke, Reichard, & Zuberbühler, 2006).

Finally, recent work with black-fronted titi monkeys in South Eastern Brazil has shown that, upon encountering predators, adult monkeys respond with call sequences consisting of two brief, high-pitched calls with different frequency contours. Call A was mainly given to raptors but also to other threats within the canopy, particularly predatory capuchin monkeys, while call B was mainly given to predatory or non-predatory disturbances on the ground (Caesar, Byrne, Young, & Zuberbühler, 2012). In playback experiments, conspecifics responded to these calls in adaptive ways: listeners preferentially looked upward when hearing raptor-related calls and towards the presumed caller when hearing terrestrial predator-related calls. If locomotor responses occurred, then always in the expected direction, suggesting that monkeys discriminated between calls based of their acoustic features to make inferences about the nature of the disturbance (Caesar, Byrne, Hoppitt, Young, & Zuberbühler, 2012).

A second context in which call combinations have been observed is foraging. In bonobos, individuals produce five acoustically different call types during interactions with food. The calls are usually given as part of non-random sequences related to the type of food encountered by the caller (Clay & Zuberbühler, 2009). Overall, individual call types are poor indicators of food quality, but context specificity is greater at the call sequence level. To investigate whether receivers could extract meaning from different call sequences, we first trained individuals at Twycross Zoo to find kiwi (preferred) and apples (less preferred) at two different locations of their enclosure. In playback experiments, we then broadcasted sequences of calls, originally produced by a familiar individual to either kiwi or apples. All sequences contained the same number of calls but varied in call composition. Results showed that subjects devoted significantly more search effort to the area indicated by the call sequence, suggesting that listeners attended to the sequences to make inferences about the food encountered by a caller (Clay & Zuberbühler, 2011).

A largely unresolved issue is how sequentially organised calling behaviour develops and how much voluntary control individuals have when generating call variants or call sequences. In one study, the alarm calling behaviour of two adjacent populations of Diana monkeys at Taï forest (Ivory Coast) and Tiwai Island (Sierra Leone) with and without leopards was compared, which revealed consistent differences in how males assembled calls into sequences. At Tiwai, a habitat without leopards, males responded to leopards and general disturbances in the same way, while at Taï, a habitat with leopards, males discriminated by giving different call sequences, suggesting that ontogenetic experience determines how calling behaviour (Stephan & Zuberbühler, 2008), but how much primates are able to control their vocal behaviour is still very much unknown.

Vocal Interactions

A final way to investigate the importance of call combinations is at the level of vocal interactions. Here, especially recipients are cognitively challenged if they are motivated to extract social information by attending to vocal interactions between other group members. The classic study probably is Bergman et al.'s (2003) study on baboons extracting complex social information from listening to call exchanges by other group members. In another well-studied system, the Diana monkey contact calls, it has been found that females continuously produce and respond to each other's contact calls. Calling is contagious in that calls typically elicit vocal responses from out-of-sight group members within a few seconds. Call rates are elevated when the visibility is poor, when the group spread is large, and when the group is not associated with other monkey species, suggesting that calling is a reflection of perceived threat (Uster & Zuberbühler, 2001). Whether recipients are able to extract information from this system of co-ordinated vigilance is currently unknown.

In Campbell's monkeys, vocal interactions have also been studied. Individuals do not produce contact calls randomly by respect a minimum inter-call interval, with the result that callers rarely overlap with each other (Lemasson, Gandon, & Hausberger, 2010). Adults are particularly good at attending to this 'turn-taking rule' (Lemasson, Glas, Barbu, Lacroix, Guilloux, Remeuf, & Koda, 2011), but this is not the case for younger individuals who often do not adhere to them, that the same was recently found in a study on Japanese macaques (Lemasson, Guilloux, Rizaldi, Barbu, Lacroix, & Koda, 2013). When appropriate and inappropriate vocal exchanges (i.e. turn-taking respected or violated) to young and adult individuals, only adults showed evidence of discrimination. At the same time, adult group members received more vocal responses than youngsters (Lemasson, Gandon, & Hausberger, 2010).

Conclusions

This review of some of the current empirical evidence demonstrates that non-human primates are able to increase the effective size of their usually small vocal repertoires by varying the acoustic structure of basic call types, by varying the rate and frequency of calls given as part of longer utterances, and by combining calls into more complex sequential structures. These phenomena are seen in communication contexts that serve a purely social function and in response to relevant external events. Although non-human primate communication is often said to be fundamentally different from human speech, the main source of difference appears to be in terms of vocal control. The differences to humans are especially striking if comparisons are made to some marine mammals and bird species, some of which appear to have near-human sophisticated control over their vocal production. What are the biological roots and evolutionary pressures to control sound production and learn the vocal utterances of others? The studies reviewed here demonstrate that primates have some control over the acoustic fine structure of some of their calls, both when interacting with each other and when responding to external events. We propose that the human capacity to control vocal production has evolved gradually from existing preadaptations, some of which still present in the vocal behaviour of modern primates. We also propose that human vocal behaviour, before the advent of advanced motor control, strongly resembled the vocal behaviour of modern primates.

References

Arnold, K., & Zuberbühler, K. (2006a). Semantic combinations in primate calls. *Nature, 441*(7091), 303–303.

Arnold, K., & Zuberbühler, K. (2006b). The alarm-calling system of adult male putty-nosed monkeys, Cercopithecus nictitans martini. *Animal Behaviour, 72*, 643–653.

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology, 18*(5), R202–R203.

Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain & Language, 120*(3), 303–309.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science 302*(5648), 1234–1236.

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2012). Age and sex-specific patterns of vocal behavior in De Brazza's monkeys (*Cercopithecus neglectus*). *American Journal of Primatology, 74*(1), 12–18.

Bouchet, H., Blois-Heulin, C., & Lemasson, A. (2013). Social complexity parallels vocal complexity: A comparison of three nonhuman primate species. *Frontiers in Psychology 4*, 390.

Bouchet, H., Blois-Heulin, C., Pellier, A.-S., Zuberbühler, K., & Lemasson, A. (2012). Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys (*Cercocebus torquatus*). *Journal of Comparative Psychology, 126*(1), 45–56.

Bouchet, H., Pellier, A.-S., Blois-Heulin, C., & Lemasson, A. (2010). Sex differences in the vocal repertoire of adult red-capped mangabeys (*Cercocebus torquatus*): a multi-level acoustic analysis. *American Journal of Primatology, 72*, 360–375.

Caesar, C., Byrne, R. W., Hoppitt, W., Young, R. J., & Zuberbühler, K. (2012). Evidence for semantic communication in titi monkey alarm calls. *Animal Behaviour, 84*(2), 405–411.

Caesar, C., Byrne, R. W., Young, R. J., & Zuberbühler, K. (2012). The alarm call system of wild black-fronted titi monkeys, Callicebus nigrifrons. *Behavioral Ecology and Sociobiology, 66*(5), 653–667.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012a). Convergence and divergence in Diana monkey vocalizations. *Biology Letters, 8*(3), 382–385.

Candiotti, A., Zuberbühler, K., & Lemasson, A. (2012b). Context-related call combinations in female Diana monkeys. *Animal Cognition, 15*(3), 327–339.

Chen, H.-C., Kamolnorranath, S., & Kaplan, G. (2008). Female crested gibbons (genus *Nomascus*) sing male song. *Vietnamese Journal of Primatology, 2*, 47–53.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clarke, E., Reichard, U., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One, 1*(1), e73.

Clay, Z., & Zuberbühler, K. (2009). Food-associated calling sequences in bonobos. *Animal Behavior, 77*(6), 1387–1396.

Clay, Z., & Zuberbühler, K. (2011). Bonobos extract meaning from call sequences. *PLoS One, 6*(4).

de Boer, B., & Fitch, W. T. (2010). Computer models of vocal tract evolution: An overview and critique. *Adaptive Behavior, 18*(1), 36–47.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature, 418*(6900), 869–872.

Fitch, W. T. (2000). The phonetic potential of nonhuman vocal tracts: Comparative cineradiographic observations of vocalizing animals. *Phonetica, 57*(2–4), 205–218.

Fitch, W. T., & Zuberbühler, K. (2013). Primate precursors to human language: beyond discontinuity. In E. Altenmuller, S. Schmidt, & E. Zimmerman (Eds.), *Evolution of emotional communication: from sounds in nonhuman mammals to speech and music in man*. Oxford: Oxford University Press.

Geissmann, T. (1983). Female capped gibbon (Hylobates pileatus Gray 1891) sings male song. *Journal of Human Evolution, 12*, 667–671.

Gouzoules, H., Gouzoules, S., & Marler, P. (1984). Macaque Rhesus (Macaca mulata) screams: Representation signs in the recruiting of agonistique aid. *Animal Behaviour, 32*, 182–193.

Green, S. (1975). Communication by a graded vocal system in Japanese monkeys. In L. A. Rosenblum (Ed.), *Primate behavior* (Vol. 4). New York: Academic.

Lemasson, A., Gandon, E., & Hausberger, M. (2010). Attention to elders' voice in nonhuman primates. *Biology Letters, 6*, 325–328.

Lemasson, A., Gautier, J. P., & Hausberger, M. (2003). Vocal similarities and social bonds in Campbell's monkey (Cercopithecus campbelli). *Comptes Rendus Biologies, 326*(12), 1185–1193.

Lemasson, A., Glas, L., Barbu, S., Lacroix, A., Guilloux, M., Remeuf, K., et al. (2011). Youngsters do not pay attention to conversational rules: Also in nonhuman primates? *Nature Scientific Reports, 1*, 22.

Lemasson, A., & Hausberger, M. (2004). Patterns of vocal sharing and social dynamics in a captive group of Campbell's monkeys (Cercopithecus campbelli campbelli). *Journal of Comparative Psychology, 118*(3), 347–359.

Lemasson, A., & Hausberger, M. (2011). Acoustic variability and social significance of calls in female Campbell's monkeys (Cercopithecus campbelli campbelli). *Journal of the Acoustical Society of America, 129*(5), 3341–3352.

Lemasson, A., Hausberger, M., & Zuberbühler, K. (2005). Socially meaningful vocal plasticity in adult Campbell's monkeys (Cercopithecus campbelli). *Journal of Comparative Psychology, 119*, 220–229.

Lemasson, A., Ouattara, K., Bouchet, H., & Zuberbühler, K. (2010). Speed of call delivery is related to context and caller identity in Campbell's monkey males. *Naturwissenschaften, 97*(11), 1023.

Lemasson, A., Ouattara, K., Petit, E. J., & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology, 11*.

Lemasson, A., Guilloux, M., Rizaldi, Barbu, S., Lacroix, A., & Koda, H. (2013). Age- and sex-dependent contact call usage in Japanese macaques. *Primates, 54*(3), 283–291.

Lieberman, P. (1984). *The biology and evolution of language*. Cambridge MA: Harvard University Press.

May, B., Moody, D. B., & Stebbins, W. C. (1988). The significant features of Japanese macaque coo sounds: A psychophysical study. *Animal Behaviour, 36*, 1432–1444.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys use affixation to alter call meaning. *PLoS One, 4*(11), e7808.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America, 106*(51), 22026–22031.

Ouattara, K., Zuberbühler, K., N'goran, K. E., Gombert, J.-E., & Lemasson, A. (2009). The alarm call system of female Campbell's monkeys. *Animal Behaviour, 78*, 35–44.

Riede, T., Bronson, E., Hatzikirou, H., & Zuberbühler, K. (2005). Vocal production mechanisms in a non-human primate: morphological data and a model. *Journal of Human Evolution, 48*(1), 85–96.

Schel, A. M., Candiotti, A., & Zuberbühler, K. (2010). Predator-deterring alarm call sequences in Guereza colobus monkeys are meaningful to conspecifics. *Animal Behaviour, 80*(5), 799–808.

Schel, A. M., Tranquilli, S., & Zuberbühler, K. (2009). The alarm call system of black-and-white colobus monkeys. *Journal of Comparative Psychology, 123*(2), 136–150.

Schel, A. M., & Zuberbühler, K. (2012a). Predator and non-predator long-distance calls in Guereza colobus monkeys. *Behavioural Processes, 91*(1), 41–49.

Schel, A. M., & Zuberbühler, K. (2012b). Dawn chorusing in guereza colobus monkeys. *Behavioral Ecology and Sociobiology, 66*(3), 361–373.

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (Pan troglodytes schweinfurthii) distinguish between different scream types: evidence from a playback study. *Animal Cognition, 12*(3), 441–449.

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences of the United States of America, 104*(43), 17228–17233. doi:10.1073/pnas.0706741104.

Slocombe, K. E., & Zuberbühler, K. (2005a). Agonistic screams in wild chimpanzees (Pan troglodytes schweinfurthii) vary as a function of social role. *Journal of Comparative Psychology, 119*(1), 67–77.

Slocombe, K. E., & Zuberbühler, K. (2005b). Functionally referential communication in a chimpanzee. *Current Biology, 15*(19), 1779–1784.

Slocombe, K. E., & Zuberbühler, K. (2006). Food-associated calls in chimpanzees: responses to food types or food preferences? *Animal Behaviour, 72*, 989–999.

Stephan, C., & Zuberbühler, K. (2008). Predation increases acoustic complexity in primate alarm calls. *Biology Letters, 4*(6), 641–644.

Sugiura, H. (1993). Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Animal Behaviour, 124*(3–4), 207–225.

Sugiura, H. (1998). Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Animal Behaviour, 55*, 673–687.

Sugiura, H., & Masataka, N. (1994). Flexibility in vocal behaviour (coo calls) in Japanese macaques. In J. Roeder, B. Thierry, J. R. Anderson, & N. Herrenschmidt (Eds.), *Current primatology* (Social development, learning and behaviour, Vol. 2, pp. 215–220). University of Louis Pasteur: Strasbourg.

Uster, D., & Zuberbühler, K. (2001). The functional significance of Diana monkey 'clear' calls. *Behaviour, 138*, 741–756.

Zuberbühler, K. (2000). Interspecific semantic communication in two forest monkeys. *Proceedings of the Royal Society of London Series B-Biological Sciences, 267*(1444), 713–718.

Zuberbühler, K. (2002). A syntactic rule in forest monkey communication. *Animal Behaviour, 63*, 293–299.

Part IV
About Formal Grammars and Artificial Intelligence

Chapter 10
Unification and Efficient Computation in the Minimalist Program

Sandiway Fong

Introduction

From the perspective of computational modeling, in recent years, the shift from a declarative to an operational approach to the description of linguistic theories has important implications for both efficient computation and the space of possible implementations.

The declarative approach is exemplified by the Principles and Parameters (P&P) approach (Chomsky, 1981), in which linguistic constraints (or filters) are abstractly stated over syntactic configurations at various levels of representation and, perhaps derivationally, between levels of representation. These filters may call upon a variety of linguistic devices that may be generated in the course of a derivation, including indices, chains (of movement), and various empty categories, e.g., traces. The (not unsubstantive) problem of generating appropriate syntactic descriptions, perhaps by some top-down, bottom-up, or mixed assembly of syntactic objects, is (apparently) left to the grammar designer. There is considerable freedom of implementation. For instance, apart from basic dependencies between various principles, the order of application of the constraints is undetermined. Perhaps the most straightforward realization of the P&P approach can be found in the

S. Fong (✉)
Departments of Linguistics and Computer Science,
University of Arizona, Tucson, AZ 85721, USA
e-mail: sandiway@email.arizona.edu

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_10,

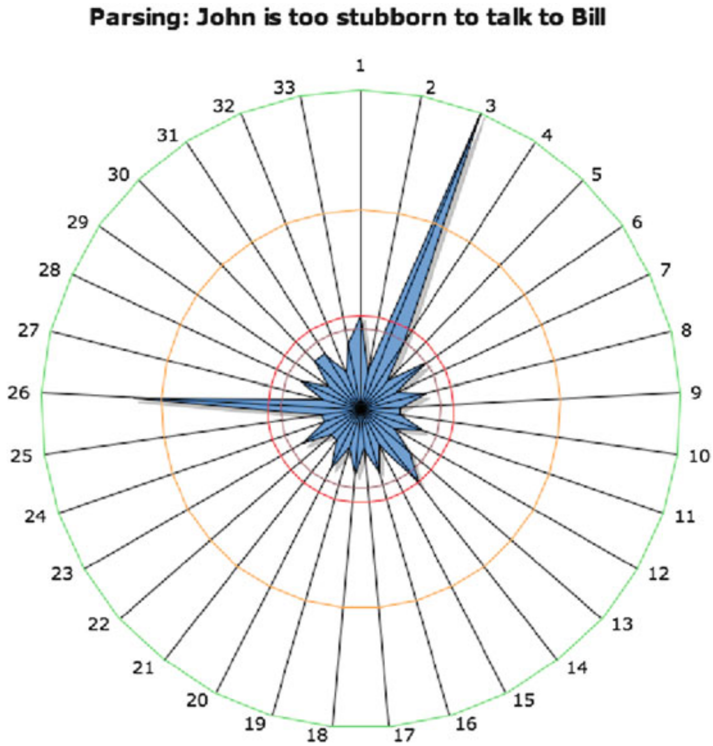


Fig. 10.1 Example of the generate-and-test paradigm

largely generate-and-test paradigm given in Fong (1991).[1] The radar plot in Fig. 10.1 illustrates the degree of overgeneration and thus computational inefficiency, typically observed in the implemented parsing system. For the sentence *John is too stubborn to talk to Bill*, only one, the correct parse, out of 33 candidate linguistic descriptions successfully emerges from constraint testing.[2]

[1] In theory, the generate-and-test paradigm, i.e., generate linguistic descriptions and explicitly rule out illicit cases, can be contrasted with a constraint-based approach, i.e., one that emits only licit linguistic descriptions. From a computational complexity perspective, the constraint-based approach offers the possibility of avoiding the inefficiency inherent in overgenerating linguistic descriptions. Careful comparisons are necessary to verify that the bookkeeping necessary to maintain and manipulate constraint descriptions does not trump the inefficiency introduced by simple overgeneration. (See also footnote 8.) However, the author is not aware of any substantial purely constraint-based implementations of the Principle and Parameters approach.

[2] In practice, serial constraint testing can be organized to favor the early elimination of illicit candidates (to reduce implementational inefficiency). Note that most of the candidates in Fig. 10.1 are eliminated early on by the Case and Theta group of constraints. Only one other parse (in addition to the correct parse) makes any significant headway out of the Case and Theta group.

By contrast, recent theories in the minimalist program, e.g., in Chomsky (2001) and thereafter, seemingly leave little wiggle room for implementational variation or inefficiency. As will be described later in this paper, the assembly of licit syntactic objects has been made precise down to a stepwise (fundamental) operational level. Linguistic constraints such as those in the P&P approach must be recoded or reanalyzed in terms of the interaction between just two fundamental operations, bottom-up Merge (*external* (encoding syntactic composition) and *internal* (encoding sub-object displacement)) and Agreement (between recently merged objects that function as a *probe* locally seeking already merged sub-objects, termed *goals*). Both probe and goal objects have uninterpretable features that must be valued (and discharged) in the course of assembly.

Computationally speaking, it is assumed that Agree applies as early as possible in the derivation as Merge proceeds to build a composite syntactic object starting from a preselected lexical array of basic functional and substantive objects. The resulting syntactic object (SO) is *simple* in the sense that (artifactual) devices not originally present in the lexical array cannot be introduced or built in the course of assembly, e.g., no traces, movement chains, or indices can be added to syntactic descriptions. Moreover, not only must a probe "fire" as soon as it is merged, but probe-goal search is limited to local domains defined by Phase theory. A target with an uninterpretable feature beyond the range, measured in Phases, of an intended probe will cause narrow syntax computation to crash, i.e., terminate unsuccessfully.

We must make precise the notions of efficiency and complexity to be adopted in the paper. In computer science, there are various methods of measuring complexity and efficiency. For example, depending on the situation, it may be appropriate to adopt asymptotic analysis: if we assume the size of the data set, e.g., the length of the input sentence, may increase without bound, we can ignore constant factors and compare the fundamental rate of growth of the number of operations needed to compute an answer over a fixed grammar with respect to different algorithms. At a more abstract level, i.e., independent of particular grammar, it is sometimes possible to compare the generative power and computational complexity of different grammar formalisms: in general, we find there is a not unexpected trade-off between a formalism's expressive power (what is encodable) and computational complexity. For example, traditional formal language theory tells us there exists a sliding scale of complexity from finite-state automata (equivalently, regular grammars or regular expressions), through push-down (PD) automata (equivalently, context-free grammars), to nested-PD automata (equivalently, context-sensitive grammars). It has been suggested that natural languages may be characterized as nestling between context-free and context-sensitive bounds of expressive power. In particular, mildly context-sensitive grammars allow the expression of limited non-context-free dependencies found in natural language but retain polynomial time parse-ability (a desirable computational property denied to unrestricted context-sensitive grammars). It has been shown that certain mildly context-sensitive grammar formalisms are all formally equivalent in expressive power (Vijay-Shanker and Weir, 1994).

With respect to the P&P approach, we can compute the degree of overgeneration as an approximation to system and implementational (in)efficiency.[3] However, since the linguistic devices employed in the P&P approach are many and varied, a usefully limited characterization of its complexity is not practical. With respect to the minimalist program, a (simplified) mathematical characterization of the uninterpretable feature-checking mechanism can be found in Stabler's (1997) Minimalist Grammar (MG) formalism. This framework enjoys similar expressive power and complexity to the mildly context-sensitive grammars mentioned earlier. However, it remains to be seen whether MG can encode the full range of Agree and Phase theory constraints described in Chomsky (2001) and later publications. Furthermore, the idea of overgeneration, introduced above for the P&P approach, is not a useful measure of efficiency in the case of the Merge/Agree system. Much like a jigsaw puzzle, there is only one defined pattern per preselected lexical array. Thus, correctly implemented, there should be only one way in which those objects can be fit together.[4] The notion of efficiency therefore is only meaningful with respect to the details of the Agree operation. In particular, we can evaluate the extent or depth of the search for goals by probes and count the number of agree relations computed in the course of a derivation. This paper argues that adopting the mechanism of unification for feature matching will result in improved efficiency in these terms. And that minimizing the number of operations and localizing goal search as far as possible is in keeping with the spirit and goals of the minimalist program. But first, we make precise the computational details of the Merge/Agree system in the next section.

Background

Chomsky (2001) sets out the following operations as being basic to computation:

1. Merge "an *indispensible operation of a recursive system*" comes in two flavors:
   (a) *External Merge* takes two syntactic objects (SO) α, β and forms the set-merged SO: $\{\alpha, \beta\}$. For labeling, one of either α or β must project, i.e., label($\{\alpha, \beta\}$)=label(α) or label(β). An example, with $\alpha = [_V$ like] and β = $[_n$ mary], is given in Fig. 10.2a.
   (b) *Internal Merge* implements displacement. Selecting SOs α and γ, γ properly contained in α, form the aggregate set-merged SO $\{\alpha, \gamma\}$. Furthermore, label($\{\alpha, \gamma\}$)=label(α).

[3] Generally, overgeneration could be due to insufficient grammatical constraints or to the implementation itself. Given a sentence with a single licit parse, in the former case, extra parses will be generated. In the latter case, the system will produce multiple candidates, but only the correct parse will survive constraint checking.

[4] This is true up to a point. For example, from the same lexical array, one can assemble both *John likes Mary* and *Mary likes John* using the same sequence of steps, depending on which nominal, *John* or *Mary*, is selected to be the subject and direct object, respectively. See the sequence of operations shown in Fig. 10.3.

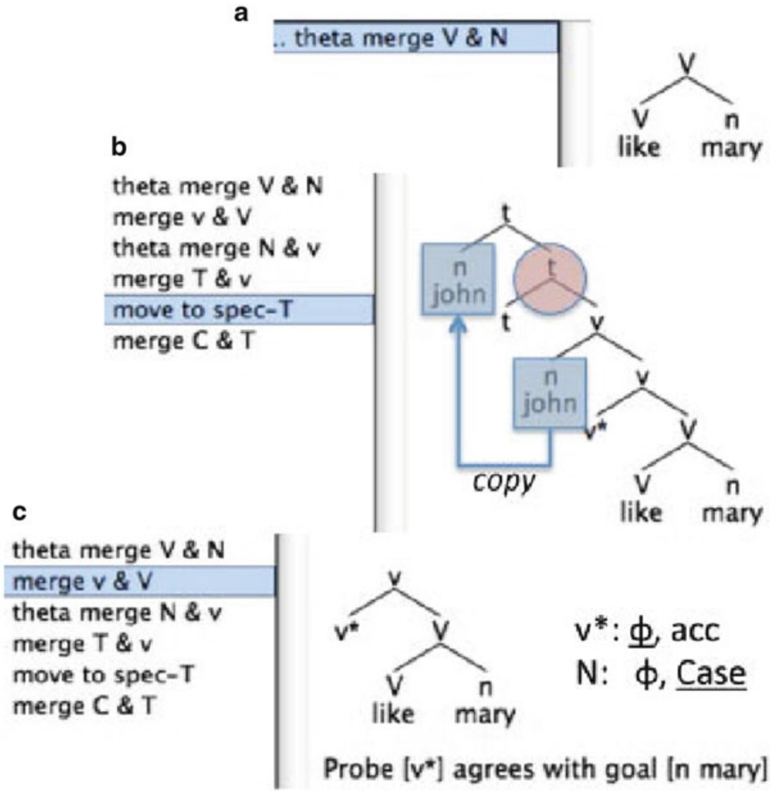


Fig. 10.2 Basic operations

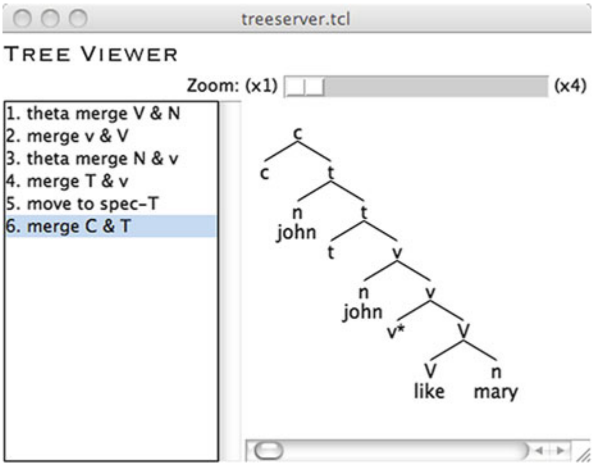
Figure 10.2b illustrates internal merge in the case of subject raising to tense (t). $\alpha = [_t\ t\ [_v\ [_n\ \textit{john}][_v\ v\ [_V\ [_V\ \textit{like}][n\ \textit{mary}]]]]]$, and subject $\gamma = [_n\ \textit{john}]$. Note there are two copies of γ. No trace is generated.

2. *Agree* obtains between an active probe SO α (*active* = still having uninterpretable features) and an active goal SO β. β should be the closest locally available goal in the c-command domain of α[5]. Features of α and β are matched, and uninterpretable features of both probe and goal are deleted.

Figure 10.2c illustrates probe-goal agreement for $\alpha = v^*$ (transitive v) and $\beta = [_n\ \textit{mary}]$. Goal $[_n\ \textit{mary}]$ has interpretable φ-features, i.e., person (3rd), number (singular), and grammatical gender (feminine) but uninterpretable Case. The probe v* has uninterpretable φ-features but can value accusative Case. v*'s uninterpretable φ-features are valued through matching with the goal's φ-features. The goal $[_n\ \textit{mary}]$ in turn receives accusative Case from v*. After Agree, neither probe nor goal remains active, as their uninterpretable features have been valued.

[5] The term "local" refers to the limits on search extent imposed by Phase theory on the unrestricted c-command domain.

Fig. 10.3 Parse of John likes Mary



A convergent derivation using Merge and Agree obtains when the initial selection of SOs has been fully utilized and no uninterpretable features remain at large.

Sentential structure is assumed to involve a sequence of (top-down) selection from an inventory of primitive functional elements including the c (complementizer), t (tense), and v (little v), V pairs. These functional categories come in various flavors, e.g., V may select for a thematic direct object, as with transitive and unaccusative verbs. V may be objectless, as with unergative verbs. V itself is directly selected by v, which may select for a thematic subject, as in the case of unergative or transitive verb (v*, which can value accusative Case), or be subjectless, as in the case of passives. Tense (t), which has an EPP or edge feature and requires a subject, may come in uninterpretable φ-complete (tensed) or φ-incomplete (infinitival) flavors; however, only φ-complete tense can value nominative Case.

An example of the parse produced for a simple transitive sentence *John likes Mary* is given on the right in Fig. 10.3. Although there are two copies of *John*, only the highest copy is pronounced.[6] The convergent Merge/Agree derivation sequence is summarized on the left and traced in detail in Fig. 10.4.[7]

In the basic case, as in Fig. 10.4, there is a simple, one-to-one correspondence between probes and goals, and this leaves no room for computational optimization. However, in more complex scenarios, Agree relations may simultaneously hold between a single probe and multiple goals: in particular, when either the probe or goal may be φ-incomplete. In these cases, as we will see in the next section, unification can improve the efficiency of the computational system.

[6] Inflectional morphology is not implemented in Fig. 10.3. We assume a procedure that spells out *likes* from t (tensed) + v* (φ: 3rd-sg-fem)+ V (*like*).

[7] In Fig. 10.4, the !*F* notation, where *F* is a feature, as in [n!case …], is used to indicate that a SO has a (currently) unvalued uninterpretable feature. In steps 1 and 3, *Mary* and *John* have unvalued Case, respectively. By steps 2 and 4, the uninterpretable Case in each case has been valued by probes v* and t, respectively.

```
0. lexical array              [c][t][n john][v*][V like][n mary]
1. theta merge V & n          [V[V like][n!case mary]]
2. merge v & V                [v[v*][V[V like][n mary]]]
Probe [v*] agrees with goal [n mary]
3. theta merge n & v          [v[n!case john][v[v*][V[V like][n mary]]]]
4. merge t & v                [t[t][v[n john][v[v*][V[V like][n mary]]]]]
Probe [t] agrees with goal [n john]
5. move to spec-T             [t[n john][t[t][v[n john][v[v*][V[V like][n mary]]]]]]
6. merge C & T                [c[c][t[n john][t[t][v[n john][v[v*][V[V like][n mary]]]]]]]
```

Fig. 10.4 Merge/Agree sequence for John likes Mary

Unification and Computation

We will illustrate the advantage of unification on the following examples from Chomsky (2001:4b–c):

3. (a) There are likely to be awarded several prizes
   (b) Several prizes are likely to be awarded.
4. (a) We expect there to be awarded several prizes.
   (b) We expect several prizes to be awarded.

Examples (3.a–b) and (4.a–b) contain the same passivized embedded clause *to be awarded several prizes*. The matrix predicate is a *raising* predicate, *be-likely* in (3.a–b) and an *Exceptional Case Marking* (ECM) verb *expect* in (4.a–b). Within the embedded clause, passive *v* is non-Case-valuing, and the direct object (DO) is *several prizes*. The adjectival past participle *-ed* (PRT) is assumed to be φ-incomplete and have uninterpretable Case (unrealized in English).[8] The embedded clause includes a subject position, overtly occupied by a copy of *several prizes* in (4.b) and by pleonastic *there* in (4.a); the embedded subject position is covert in (3.a–b). Finally, tense is φ-incomplete (infinitival) and cannot value Case either.

Thus *several prizes* must get Case from the tense that heads the matrix clause in (3.a–b) and matrix v* in (4.a–b), respectively. In fact, since matrix tense and matrix v* are the only φ-complete probes available, it (namely, tense and v*) must participate in multiple agree relations with various goals in the embedded clause (including the DO and PRT) in order for computation to converge.

Consider the derivation of (3.b), illustrated in Fig. 10.5. The DO *several prizes* raises to embedded tense and then to the highest position at matrix tense where it is pronounced. Along the way, the following Agree relations are computed: at position (1), adjectival *-ed* (PRT) undergoes feature matching with the DO. The PRT's uninterpretable φ-features are valued by the DO, which possesses interpretable φ-features. However, both PRT and DO lack Case. At position (2), the embedded tense (tdef = infinitival) agrees with the DO and its uninterpretable φ-features are valued. However, tdef is φ-incomplete and cannot value Case. Since tense has an

[8] PRT is assumed by Chomsky (2001) to have uninterpretable number and gender φ-features.

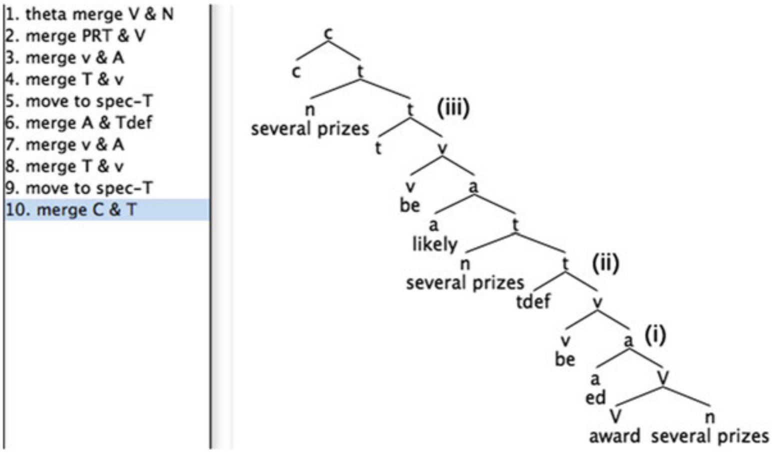


Fig. 10.5 Several prizes are likely to be awarded

EPP (or edge) feature, the DO raises. At position (3), matrix tense (t) probes and finds the raised DO. Since t is φ-complete, it values Case for the DO and its uninterpretable φ-features are valued. In Chomsky (2001), in order for the derivation not to crash, t must continue to probe beyond the raised DO to value PRT's Case feature as well. Thus t is in a multiple agreement relation. However, if feature matching is implemented using unification, the matrix tense's secondary search can be avoided altogether, thus making for more local (and efficient) computation. Suppose unification is adopted. Then, at stage (1), the PRTs and DOs still unvalued uninterpretable Case features can be unified together.[9] Subsequently, in stage (3), PRT's Case feature will be valued at the same time as the DO's, without the search extension of Chomsky (2001).

Consider next the derivation of (3.a), illustrated in Fig. 10.6. The sequence of operations, stages (1–3), largely parallels that of (3.b) discussed previously, except that the initial conditions will be different: in (3.a), the initial lexical array contains pleonastic *there*.[10] Instead of the DO *several prizes* raising to embedded tense as in (3.b), *there* insertion via External Merge to embedded tense is triggered in (3.a).[11] At stage (3), matrix tense (t) will probe and encounter (first) *there* at embedded tense. Both *t* and *there* have uninterpretable φ-features. Thus, unlike (3.b), matrix *t* must continue to probe all the way down until it encounters the in situ DO *several prizes*, whereupon its uninterpretable φ-features, and the DO's uninterpretable Case feature, will be valued.

[9] Implementation note: more precisely, an unvalued uninterpretable feature will have an uninstantiated logical variable as its value. Thus, when two unvalued uninterpretable features are unified, their values are represented by the same variable.

[10] Chomsky (2001) assumes *there* is φ-incomplete, containing only a person feature.

[11] In this implementation, it is assumed that External Merge is preferred (where available) to Internal Merge.

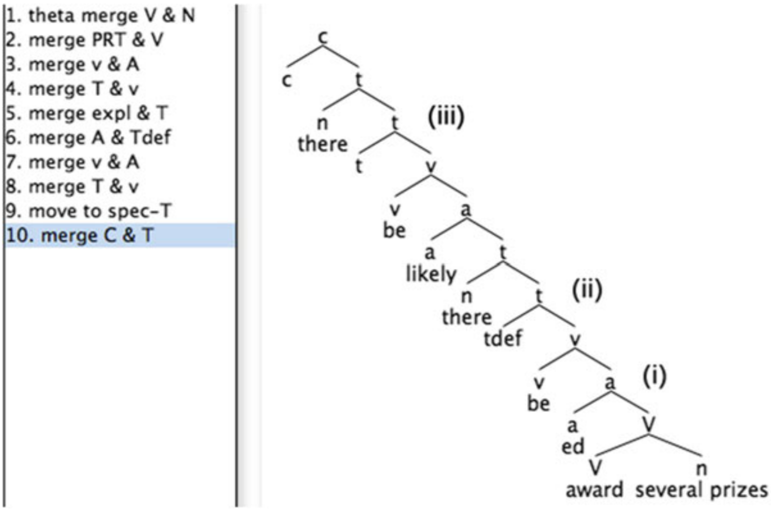


Fig. 10.6 There are likely to be awarded several prizes

In the implementation, at this point several other operations must also complete for the derivation to converge properly. Prior feature matching by unification will result in adjectival -*ed*'s (PRT) Case feature being valued at the same time. Pleonastic *there*'s uninterpretable φ-feature will be also valued (as its φ-feature was already unified with matrix *t*'s). Without unification, feature matching must contain more steps. In particular, *there* must be a probe and directly enter into an agree relation with *several prizes*, after agreeing with (and probing beyond) PRT. Therefore, unification provides for an efficiency gain since there will be fewer relations computed, despite matrix tense having to reach all the way down to find the in situ DO.

Not only does unification requires less search to value the same uninterpretable features, it also simplifies bookkeeping from the viewpoint of recursive computation. As mentioned earlier, it is assumed that Agree applies as early as possible in the derivation as Merge proceeds. Once a probe has agreed with a goal within a (still partially assembled) syntactic object, there is no need to revisit (or later revive) that now-established relationship as syntactic object building proceeds in compositional fashion beyond initial Merge of the probe. In other words, there is no need to re-descend and check sub-object features: still unvalued uninterpretable features (linked earlier by unification) will be automatically valued at the earliest opportunity.

Consider now the derivation of (4.b), illustrated in Fig. 10.7. Stages (1) and (2) are similar to that of (3.b), illustrated previously in Fig. 10.5. However, unlike (3.b), stage (3) for (4.b) makes use of v* introduced along with the ECM verb *expect*. The probe v*, which values accusative Case, agrees with the DO *several prizes* previously raised to embedded tense.[12] In Chomsky (2001), v* must continue to

[12] Example pairs (3.a–b) and (4.a–b) differ with respect to the Case assigned to *several prizes*, nominative and accusative by matrix tense and v*, respectively. This difference is not manifested here. However, cf. *we expect him/*her to be nominated*.

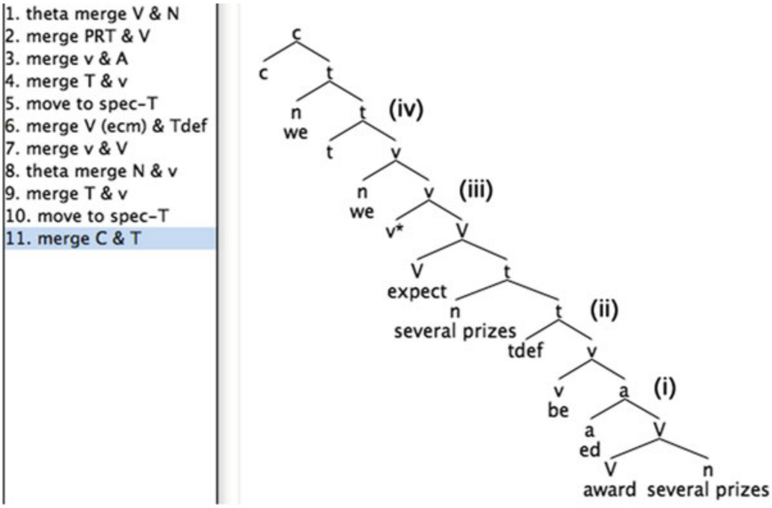


Fig. 10.7 We expect several prizes to be awarded

probe past the DO and value the Case feature of the PRT -*ed*. Assuming unification has already taken place in stage (1) between the unvalued uninterpretable Case features for both PRT and DO, there is no need (in our implementation) to extend the domain of search for v* past embedded tense. Thus unification has a locality advantage with respect to probe-goal search. To complete the derivation, we note that in stage (4), matrix tense (t) probes and values the nominative Case for the matrix subject pronoun *we*.

Finally, in the case of (4.a), the derivation parallels that of (4.b) described above, with the initial conditions changed by the presence of pleonastic *there*, as with (3.a). At stage (4) (not shown), v* agrees with *there* and must probe all the way down to have its φ-features valued by in situ *several prizes*. Similar efficiency gains to (3.a) are realized here: in particular, *there* does not need to probe *several prizes* in this implementation.

References

Chomsky, N. A. (1981). *Lectures in government and binding*. Dordrecht: Foris.

Chomsky, N. A. (2001). Derivation by phase. In M. Kenstowicz (Ed.), *Ken Hale: A life in language* (pp. 1–52). Cambridge, MA: MIT Press.

Fong, S. (1991). *Computational Properties of Principle-Based Grammatical Theories*. Ph.D thesis, Artificial Intelligence Laboratory, MIT.

Stabler, E. P. (1997). Derivational minimalism. In C. Retoré (Ed.), *Logical aspects of computational linguistics (LACL '96), Lecture Notes in Artificial Intelligence, 1328* (pp. 68–95). Berlin: Springer.

Vijay-Shanker, K., & Weir, D. J. (1994). The equivalence of four extensions of context-free grammars. *Mathematical Systems Theory, 27*, 27–51.

Chapter 11
Recursion in Generative Grammar

Robert Freidin

Introduction

The topic of recursion first appears in the writings on modern generative grammar on virtually the first page of the first work, in the second footnote to Noam Chomsky's *The Morphophonemics of Modern Hebrew* (henceforth, *MMH*).[1] It arises in reference to the criteria of adequacy a grammar of a language must meet in the case of linguistic analysis, identified as "requirements of simplicity, economy, compactness, etc." *MMH* characterizes the linguistic analysis of a language L as a process that determines "the set of 'grammatical' or 'significant' sentences of L" and goes on to relate this to recursion: "or, in other words, it is a process of converting an open set of sentences—the linguist's incomplete and in general expandable corpus into a closed set—the set of grammatical sentences and of characterizing this latter set in some interesting way" (*MMH*:1). The footnote on "closed" qualifies the set as "not necessarily finite" and goes on to note that "the resulting grammar will in general contain a recursive specification of a denumerable set of sentences" (*MMH*:67).

In *MMH*, the only actual source of recursive linguistic structure mentioned comes in the form of an illustration of a phrase structure grammar based entirely on rewrite rules; the first rule of which is (1) where "Connective" represents a coordinating conjunction (*and*, *or*, *but*, etc.) and the angle brackets enclose optional elements.

[1] The first version of *MMH*, written in 1949, served as Chomsky's senior undergraduate thesis at the University of Pennsylvania. The manuscript was expanded as his 1951 master's thesis and published by Garland Publishing in 1979 as part of their Outstanding Dissertations in Linguistics series.

The term "recursion" has been used in various ways in different domains of inquiry. See Fitch (2010) for useful discussion.

R. Freidin (✉)
Princeton University, Princeton, NJ, USA
e-mail: freidin@Princeton.EDU

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_11, © Springer Science+Business Media New York 2014

(1) Sentence → Elementary sentence <Connective + Sentence>

(1) is clearly a recursive rule given that a part of its output, the category "Sentence," serves as the input to the rule itself. As *MMH* notes, (1) must reapply until "Sentence" is eliminated. Thus, the set of sentences that can be generated by a grammar containing (1) will be infinite because it will allow infinite use of the finite rules it contains, thereby giving concrete form to von Humboldt's insight that language makes infinite use of finite means.

Chomsky (1967) notes that the property of infinity in language follows from the simple fact that it is in general impossible to specify any sentence as the longest sentence of the language in question.[2]

> *It is, first of all, quite clear that the set of paired phonetic and semantic representations generated by the grammar will be infinite. There is no human language in which it is possible, in fact or in principle to specify a certain sentence as the longest sentence meaningful in this language* (p. 400).

In English, for example, any candidate longest sentence can easily be extended by embedding it in a longer sentence beginning *I believe that* or coordinating it with another sentence, both instances of recursive embedding of clause structures in other clause structures.[3] Chomsky goes on to point out that the unbounded character of language results from grammatical devices that make it possible and furthermore that normal language use crucially depends on this.

> *The grammar of any language contains devices that make it possible to form sentences of arbitrary complexity, each with its intrinsic semantic interpretation. It is important to realize that this is no mere logical nicety. The normal use of language relies in an essential way on this unboundedness, on the fact that language contains devices for generating sentences of arbitrary complexity* (p. 400).

The grammatical mechanisms for generating sentences of arbitrary complexity are going to be recursive devices. As noted in *Syntactic Structures* (Chomsky (1957)), if a grammar contains "recursive devices of some sort, it will produce infinitely many sentences" but "if a grammar does not have recursive devices… it will be prohibitively complex" (p. 24).[4] "In general, the assumption that languages are infinite is made in order to simplify the description of these languages" (pp. 23–24).

In over half a century of research on generative grammar that followed, the conceptual changes in the study of language have radically altered the

[2] See also Hauser, Chomsky, and Fitch (2002) for a more recent statement of this point. Whether the Brazilian language Pirahã is an actual counterexample remains controversial (see Everett, 2005, Nevins, Pesetsky, & Rodriguez 2007, and the following interchange). See Chomsky (2013b) for further comment.

[3] Note that the critique of the "infinitude claim" in Pullum and Scholz (2010) does not address this argument. See also the next footnote.

[4] Repeating points made in Chomsky (1956, §2). In their conclusion, Pullum and Scholz (2010) make a similar point about grammars of finite languages (presumably without recursive devices)—namely, that they would include "several highly arbitrary and numerical conditions" (quoting Harris, 1957, p. 208), an unwanted result. They claim that "ideally grammars should be stated in a way that insists neither on finitude or on infinitude" (p. 133). But this seems to miss the point that unboundedness is a consequence of recursive devices and that to prohibit these devices from grammars inevitably yields grammars with arbitrary conditions that will be "prohibitively complex."

perspective in which recursion remains a key topic. Perhaps the most fundamental shift in focus involves the conception of a language as a system of knowledge in the mind of the speaker—what Chomsky (1986) (and elsewhere) designates as an I-language, where I stands for "internal" and hence "individual."[5] These systems of knowledge have been modeled as a computational system plus a lexicon that generates the unbounded linguistic expressions of the language as paired representations for sound (Phonetic Form (PF)) and meaning (called Logical Form (LF) and abstracting away from knowledge of the world, beliefs, and much else).[6] This constitutes a shift in focus from the description of utterances of a language to the underlying grammar, its properties, and function.[7] Thus the primary goal is no longer a description of the external forms of language as represented in PF.

The current model of the computational system consists of two minimal operations, Merge and Delete. Merge is a structure creating operation that combines two[8] syntactic objects—lexical items or syntactic objects formed from them—to form a new one. Consider, for example, the derivation of (2) under Merge.

(2) We have made a decision.

The noun *decision* and the determiner *a* are combined to form a phrase as indicated by the brackets in (3), and that syntactic object is then merged with the verb *made* to form another phrase, and so on as indicated in (3).

(3) [we [have [made [a decision]]]]

The constituents constructed would be labeled by the syntactic category of their heads, in effect projecting the label of the head onto the constituent formed by Merge. Thus (3) would be labeled as in (4), where T stands for "tense," the finite element that heads a finite clause—in this case the finite perfective auxiliary *have*.[9]

---

[5] This is in sharp contrast to the notion of a language as a corpus of linguistic expressions (finite or infinite) that can be characterized independently of the mind of the speaker (what Chomsky (1986) (and elsewhere) characterizes as an E-language (E for "external"). Note that this approach privileges the external forms of language, emphasizing the properties of Phonetic Form (PF), which can be highly idiosyncratic, over the properties of Logical Form (LF), which are virtually entirely covert and apparently general crosslinguistically. The notion of E-language raises many problems of definition and may not be coherent (see Chomsky (1986) for discussion and Freidin (2012) for a discussion of problems that arise for a characterization of an "English language").

[6] Chomsky (1993) conjectures that there may be only one computational system for human language and one lexicon, setting aside the idiosyncratic properties of PF crosslinguistically—starting with the phonetic labels of lexical items, which are essentially historical accidents.

[7] For a discussion of the conceptual shifts in modern generative grammar, see Freidin (1995, 2012, 2013)—including the references cited in these chapters.

[8] Possibly more in the case of coordination, see Freidin (2012) for discussion.

[9] For a discussion on whether the labeling function (projection) is part of Merge or a separate operation, see Chomsky (2013a). For the remainder of this note, it will be assumed that Merge incorporates both a labeling function and a grouping function.

(4) $[_T$ we $[_T$ have $[_V$ made $[_N$ a decision]]]]

The outer bracket subscripted T identifies the maximal phrasal projection of *have*, while the inner T bracket marks an intermediate phrasal projection. In this derivation, all applications of Merge take two independent syntactic objects and combine them into a single construct, what is referred to as External Merge (EM).

EM is one option for the application of Merge. The other applies when one of the syntactic objects merged is contained in the other, called Internal Merge (IM).[10] This yields the displacement property of human language, where constituents pronounced in one syntactic position are interpreted as occupying another. Consider, for example, a simple passive construction in English related to (2).

(5) A decision has been made.

At LF, the syntactic subject of (5) is interpreted as the object of the verb *made*, in this case the passive past participle as opposed to the perfective participle in (2). Thus the derivation of (5) involves merging the verb and its direct object *a decision* as in the derivation of (2), followed by the merger of that verb phrase with the past participle form of the passive auxiliary *been*, and then the merger of the verb phrase headed by *been* with the finite perfective auxiliary *has*, yielding (6).

(6) $[_T$ has $[_V$ been $[_V$ made $[_N$ a decision]]]]

The next step in the derivation of (5) involves an application of IM, where the NP *a decision* is merged with (6), yielding (7).

(7) $[_T$ $[_N$ a decision] $[_T$ has $[_V$ been $[_V$ made $[_N$ a decision]]]]]

Note that the merger of the NP *a decision* with (6) creates the subject position (5) in the same way that the merger of *we* in the derivation of (2) creates the subject position of that sentence. (7) contains two copies of the NP *a decision*, forming a "chain" that is relevant to interpretation at the LF interface. At PF, such chains are assumed to be illegible and therefore must be eliminated by deletion of extraneous copies, chain reduction that linearizes the phonetic representation, given that a single syntactic object cannot be pronounced in more than one position in PF.[11] Thus at PF, (7)

[10] See Chomsky (2013b) for further discussion.

[11] The operation Delete also creates ellipsis constructions where the deleted material constitutes a repetition rather than a copy. For example, at LF the following examples have the same structure and hence the same interpretation:

(a) Mary has bought a new piano and Fred has bought a new piano too.
(b) Mary has bought a new piano and Fred has ~~bought a new piano~~ too.

At PF, the VP *bought a new piano* in the second conjunct is not pronounced, though it is interpreted at LF as a separate action from the action expressed in the identical VP of the first conjunct. Thus Delete affects PF in two ways, obligatorily deleting multiple copies and optionally deleting repetitions. Given that deletion is restricted to the derivation of PF and does not affect LF, this observation motivates derivations via Merge where the derivation splits into two paths, one to LF and the other to PF, a point (or points) called Spell-Out. From Full Interpretation, it follows that is mapped onto (8) where the deleted material is indicated with a strikethrough, matching (5).

(8) $[_{T}$ $[_{N}$ a decision] $[_{T}$ has $[_{V}$ been $[_{V}$ made ~~$[_{N}$ a decision]~~]]]]

This follows from Full Interpretation (Chomsky, 1986), which prohibits superfluous symbols in interface representations. As Chomsky has often noted (cf. most recently Chomsky, 2013a), prohibiting the IM option would require stipulation—pretty obviously unmotivated, one might add, given the prevalence of displacement phenomena in human language.

In addition to the assumption of binary Merge (but see footnote 8), the grouping function of Merge is further constrained by a No Tampering Condition (NTC, Chomsky, 2007, 2008) which prohibits Merge from altering the internal constituent structure of syntactic objects, insuring that Merge applies only to the roots of the syntactic objects combined. In effect, the NTC imposes bottom-up strictly cyclic derivations.[12] Projection (the labeling operation) is limited to properties that exist in the lexicon (e.g., syntactic category features) by the Inclusiveness Condition (Chomsky, 1995b) and moreover to the label of one of the two constituents merged, the head.

The operation Merge and its application approach an ideal of minimal computation, directly addressing the criteria of adequacy that Chomsky identified in *MMH*: "requirements of simplicity, economy, compactness, etc." This is obvious, especially when we compare Merge to the machinery of earlier generative models of the computational system. Consider, for example, the computational machinery required for the derivation of simple passive constructions in Chomsky (1957).[13] In that model, the computational system consists of a phrase structure grammar that generates a set of basic structures, those that underlie simple active affirmative indicative sentences, plus a set of transformations that map these structures onto complex and compound structures, as well as sentences involving passive voice, negation, and non-indicative forms (e.g., interrogatives and imperatives). The phrase structure rules involved in the derivation of (2) are given in (9).

(9)
1. Sentence → NP + VP
2. VP → Verb + NP
3. Verb → Aux + V
4. Aux → Tense (modal) (*have* + *en*) (*be* + *ing*)
5. NP → $\{NP_{sing}, NP_{plural}\}$
6. NP_{sing} → T+N+ø
7. NP_{plural} → T+N+*s*

at Spell-Out phonetic features and semantic features are separated, the former restricted to the PF part of the derivation and the latter to LF part.

[12] It is worth noting that the syntactic cycle was the first bottom-up syntactic principle proposed (Chomsky, 1965). For discussion on how the empirical effects of cyclic derivation might be derived from other considerations, see Freidin (1978, 1999).

[13] The formulation of phrase structure rules in Chomsky (1957) constitutes a significant simplification of the formulations in *The Logical Structure of Linguistic Theory* (Chomsky (1955–6)). However, the essential character of phrase structure rules remains the same.

8. T → *the*, *a*
9. N → *man*, *ball*, *committee*, *decision*, *we*, etc.
10. V → *hit*, *take*, *make*, *walk*, *read*, etc.

(9) consists of two subtypes of phrase structure rule, phrasal rules (9.1–7) that specify the phrase structure of sentences and lexical rules (9.8–10) that insert lexical items into phrase structures. Rule 9.4 actually performs both functions. Lexical rules connect the computational system to the lexicon.

The derivation of (2) via (9) also involves obligatory transformational operations that map an underlying structure onto the PF of (2), including one that attaches the past participle affix *en* onto the verb *make* resulting in the PF *made* and another that converts the uninflected perfective auxiliary *have* into an inflected auxiliary *has*. The derivation of the passive (5) involves a passive transformation (10) that maps the underlying structure of (5) onto its PF where the object of the verb is pronounced as the subject of the sentence and a passive auxiliary is inserted into the structure.

(10) *Passive*—optional:
Structural analysis: *NP—Aux—V—NP*
Structural change: $X_1 - X_2 - X_3 - X_4 \rightarrow X_4 - X_2 + be + en - X_3 - by + X_1$

(10) identifies four consecutive constituents and rearranges two of them as well as inserting two lexical elements (the passive auxiliary *be* and its associated passive past participle affix *-en* plus the passive preposition *by*). The derivation of truncated passives (i.e., without a passive *by*-phrase) as in (5) involves another transformation that optionally deletes a *by*-phrase whose interpretation is recoverable in some way from the interpretation of the passive main verb (e.g., *by someone*).

To the extent that both phrase structure rules and transformations are formulated in terms of lexical items specific to particular languages, the formulation of the computational system is language specific. Chomsky (1965) goes a long way towards eliminating the language-specific character of phrase structure grammar by replacing complex lexical phrase structure rules with a general transformational operation that substitutes partially specified terminal elements in phrase structure derivations with fully specified lexical items where the former are nondistinct from the latter. This move was an important step in separating the lexicon from the computational system, allowing for a general formulation of the latter that makes no reference to particular languages.

The next step in eliminating language-particular formulations of the computational system comes with the replacement of phrase structure rules, which stipulate the linear order of constituents that is specific to particular languages, with Merge (Chomsky, 1995a). The general assumption is that Merge creates hierarchical structure that is linearized at PF based on language-particular requirements (e.g., the head parameter which determines head-complement order). Thus at LF, languages like English and Japanese, which have very different PF linear order, have essentially the same LF structures—abstracting away from differences in lexical properties.

The formulation of Merge is arguably the minimal specification of a recursive operation that generates fundamental linguistic structure. It subsumes in a single minimal operation the structure-building character of phrase structure rules, the displacement property as formulated in movement transformations, and the connection between the lexicon and the computational system (the function of lexical insertion). Given that every human language has phrase structure, Merge is obviously part of the grammar of any language, therefore universal in the Greenbergian sense. Given that it cannot be learned from overt properties of language (i.e., those that come from nurture as opposed to nature), Merge must be part of human genetic endowment, hence universal across the species and therefore part of the initial state of the language faculty (what has been called universal grammar).[14] It is a stunning achievement of pursuing a minimalist program as initially formulated in Chomsky (1993).[15]

From the perspective of Merge, the computational system is general crosslinguistically, and the lexicon is the sole source of crosslinguistic variation. This leads to Chomsky's strong minimalist thesis (11) (see Chomsky (2010) for discussion).

(11) Interfaces + Merge = Language

(11) has been restated as (12), the title of a volume edited by Sauerland and Gärtner (2007), with the addition of a crucial question mark "because there is so much that is not understood" (Chomsky, 2013b).

(12) Interfaces + Recursion = Language?

In effect, the essence of language is reduced to minimal recursion in the computational system and whatever constraints are imposed on its operation from the cognitive components that process the two interface representations, PF and LF. Whether this extremely abstract equation can be proven in some substantial way remains an open question.

There are however two considerations that might bear on this version of a strong minimalist thesis. One concerns that fact that the computational system must include an operation Delete for chain reduction and ellipsis that both affect PF (but crucially not LF) representations. (12) might be viable if we exclude externalization of language (i.e., PF) from the definition of "language." Whether this is the right approach requires strong motivation (but cf. Chomsky (2013b) for discussion of externalization).

Another consideration concerns the issue of the unbounded character of language. Merge provides the mechanism for generating recursive structures, which establish this unbounded character. Merge is necessary, but not necessarily sufficient. To some extent, recursive structure is grounded in the lexicon. For example, verbs and adjectives that take clausal complements ground clausal recursion as a lexical

[14] In this regard, it is worth noting that it took almost four decades of research for modern generative grammar to discover that Merge is that driving force of linguistic structure.

[15] For a detailed history of the generative analysis of passive constructions, see Freidin (1995); for a brief history of generative grammar, see Freidin (2011); and for a sketch of Chomsky's contribution to linguistics, see Freidin (2013).

property. If some human language lacked such lexical items or realized such complements as nominalizations rather than clauses with verbs (cf. Everett, 2005), some part of clausal embedding could be eliminated. Coordination, which also yields recursive embedding, is linked to the lexicon via conjunctions. So complementation and coordination are phenomena potentially involving recursive structure that are linked to the lexicon. However, the phenomenon of modification (e.g., via relative clauses or prepositional phrases that function as adjuncts to nouns) is not a property of any particular lexical item. It requires recursive Merge, but its existence in language does not apparently follow from Merge alone. If so, then perhaps the unbounded character of language derives from something more than the computational system plus a lexicon.[16]

References

Chomsky, N. (1955–6). *The logical structure of linguistic theory*. Published 1975, New York: Plenum Press.

Chomsky, N. (1956). Three models for the description of language. *I.R.E. Transactions on Information Theory*, *IT-2*.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1967). The formal nature of language. In E. Lenneberg (Ed.), *Biological foundations of language*. New York: Wiley.

Chomsky, N. (1979). *The morphophonemics of modern Hebrew*. New York: Garland.

Chomsky, N. (1986). *Knowledge of language*. New York: Praeger.

Chomsky, N. (1993). A minimalist program for linguistic theory. In K. Hale & S. J. Keyser (Eds.), *The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger*. Cambridge, MA: MIT Press.

Chomsky, N. (1995a). Bare phrase structure. In H. Campos & P. Kempchinsky (Eds.), *Evolution and revolution in linguistic theory*. Washington, DC: Georgetown University Press.

Chomsky, N. (1995b). *The minimalist program*. Cambridge, MA: MIT Press.

Chomsky, N. (2007). Approaching UG from below. In U. Sauerland & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*. Berlin: Mouton de Gruyter.

Chomsky, N. (2008). On phases. In R. Freidin, C. Otero, & M.-L. Zubizaretta (Eds.), *Foundational issues in linguistic theory*. Cambridge, MA: MIT Press.

Chomsky, N. (2010). Some simple evo devo theses: How might they be true for language? In R. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language: Biolinguistic perspectives*. Cambridge: Cambridge University Press.

Chomsky, N. (2013a). Problems of projection. *Lingua, 130*, 33–49.

Chomsky, N. (2013b). Minimal recursion: Exploring the prospects. In T. Roeper & M. J. Speas (Eds.), *Recursion: Complexity in cognition*. New York: Springer.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã. *Current Anthropology, 46*, 621–646.

[16] This discussion develops some ideas that were presented at the conference on language and recursion at the University of Mons, March 2010.

Fitch, W. T. (2010). Three meanings of "recursion": Key distinctions for biolinguistics. In R. Larson, V. Déprez, & H. Yamakido (Eds.), *The evolution of human language: Biolinguistic perspectives*. Cambridge: Cambridge University Press.
Freidin, R. (1978). Cyclicity and the theory of grammar. *Linguistic Inquiry, 9*, 519–549.
Freidin, R. (1995). Conceptual shifts in the science of grammar: 1951–1992. In C. Otero (Ed.), *Noam Chomsky: Critical assessments* (Vol. 1). London: Routledge.
Freidin, R. (1999). Cyclicity and minimalism. In S. D. Epstein & N. Hornstein (Eds.), *Working minimalism*. Cambridge, MA: MIT Press.
Freidin, R. (2011). A brief history of generative grammar. In G. Russell & D. Graff (Eds.), *The Routledge companion to the philosophy of language*. London: Routledge.
Freidin, R. (2012). *Syntax: Basic concepts and applications*. Cambridge: Cambridge University Press.
Freidin, R. (2013). Noam Chomsky's contribution to linguistics: A sketch. In K. Allen (Ed.), *The Oxford handbook of the history of linguistics*. Oxford: Oxford University Press.
Harris, Z. (1957). Co-occurrence and transformation in linguistic structure. *Language, 33*, 283–340.
Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.
Nevins, A., Pesetsky, D., & Rodriguez, C. (2007). Pirahã exceptionality: A reassessment. *Language, 85*, 355–404.
Pullum, G., & Scholz, B. (2010). Recursion and the infinitude claim. In H. van der Hulst (Ed.), *Recursion in human language*. Berlin: De Gruyter Mouton.
Sauerland, U., & Gärtner, H.-M. (Eds.). (2007). *Interfaces + recursion = language?: Chomsky's minimalism and the view from syntax-semantics*. Berlin: Mouton de Gruyter.

Chapter 12
Computational Language Related to Recursion, Incursion and Fractal

Daniel M. Dubois

Introduction

This paper is not a review paper in the field of language and recursion. The excellent paper written by Francis Lowenthal (2010) is a good introduction to the scientific works in the field of the relation between language and recursion, with many useful references.

I present in this paper some mathematical and computational tools in view of analysing if the languages have a fractal geometry generated by recursion. At the end of the paper, the concept of incursion is presented as an extension to the recursion.

The first section deals with a presentation of the recursion related to Turing machine, formal language and fractal. Wittgenstein made hard critics about the formal language after having set out a logically formal language. Noam Chomsky is a pioneer in the application of recursion and Turing machine in computational linguistic. After a comparison of computer and programming languages, some reflections are given about the natural languages related to recursion and fractal.

The second section presents some mathematical formulas in computational linguistics. We survey the Zipf and the Mandelbrot scaling laws of languages, the Shannon–Weaver entropy, the Kullback–Leibler relative entropy, deduction of invariants for cyclic time series from the Shannon–Weaver entropy in natural language processing and the Rényi generalized fractal dimensions. Finally, this section presents the research work of Marcel Ausloos to fractal, multifractal and correlation dimensions of time series of texts.

D.M. Dubois (✉)
HEC-ULg, Management School of the University of Liege, Liege, Belgium

ASBL CHAOS, Centre for Hyperincursion and Anticipation in Ordered Systems,
Institute of Mathematics B37, Grande Traverse 12 BE 4000, Liege, Belgium
e-mail: Daniel.Dubois@ulg.ac.be

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*, 149
DOI 10.1007/978-1-4614-9414-0_12, © Springer Science+Business Media New York 2014

The third section introduces the computational fractal geometry with recursion and incursion, in giving details about the box counting method for computing the dimension of the Euclidean space and the fractal geometry of automata. Then, two generators of fractal automata with recursion and incursion are presented.

Recursion Related to Turing Machine, Formal Language and Fractal

This section deals with the introduction of formal logic in language and the relation with the Turing machine. By definition, recursion is the process of repeating objects by self-similarity and has a variety of meanings specific to a variety of disciplines ranging from linguistics to logic. Noam Chomsky (Hauser, Chomsky, & Fitch, 2002) suggested the idea that recursion is an essential property of human language. But the natural languages are not yet used as a communication means between man and computer. The artificial languages of computer are processed by recursion in the Turing machine.

By definition, a fractal object shows the property of self-similarity, like a hologram, and is generated by a process that undergoes iteration based on recursion.

Development of Formal Language by Wittgenstein

The attempt to set out a logically perfect language, building on Russell's work, was proposed in the Tractatus Logico-Philosophicus by Wittgenstein (1921) has become known as the formal language. But, Wittgenstein repudiated many of his own earlier views and made hard critics about the formal language, in his book Philosophical Investigations (Wittgenstein, 1953), in pointing out that the practice of human language is more complex than the simplified views of language that have been held by those who seek to explain or simulate human language by means of a formal system. It would be a disastrous mistake, according to Wittgenstein, to see language as being in any way analogous to formal logic. For him, language cannot be restricted to be a property of the brain but must be included in the whole human social system. Wittgenstein extends his critics to the conceptual approach of the mind and consciousness, as included in a social system. Wittgenstein has thought about the introduction of a dictionary in the formal language, but he rejected this idea because of the infinite regress of the definitions of words in a self-referential way.

Turing Machine and Formal Language of Noam Chomsky

There are fundamental questions on the relation between the brain and Turing machine and between the natural language and formal language. Any real computer is based on the model of the Turing machine (Turing, 1936, 1937). A Turing machine, described by Alan Turing in 1936, is a basic abstract symbol-manipulating device that, despite its simplicity, can be adapted to simulate the logic of any computer algorithm. The Turing machine mathematically models a machine that mechanically operates on a tape on which symbols are written, which it can read and write one at a time using a tape head; operation is fully determined by a finite set of elementary instructions. Any computing machines use artificial languages, based on formal language. A formal language is a set of words, given by finite strings of letters, or symbols. The inventory from which these letters are taken is called the alphabet over which the language is defined. A formal language is defined by means of a formal grammar, and based on purely syntactical rules, so there is not necessarily any meaning associated with it. A formal grammar defines (or generates) a formal language, which is a (usually infinite) set of finite-length sequences of symbols (i.e. strings) that may be constructed by applying production rules to another sequence of symbols which initially contains just the start symbol. A rule may be applied to a sequence of symbols by replacing an occurrence of the symbols on the left-hand side of the rule with those that appear on the right-hand side. A sequence of rule applications is called a derivation. Such a grammar defines the formal language: all words consisting solely of terminal symbols which can be reached by a derivation from the start symbol. Noam Chomsky (1956) studied models for the description of language. The grammar needed to specify a programming language can be classified by its position in the following hierarchy of grammars, Gn, described by Noam Chomsky (1959):

G0. Unrestricted grammars: recursively enumerable languages, which generate exactly all languages that can be recognized by a Turing machine
G1. Context-sensitive grammars: context-sensitive languages, which a somewhat more accurate model of computers that actually exist than a Turing machine, whose definition assumes unlimited tape
G2. Context-free grammars: context-free languages, which are the theoretical basis for the syntax of most programming languages
G3. Regular grammars: regular languages, which are used to define search patterns and the lexical structure of programming languages

Let us note that the set of grammars corresponding to recursive languages, which can be decided by an always-halting Turing machine, is not a member of this Chomsky hierarchy. All recursive languages are also recursively enumerable. All context-sensitive, context-free and regular languages are recursive.

So, all formal languages are recognized by a Turing machine. A Turing machine has a transition function that, for a given state and symbol under the tape head, specifies three things: the symbol to be written to the tape, the direction (left or right) in which the head should move and the subsequent state of the finite control. The theory of formal language is the field of mathematics and computer science, which is concerned only with the purely syntactical aspects of such languages, given by their internal structural patterns. Formal languages are studied in the fields of logic, computer science and linguistics. Their practical application is for the precise definition of syntactically correct programmes for a programming language.

Although it is not part of the language, the words of a formal language have a semantic dimension. Thus natural languages cannot be described by these formal languages, without semantic words, and so, could not be recognized by a Turing machine.

Computer and Programming Languages

A programming language is an artificial language designed to create programmes which communicate instructions to a computer that is a machine. In computers, all programmes are given by an ordered succession of instructions and commands, written in artificial languages, created by man. The programmes written in an artificial high-level language must be translated to the computer low-level language, called the machine code, and executable by the microprocessor. It is important to note that the input and output of data does not depend of the language, but of the devices to read and write the data. There are computations of mathematical equations, symbolic systems or translator programmes. These programmes are specifications of a computation or algorithm. They can be used to control the machine and to perform applications. Programming languages are given, for example, by the Basic, Pascal and C languages. The set of programming languages is a subset of computer languages. For example, markup languages, like HTML or XML, are referred to as computer languages to emphasize that they are not meant to be used for programming. Programming languages have a syntax (a form) and semantics (a meaning). The syntax of a language describes the structure of valid texts given by possible combinations of symbols satisfying both a lexical structure and a grammatical structure. In computer science, the syntax is defined from a formal grammar. The semantics refers to the meaning of languages, as opposed to their form given by their syntax. In computer science, there are static and dynamic semantics. The static semantics of a language defines rules on the structure of valid texts that are not included in the syntax. The dynamic semantics, also known as execution semantics, of a language defines the various texts of this language that can produce the behaviour of a programme. The static and dynamic semantics defined the formal semantics.

The technological dream is that man can communicate with a computer in his natural native language (English, French, etc.) with a true man–machine dialogue. This includes the research on language for communication with robots (e.g. Steels, 2010). A natural language is based on a dictionary given by a set of words with a fractal structure, where each word is defined by the other words of the same natural language. In computer science, there is no such dictionary. The artificial language is based on computer words that are defined in a lexicon not written in the artificial language itself. So the machine is not able to understand its own artificial language, which is a set of instructions and commands, without a dictionary written with its own words of its lexicon.

Natural Languages Related to Recursion and Fractal

Any natural language is a communication tool between humans to exchange information.

For Edgar Morin (1986, p. 105), the hypercomplex organization of the mind–brain is inseparable of the trinitary complex [Dialogic, Recursion, Hologram]. In reading a text, the formation of the sense is a dialogic-recursive process between [Words, Text and Context]. The semantic structure of language is hologrammatic. Recall that, for Edgar Morin, the concept of hologrammatic structure is taken in a metaphoric sense, meaning that each part of the structure is at the image of the whole: if one cut a piece from a hologram plate, this piece permits still to view the whole image but with less precision. So, the fractal geometry is adequate for modelling a hologrammatic structure, because the main property of a fractal is the self-similarity: each part of a fractal structure is at the image of the whole. Moreover, the construction of a fractal is de facto made by a recursive and dialogic process (Dubois, 1990, p. 251). But if a fractal is created by a recursive algorithm, a recursive algorithm does not only create fractals.

For Kalevi Kull (1998), translation is the comprehension of the meaning of a script and the subsequent production of an equivalent script, likewise called a translation that communicates the same message in another language.

The base of natural languages is the dictionary of words that define each other. Each word is defined by a set of other words, that themselves are defined by other words, etc., to obtain all the words of the dictionary, itself defined by the set of all the words. The fractal geometry presents the same structure where the whole is in the parts that are in the whole and where the parts are able to regenerate the whole in a plus or minus good manner, following the scale of the considered parts. Moreover, a word has not always only one well-defined sense; it is its context that removes the ambiguity: this semantic aspect is the object of deep research. Then, a text has a sense only in the context of a situation; this is the pragmatic aspect, equally very studied. The semantic and pragmatic aspects are inherent to the fractal structure of the natural languages.

A recent paper defends the idea of loops and self-references in the construction of dictionaries (Levary, Eckmann, Moses, & Tlusty, 2012); let us cite these authors: "*The lexicon is a natural object that encompasses all the relations between words and meanings that exist in a language… Dictionaries provide snapshot representations of the lexicon*, *and*, *in particular*, *the relationship between words and concepts*." They found that new concepts are introduced into language by the formation of definitional loops in the dictionary: new concepts must be self-contained, so the collection of words used to represent them must be self-referential (Levary et al., 2012).

Some Mathematical Formulas in Computational Linguistic

This section will give some formulas which are used to quantify properties of languages.

The Shannon–Weaver Entropy in Natural Language Processing

In cybernetics, the observer of any system can know the content in information or value of the message represented by the set of probabilities $\{p_i\}$ at any instant from the Shannon–Weaver formula (1949) measuring the uncertainty by the entropy:

$$H = -\sum_{i=1}^{S} p_i \log_b p_i \quad \text{with} \quad \sum_{i=1}^{S} p_i = 1 \tag{12.1}$$

where p_i, $i=1,..,S$ is the set of S probabilities and b is the base of the logarithm function $\log_b$. This formula is also called the Shannon–Weaver index of diversity.

For example, the answer {yes, no} to a question, in binary base $b=2$, shows an uncertainty of 1 bit (binary digit) the unit of information:

$$H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1\,bit \tag{12.1a}$$

So the yes or no answer to a question gives a gain of information of 1 bit.

With this entropy formula, Shannon (1951) made a fundamental contribution to natural language processing and computational linguistics. He wrote that, for English words, using logarithmic scales both for word frequency (probability) and word order (rank), the curve is approximately a straight line with a slope -1; thus, if p_n is the probability of the nth most frequent word, we have, roughly, $p_n=0.1/n$. He remarked that there was a problem since the total probability $\sum_{n=1}^{\infty} 0.1/n$ is infinite, so he found that the critical n is the word of rank 8727. The entropy is then

$$H = -\sum_{n=1}^{8727} p_n \log_2 p_n = 11.82\ bits\ per\ word \tag{12.1b}$$

or 2.62 bits by letter since the average word length in English is 4.5 letters.

Shannon continued in saying that Zipf (1949) pointed out that this type of formula, $p_n=k/n$, gives a rather good approximation to the word probabilities in many different languages.

The Zipf and Mandelbrot Scaling Laws of Languages

In computational linguistics, elementary formulas about the quantitative entities of some properties of language texts were defined and which would be obeyed for all human languages which are very diverse and qualitatively complex. The most famous quantitative law of language is the Zipf law. In his works (Zipf, 1935, 1949, 1965), Zipf studied the following scaling law:

$$f(w) = A \Big/ \left[r(w)\right]^{\alpha} \tag{12.2}$$

that deals with the frequency distribution of words for long texts in English, which computes the frequency of any word $f(w)$ is inversely proportional to its rank $r(w)$ in the frequency table, where the exponent $\alpha \geq 1$ is slightly than 1, and where A is a normalizing constant. Thus the most frequent word will occur approximately twice as often as the second most frequent word, three times as often as the third most frequent word, etc. It was checked that this formula holds for other natural human languages and also for the artificial programming languages.

Mandelbrot (1954, 1966) corrected this Zipf law into the form

$$f(w) = A \Big/ \left[1 + Cr(w)\right]^{\alpha} \tag{12.3}$$

which fits language data better. The correction contains a second parameter C that needs to be adjusted to fit the data. Mandelbrot studied such scaling laws in his book (Mandelbrot, 1983), which inspired him to discover fractal geometry. It has been shown that this Zipf–Mandelbrot law is also obeyed by random processes that can be mapped onto texts.

The Kullback–Leibler Relative Entropy

The Kullback–Leibler (1951) divergence, also called relative entropy, information divergence or information gain, is a measure of the difference between two discrete probability distributions:

$$D(P \parallel Q) = \sum\nolimits_{i=1}^{S} p_i \ln p_i / q_i \tag{12.4}$$

if P and Q both sum to 1 and if $q_i = 0$ implies $p_i = 0$ for all i. It is a measure of the information lost when Q is used to approximate P, so P represents the actual distribution of data and Q represents a model of P.

Deduction of Invariants for Cyclic Time Series from the Shannon–Weaver Entropy

It is well-known that the Shannon–Weaver entropy, given by (12.1), deals with discrete systems. The Shannon–Weaver entropy was also used as an index of diversity. A generalization of this entropy was proposed in view of defining some invariants for cyclic time series (Dubois, 1973) represented by probabilities depending on time t, $p_i = p_i(t)$, with the temporal entropy:

$$H = H(t) = -\sum\nolimits_{i=1}^{S} p_i(t) \log_b p_i(t) \tag{12.5}$$

for which the total message, H^*, integrated over the cyclic time, T, is given by

$$H^* = \frac{1}{T}\int_0^T H(t)\,dt \tag{12.5a}$$

and, with the following normalization condition, $\frac{1}{T}\int_0^T \sum_{i=1}^{S} p_i(t)\,dt = 1$

A state of reference was defined by the set of averaged probabilities $\{p_i^*\}$: $p_i^* = \frac{1}{T}\int_0^T p_i(t)\,dt.$

The content of the message of this state of reference I_0 is given by the expression

$$I_0 = -\sum_{i=1}^{S} p_i^* \log_b p_i^* \tag{12.6}$$

After some mathematical developments, I obtained a first function (Dubois, 1973):

$$D_0 = \sum_{i=1}^{S}\left[p_i \log_b\left(p_i / p_i^*\right) + p_i^* - p_i\right] \tag{12.7}$$

where D_0 is a positive definite function that is zero at the state of reference $p_i = p_i^*$. This formula is also an extension of the Kullback–Leibler relative entropy. From the expression

$$H^* = I_0 - \frac{1}{T}\int_0^T D_0\,dt \tag{12.8}$$

one understands that the content of the message of a cyclic system H^* is the difference between the content of the message of the state of reference I_0 and the average of the function D_0 on a cycle; thus, D_0 is a dynamic characteristic of the system.

A second function was also obtained as

$$D_0^{(2)} = \sum_{i=1}^{S}\left[p_i^* \log_b\left(p_i^* / p_i\right) + p_i - p_i^*\right] \tag{12.9}$$

where $D_0^{(2)}$ is also a positive definite function that is zero at the reference state.

Let us recall that following Lyapunov (La Salle & Lefschetz, 1961), a system is stable if it is possible to find a positive definite function, L, for which the time derivative is negative or null. When the time derivative is null, the function is equal to a positive constant, $L = K$, defining a cyclic system with a non-asymptotic stability, and the system keeps in memory all its fluctuations.

So these two positive definite functions, D_0 and $D_0^{(2)}$, are called index of fluctuations and are also called index of diversity stability (Dubois, 1975). The time derivatives of these positive definite functions (12.7, 12.9) are given by (Dubois, 1973)

$$\frac{dD_0}{dt} = \sum_{i=1}^{S} \frac{dp_i}{dt} \log_b\left(p_i / p_i^*\right) \quad \text{and} \quad \frac{dD_0^{(2)}}{dt} = \sum_{i=1}^{S} \frac{dp_i}{dt}\left(1 - p_i^* / p_i\right). \tag{12.10a, b}$$

Two invariants K_1 and K_2 can be deduced in annulling the Eqs. (12.10a, b): $\frac{dD_0}{dt} = 0, \frac{dD_0^{(2)}}{dt} = 0\cdot$

On one hand, from the first INVARIANT: $D_0 = K_1$, after some mathematical developments, for the case $S = 2$, two differential equations were deduced (Dubois, 1973):

$$\frac{dp_1}{dt} = -F \log_b \frac{p_2}{p_2^*} \quad \text{and} \quad \frac{dp_2}{dt} = F \log_b \frac{p_1}{p_1^*} \tag{12.11a, b}$$

where F is any function of p_i, t.

On the other hand, from the second INVARIANT: $D_0^{(2)} = K_2$, also, after some mathematical developments, for the case $S = 2$, two other differential equations were deduced (Dubois, 1973):

$$\frac{dp_1}{dt} = -Fp_1\left(p_2 - p_2^*\right) \quad \text{and} \quad \frac{dp_2}{dt} = Fp_2\left(p_1 - p_1^*\right) \tag{12.12a, b}$$

where F is any function of p_i, t. The remarkable fact is that this system of equations, for $F = 1$, is the well-known Volterra (1931) model of prey, $p_1(t)$, and predator, $p_2(t)$, and also the well-known Lotka (1956) model of an autocatalytic chemical kinetics, which show cyclic oscillations.

Let us present now the Rényi formulas that generalize the fractal, Shannon–Weaver and correlation dimensions.

The Rényi Generalized Fractal Dimensions

The Rényi qth order dimensions at the scale s:

$$D_q = I_q(s) / \ln(1/s) \tag{12.13a}$$

where

$$I_q(s) = \left(1/(1-q)\right) \ln \sum p_k^q \tag{12.13b}$$

is the qth order Rényi information. In (12.13a and 12.13b), p_k is the probability in the k^{th} box of size scale length s, in assuming that they sum up to 1, $\sum p_k = 1$. For $q = 0$,

$$I_0(s) = \ln N(s) \tag{12.13c}$$

where $N(s)$ is the number of non-empty boxes of size scale length s. So $D_0 = D$ is the usual box counting method (e.g. Peitgen, Jürgens, & Saupe, 1992) of the fractal dimension, D. A practical example is given in a following section for Euclidean and fractal dimensions. For $q = 1$,

$$I_1(s) = -\sum p_k \ln p_k \tag{12.13d}$$

is the Shannon information and D_1 is the information dimension. For $q = 2$, D_2 is the correlation dimension, for which the Grassberger and Procaccia (1983a, 1983b) dynamic method has been proven to be particularly useful for obtaining information about the dynamics of a system solely from the knowledge of its time series. A linguistic application is given in the next section. In general, $D_p \geq D_q$ for any $p < q$. For fractals with uniform structure, all dimensions D_q have the same value.

Fractal, Multifractal and Correlation Dimensions of Time Series of Texts

The human written language, composed of words, punctuation signs and blanks in printed texts, which obeys rules of grammar, is an important field for the application of scientific methods in order to achieve a better understanding of computational linguistic. Let us cite some recent results. Ausloos (2008) discussed the comparison of two English texts written by Lewis Carroll, one, Alice in Wonderland, also translated into Esperanto, the other, Through the Looking Glass, in order to observe whether natural and artificial languages significantly differ from each other. He constructed one dimensional time series like signals using only word frequencies (FTS) or word lengths (LTS). These data were studied with the Zipf and Zipf–Mandelbrot scaling laws for sorting out correlations in the FTS and the Grassberger and Procaccia (1983a, 1983b) technique for finding correlations in LTS, which correspond to equilibrium and dynamic methods respectively. He found the usual value of $\alpha = 1$ in the Zipf law for LTS of word frequencies in natural and artificial languages. Nevertheless, Ausloos (2008) was surprised of the low values of $\alpha = 0.50$ and $\alpha = 0.33$ in the Zipf law for FTS of sentence lengths. For the LTS analysis with the Grassberger–Procaccia, Ausloos (2008) found a remarkable power law which does not indicate any saturation. Ausloos (2010) studied also the punctuation effects on sentence lengths for these same texts. He argued that sentences seem to be more reliable than word distributions in discussing an author style. Let us cite also the measure of complexity of these texts and translation in Esperanto with multifractals (Ausloos, 2012a) and with generalized Hurst exponent (Ausloos, 2012b). This is not the purpose of this paper to review the fractal properties of languages, many references can be found in the cited papers of Ausloos.

Computational Fractal Geometry with Recursion and Incursion

This section will explain the box counting method to measure Euclidean and fractal dimensions and will show how to generate fractal geometry with recursion and incursion.

Euclidean and Fractal Dimensions of Automata by Box Counting Method

Mandelbrot (1983) defined a fractal dimension dealing with fractional dimension instead of integer dimension. There is no more a characteristic length to define the scale of the system. There is a scale invariance: in looking at a fractal pattern at different scales, a similar or affine pattern appears, like a hologram. The fractal dimensions can be computed by box counting (e.g. Peitgen et al., 1992) from the relation:

$$D = \left[\ln N\left(d_{n+1}\right) - \ln N\left(d_n\right)\right] / \ln\left(d_n / d_{n+1}\right) \tag{12.14}$$

where $N(d_n)$ represents the number of boxes, of length d_n, containing at least one automaton at state different of 0, to cover the pattern, in considering different scales n. Let us apply this box counting method to the dimensions of the Euclidean geometry and to a fractal geometry (Dubois & Belly, 2000). The fractal geometry will be represented by the Sierpinski gasket (1915, 1916).

In Fig. 12.1a, the Sierpinski gasket fills the space with points, while in Fig. 12.1b, the Sierpinski fractal fills the space with binary digits: black square box in Fig. 12.1a is symbolized by the binary digit 1 in Fig. 12.1b, while a blank square box in Fig. 12.1a is symbolized by the binary digit 0 in Fig. 12.1b.

Euclidean geometry deals with integer space–time dimensions. If all the area of a space can be filled, without blank area, the fractal dimension is the Euclidean dimension: $D=1$ for a line, $D=2$ for a surface and $D=3$ for a volume. Let us consider the following Euclidean space of dimension 2 in Fig. 12.2a, b. The space in Fig. 12.2a is divided in 4 boxes of length 1/2, and Fig. 12.2b is divided in 16 boxes of length 1/4. All the boxes in grey can be filled. In Fig. 12.2a, $d_1=1/2$ and $N(d_1)=4$, and in Fig. 12.2b, $d_2=1/4$ and $N(d_2)=16$, so the Euclidean dimension is equal to

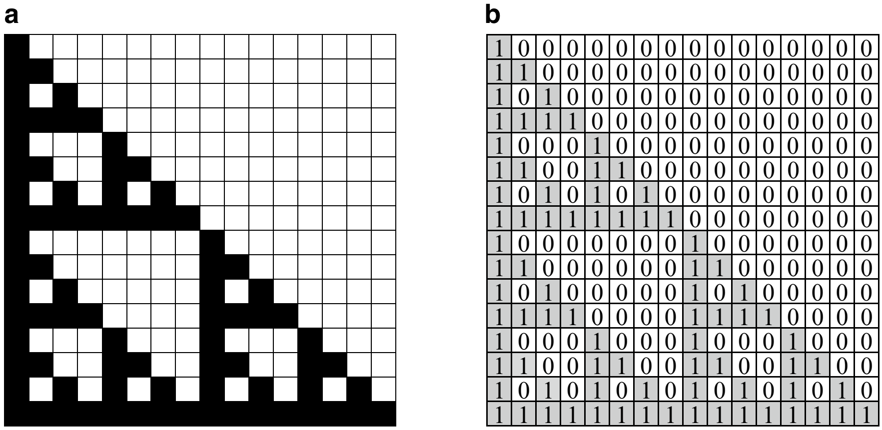


Fig. 12.1 Fractal Sierpinsky gasket. (**a**) The fractal Sierpinsky gasket is represented by *boxes* filled with *black points*. The *white boxes* are blanks. (**b**) The fractal Sierpinsky gasket is given by automata filled with binary digits 1. The 0 automata are blanks

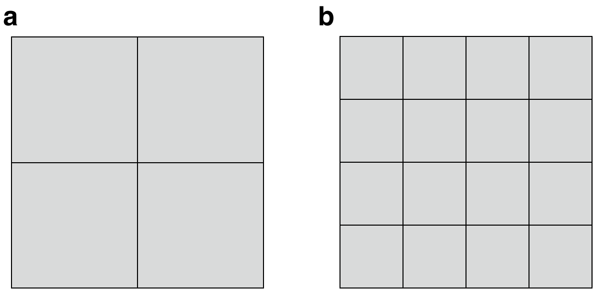


Fig. 12.2 Euclidian space covered by *boxes*. (**a**) Euclidean space covered by 4 *boxes*. (**b**) Euclidean space covered by 16 *boxes*

$$\begin{aligned} D &= \left[\ln N(d_2) - \ln N(d_1)\right] / \ln(d_1 / d_2) \\ &= \ln(16/4) / \ln(4/2) = \ln(4)/\ln(2) = 2 \end{aligned} \tag{12.14a}$$

and the dimension is $D=2$ for the two dimensions Euclidean space.

In fractal geometry, all the area of a space cannot be filled, there are blank areas. This is not a Euclidean geometry. Let us consider the Sierpinski gasket, divided in boxes, at Fig. 12.3a, b.

The space in Fig. 12.3a is divided in 4 boxes of length 1/2, and Fig. 12.3b is divided in 16 boxes of length 1/4. All the boxes in white cannot be filled. In Fig. 12.3a, $d_1 = 1/2$ and $N(d_1) = 3$, and in Fig. 12.3b, $d_2 = 1/4$ and $N(d_2) = 9$, so the fractal dimension is equal to

$$\begin{aligned} D &= \left[\ln N(d_2) - \ln N(d_1)\right] / \ln(d_1 / d_2) \\ &= \ln(9/3) / \ln(4/2) = \ln(3)/\ln(2) = 1.585 \end{aligned} \tag{12.14b}$$

and the dimension is less than 2 for this two dimensions Sierpinski fractal space. The pattern is self-similar, like a hologram.

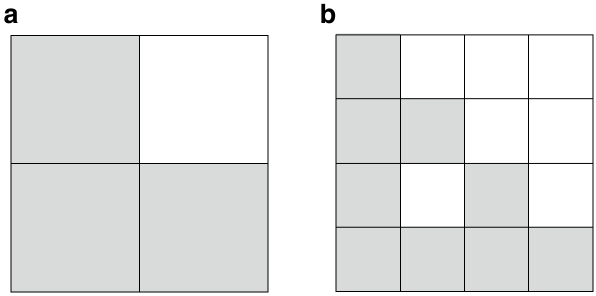


Fig. 12.3 Fractals with blank *boxes*. (**a**) Fractal with 3 *boxes*. (**b**) Fractal with 3×3 *boxes*

a

	n=	0	1	2	3	4	5	6	7	8
t=										
0		0	1	0	0	0	0	0	0	0
1		0	1	1	0	0	0	0	0	0
2		0	1	0	1	0	0	0	0	0
3		0	1	1	1	1	0	0	0	0
4		0	1	0	0	0	1	0	0	0
5		0	1	1	0	0	1	1	0	0
6		0	1	0	1	0	1	0	1	0
7		0	1	1	1	1	1	1	1	1
8		0	1	0	0	0	0	0	0	0

b

	n=	0	1	2	3	4	5	6	7	8
t=										
0		0	1	0	0	0	0	0	0	0
1		0	1	1	1	1	1	1	1	1
2		0	1	0	1	0	1	0	1	0
3		0	1	1	0	0	1	1	0	0
4		0	1	0	0	0	1	0	0	0
5		0	1	1	1	1	0	0	0	0
6		0	1	0	1	0	0	0	0	0
7		0	1	1	0	0	0	0	0	0
8		0	1	0	0	0	0	0	0	0

Fig. 12.4 Recursive and Incursive fractals. (**a**) RECURSIVE FRACTAL computed from recursive Eq. (12.15). (**b**) INCURSIVE FRACTAL computed from incursive Eq. (12.16). Let us remark that the incursive fractal is the time reverse of the recursive fractal

Generator of Fractal Automata with Recursion and Incursion

Let us recall a simple generator of Sierpinski gasket based on cellular automata. A one-dimensional network of N cellular automata is represented by a vector of automata states starting with initial values at time $t=0$. A set of rules defines how the states change at every clock time. A simple rule consists of computing the value of the state of each automaton at time $t+1$ by the sum modulo M of itself and its left neighbour at the preceding time t, for each clock time $t=0,1,2,3$, etc. The following recursive automata exhibit fractal structures:

$$X(n,t+1) = \left[X(n,t) + X(n-1,t)\right] mod\, M \tag{12.15}$$

with $t=0,1,2,\ldots$ and $n=1,2,\ldots,N$, starting with initial conditions $X(n,0)$, $n=1,2,\ldots$ at time $t=0$ and boundary conditions $X(0,t)$ at each time step $t=1,2,\ldots$, where mod M is the modulo M. For $M=2$, the logical rule is the exclusive OR, the fractal is the Sierpinski gasket given in Fig. 12.4a.

Computational property of recursive fractal automata: at each clock time, the order in which the computations are performed is without importance, what is called parallel iterations.

Dubois (1997) proposed a new type of automata (the fractal machine) where the order in which the computations are performed is to be ruled. The following incursive automata give rise to fractals of the same class as Sierpinski gasket:

$$X(n,t+1) = \left[X(n,t) + X(n-1,t+1)\right] mod\ M \tag{12.16}$$

This equation is an incursion, an inclusive or implicit recursion (Dubois, 1997), because the state of an automaton is a function of another automaton at the future time. This is computed in a sequential order, in giving initial conditions $X(n,0)$ and boundary conditions $X(0,t+1)$ at the future time $t+1$, for each time $t=0,1,2,\ldots$ Figure 12.4b gives the Sierpinski gasket from the incursive Eq. (12.16) with $M=2$, which is the logical exclusive OR.

Computational property of incursive fractal automata: at each clock time, the order in which the computations are performed is based on a path, what is called sequential iterations.

Such incursive automata will be also called hyperincursive automata because there are a lot of different paths which can be defined for computing successively the automata (Dubois, 1997; Dubois & Resconi, 1992). These hyperincursive automata are defined by the following: (1) a frame, (2) a composition rule and (3) a path. The path must be explicitly defined in order to compute the successive automata values in an ordered way. With the same frame and the same composition rule, many different processes can be generated in choosing different paths (Dubois & Resconi, 1992). Let us notice that such hyperincursive system sometimes may exhibit uncertainty and indecidability when the system defined itself its boundary conditions in a self-referential way (Dubois, 1998).

The computation with recursion in the Turing machine is irreversible; it cancels the symbols on the tape in writing new symbols. At the contrary, an incursive Turing machine is thus reversible.

Indeed, the computation with incursion in fractal automata is reversible (invertible), it is possible to rewrite the past sequences.

Conclusion

The field of computational language is a very difficult topic but with a vast potential of applications in the man–computer interaction. Some fundamental formulas are presented for quantifying some fractal properties of texts.

The main contribution to this paper deals with the introduction of an extension of the recursive generator of fractal. The recursive generator performs iterations in a parallel way; the order in which the computations are made is without importance; and in general, the recursive processes are irreversible, like in the Turing machine.

a

	n=	0	1	2	3	4	5	6	7	8
t=										
0		A	B	A	A	A	A	A	A	A
1		A	B	B	B	B	B	B	B	B
2		A	B	A	B	A	B	A	B	A
3		A	B	B	A	A	B	B	A	A
4		A	B	A	A	A	B	A	A	A
5		A	B	B	B	B	A	A	A	A
6		A	B	A	B	A	A	A	A	A
7		A	B	B	A	A	A	A	A	A
8		A	B	A	A	A	A	A	A	A

b

	n=	0	1	2	3	4	5	6	7	8
t=										
0		0	1	0	0	0	0	0	0	0
1		0	1	1	1	1	1	1	1	1
2		0	1	0	1	0	1	0	1	0
3		0	1	1	0	0	1	1	0	0
4		0	1	0	0	0	1	0	0	0
5		0	1	1	1	1	0	0	0	0
6		0	1	0	1	0	0	0	0	0
7		0	1	1	0	0	0	0	0	0
8		0	1	0	0	0	0	0	0	0

Fig. 12.5 Incursive fractals. (**a**) Incursive fractal, with the transition rules (12.17) of strings of letters A and B. After Dubois (1998). (**b**) Incursive fractal, from (**a**), with $A=0$ and $B=1$

The incursion is an inclusive or implicit recursion, and for the incursive generator, the iterations must follow a sequential order, called the path. The incursive generator shows that the generation of the fractal is reversible. In general, with incursion, several different paths exist and are to be fixed, so these incursive automata are also called hyperincursive automata.

The fractal concepts presented in this paper could be applied in the exercises proposed by Lowenthal (2010) and used in the Non Verbal Communication Device (NVCD) approach to favour language acquisition in the case of children with cerebral lesions.

For that, let us consider the system where the data and the operators are represented by the same symbols: let us consider a string of symbols represented by the letters A and B and define a composition rule for which each symbol acts on its right adjacent symbol, beginning at the left of the string and moving to the right of it, with the following transition rules of the four words with the two letters A and B (Dubois, 1998):

$$AA \rightarrow AA, BA \rightarrow BB, AB \rightarrow AB, BB \rightarrow BA \tag{12.17}$$

where the arrow → is the sign of transition of the strings of letters.

These rules represent an elementary grammar on strings of letters, which represents the logical exclusive OR, for the four words with the A and B letters. Figure 12.5a gives the result of the computation given by (12.17) compared with Fig. 12.5b.

In looking at Fig. 12.5b, with the translation $A=0$ and $B=1$ from Fig. 12.5a, it is seen that this is the same rule as for the incursive fractal in (12.16), given by Fig. 12.4b.

The concept of path is also important in the language reading and writing, because a text is an ordered sequence of words to be read or written following this sequence, for example, from the left to the right on a line, and line after line, and page after page. When writing a text, the next word on a line is added in considering all the previous written words, in an incursive way.

References

Ausloos, M. (2008). Equilibrium and dynamic methods when comparing an English text and its Esperanto translation. *Physica A: Statistical Mechanics and Its Applications, 387*, 6411–6420.

Ausloos, M. (2010). Punctuation effects in English and Esperanto texts. *Physica A: Statistical Mechanics and Its Applications, 389*, 2835–2840.

Ausloos, M. (2012a). Generalized Hurst exponent and multifractal function of original and translated texts mapped into frequency and length time series. *Physical Review, 86*(3).

Ausloos, M. (2012b). Measuring complexity with multifractals in texts. *Translation Effects. Chaos, Solitons & Fractals, 45*, 1349–1357.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, 2*, 113–124.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control, 2*(2), 137–167.

Dubois, D. M. (1973). Aspect mathématique de l'invariant en cybernétique. *Cybernetica, 16*(3), 161–176.

Dubois, D. M. (1975). Learning, adaptation and evolution of the environment ecosystem couple. *Journal of Cybernetics, 5*(2), 109–125.

Dubois, D. M. (1990). *Le Labyrinthe de l'Intelligence: de l'Intelligence Naturelle à l'Intelligence Fractale*. Paris: InterEditions.

Dubois, D. M. (1997). Generation of fractals from incursive automata, digital diffusion and wave equation systems, Invited paper. *Biosystems, 43*, 97–114.

Dubois, D. M. (1998). Hyperincursive methods for generating fractals in automata related to diffusion and wave equations. Invited paper. *International Journal of General Systems, 27*(1–3), 141–180.

Dubois, D. M., & Belly, M. (2000). Generating self-symmetrical fractals by hyperincursive automata and multiple reduction copy machine. *International Journal of Computing Anticipatory Systems, 6*, 95–115.

Dubois, D. M., & Resconi, G. (1992). *HYPERINCURSIVITY: A new mathematical theory* (p. 260). Liège: Presses Universitaires de Liège.

Grassberger, P., & Procaccia, I. (1983a). Measuring the strangeness of strange attractors. *Physica D, 9*, 189–208.

Grassberger, P., & Procaccia, I. (1983b). Characterization of strange attractors. *Physical Review Letters, 50*, 346–349.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve ? *Science, 298*, 1569–1579.

Kull, K. (1998). Organism as a self-reading text: Anticipation and semiosis. *International Journal of Computing Anticipatory Systems, 1*, 93–104.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics, 22*(1), 79–86.

La Salle, J., & Lefschetz, S. (1961). *Stability by Lyapunov's direct method with applications*. New York/London: Academic.

Levary, D., Eckmann, J. P., Moses, E. & Tlusty, T. (2012). Loops and self-reference in the construction of dictionaries. *Physical Review*, *X*(2).

Lotka, A. (1956). *Elements of mathematical biophysics*. New York: Dover Publications Inc.

Lowenthal, F. (2010). Language and recursion. In D. M. Dubois (Ed.), *Computing anticipatory systems*, *American Institute of Physics, AIP Conference Proceedings, 1303* (pp. 255–264).

Mandelbrot, B. (1954). Structure formelle des textes et communication. *Word, 10*(1).

Mandelbrot, B. (1966). Information theory and psycholinguistics: A theory of words frequencies. In P. Lazafeld & N. Henry (Eds.), *Readings in mathematical social science*. Cambridge, MA: MIT Press.

Mandelbrot, B. (1983). *The fractal geometry of nature*. New York: W.H. Freeman.

Morin, E. (1986). *La Méthode 2. La connaissance de la connaissance/1*. Paris: Le Seuil.
Peitgen, H.-O., Jürgens, H., & Saupe, D. (1992). *Chaos and fractals: New frontiers of science*. New York: Springer.
Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell Systems Technical Journal, 30*, 50–64.
Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Champaign, IL: University of Illinois Press.
Sierpinski, W. (1915). Sur une courbe cantorienne dont tout point est un point de ramification. *Comptes Rendus (Paris), 160*, 302.
Sierpinski, W. (1916). Sur une courbe cantorienne qui contient une image biunivoque et continue de toute courbe donnée. *Comptes Rendus (Paris), 162*, 632.
Steels, L. (2010). Adaptive language games with robots. In D. M. Dubois (Ed.), *Computing anticipatory systems*, *American Institute of Physics, AIP Conference Proceedings, 1303* (pp. 3–14).
Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society, 2*(42), 230–265.
Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem: A correction. *Proceedings of the London Mathematical Society, 2*(43), 544–546.
Volterra, V. (1931). *Leçons sur la théorie mathématique de la lute pour la vie*. Paris: Gauthier-Villars.
Wittgenstein, L. (1921). Tractatus Logico-Philosophicus (TLP). In C.K. Ogden, & P. Kegan (Eds). Originally published as "*Logisch-Philosophische Abhandlung*", in *Annalen der Naturphilosophische, XIV* (3/4). London: Routledge
Wittgenstein, L. (1953). *Philosophical investigations*. Oxford: Blackwell.
Zipf, G. K. (1935). *The psycho-biology of language: An introduction to dynamic philology*. Boston: Houghton Mifflin.
Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Cambridge: Addison-Wesley.
Zipf, G. K. (1965). *The psycho-biology of language, an introduction to dynamic philology*. Cambridge, MA: MIT Press.

Part V
Philosophy, Recursion and Language

Chapter 13
Consciousness, Recursion and Language

Roger Vergauwen

Introduction

Most researchers nowadays agree that recursion is an important property of human languages and at least partly explains why they can act as an effective communication device. However, one might also say that knowing a (specific) natural language also involves having a certain quale, a phenomenal conscious state of knowing the language which is presumably somehow related to a brain state. In this paper, I will concentrate on the mutual interrelationships between consciousness, recursion and language. My starting point will be an analysis of the qualia debate in contemporary philosophy of mind and the functionalist answer to Chomsky's competence model in Devitt and Sterelny's 'Martian argument' (Devitt & Sterelny, 1989).

Biological models of the mind show recursion to be present on the level of brain processes, and it can be argued to be present on the level of consciousness, but it would seem that looking at language in relation to qualia and the question whether these qualia can be causally active requires us to introduce the concepts of 'emergence' and 'downward causation'. Furthermore, natural language is characterised not only by recursion but also by 'aboutness' or 'intentionality' which might be a reason to speculate about a non-recursive element 'that lies beneath'.

R. Vergauwen (✉)
Centre for Logic and Analytic Philosophy, Institute of Philosophy,
Katholieke Universiteit Leuven, Leuven, Belgium
e-mail: roger.vergauwen@hiw.kuleuven.be

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_13,

Why Qualia?

In recent discussions in the philosophy of mind, the nature of phenomenal states or qualia—also called ‘raw feels’, ‘conscious sensations’ or the feel of *what it is like to be*—and of consciousness in general has been an important focus of research. Qualia, then, are (first-order) properties of mental states as, e.g. ‘seeing red’ and ‘feeling pain’, and the concept is sometimes used more broadly to indicate conscious experience in general. Explaining these qualia has been called ‘the hard problem of consciousness’: David Chalmers, e.g. writes: ‘If any problem qualifies as *the* problem of consciousness, it is this one. In this central sense of ‘consciousness’, an organism is conscious if there is something it is like to be that organism, and a mental state is conscious if there is something it is like to be in that state. Sometimes terms such as ‘phenomenal consciousness’ and ‘qualia’ are also used here’ (Chalmers, 1995, p. 201). Proponents of mind–body identity theories (Macdonald, 1989) claim that, in fact, there is no problem since it is perfectly possible to reduce these mental states and their properties to neurophysiological properties. In the case of functionalism, mental states are claimed to be reducible to functional or computational properties of the organism concerned. Others claim that a reduction is impossible and that, therefore, there remains an ‘explanatory gap’ (Levine, 1983). So, while some consider consciousness and qualia as some kind of epiphenomena with no real existence in their own right, others want to say that there is a fact of the matter with respect to these phenomena and that therefore they should be given a separate ontological status and that also we should be philosophical realists about the things in question. The first group, let us call them reductionist, are antirealists in the sense that they believe that there is no fact of the matter as to the existence of the aforementioned phenomena or entities unless maybe with respect to a certain background theory but not as something existing in a mind-independent way (Vergauwen, 2000, p. 366).

Functionalist Reductionism

In the philosophy of mind, functionalism is one of the theories used to clarify the mind–body problem. Functionalists hold that mental states, qualia and cognitive states in general can be viewed as the functional states of a machine or, rather, of a Turing machine. In this view, the explanation of cognitive mental states has to be done in terms of a causal-functional description of these functions. So, e.g. a typical functional explanation of ‘pain’ would imply that pain is a functional state and as such, therefore, functions as the detection of tissue damage. Of course, such a causal-functional state or role may cause the production of a certain behaviour or of another functional state which is the result of the first state. In functionalism, a mental state is defined in terms of three elements which are causally related: the input, the stimulation which causes the mental state; the causal interaction with other mental states;

and the output which is behavioural. In this sense, mental states are defined by means of a job description. Reductive explanations in science usually work this way. It is, e.g. known that organisms are able to transmit hereditary information. This was known even before the gene was discovered. Subsequent scientific investigations showed that in fact, DNA was responsible for this transmission and that therefore it did the job that was required. This kind of explanation of mental states is underpinned by a principle which is shared by many functionalists and nonfunctionalists, which is the principle of mind–body supervenience (Kim, 1998, p.10).

Principle of Mind–Body Supervenience

The mental supervenes on the physical in that any two things (objects, events, organisms, persons, etc.) exactly alike in all physical properties cannot differ in respect of mental properties.

That is, physical indiscernibility entails psychological indiscernibility or, as it is sometimes put, 'no mental difference without a physical difference'.

The principle of supervenience is combined in functionalism with another principle which is the principle of multiple realizability (Kim, 1998, pp. 70–71).

Principle of Multiple Realizability

Mental states are multiply realizable: any given mental state can be instantiated in (infinitely) many physical states.

The Martian Argument

It is from within such a functionalist framework that some have argued that Chomsky's approach, who considers linguistics to be part of cognitive psychology, is misguided. Devitt and Sterelny (1989) have presented an argument—dubbed the Martian Argument by Stephen Laurence—intended to show why this is so. Linguistics, according to Chomsky, deals with human knowledge and understanding of language and linguistic competence and is therefore concerned with questions about the cognitive skills of human beings and thus in the end ultimately about the specific human neural setup. The psychological mechanisms for language acquisition are instantiated into the mind/brain of the human linguistic users. Since grammars are about such instantiations, the instantiation of English grammar in humans is indeed English according to this conception. This we may call the competence thesis, the thesis that grammars are about the human linguistic competence. The Martian argument, now, runs as follows. Let us assume that Martians, whose psycholinguistic processes ex hypothesi differ from human ones, nevertheless manage to produce a set of sentences that are extensionally equivalent to the set of sentences in human English. This implies that the sentences that are grammatical in Martian English are also grammatical in regular English. However, since by hypothesis Martians had a different neural organisation, they really have a different linguistic competence, and now the question arises whether they should count as speaking English. On an account of what it is to speak English, the Martian speakers should indeed count as speaking English. On the level of linguistic symbols, everything they say is indistinguishable from us, and they also seem to be able to communicate via a seemingly shared language. They are, therefore, on the face of it, competent in English. We can, then, nevertheless study the shared language that we both use on the linguistic level without having to appeal to differences in competence:

'The theory of Martian competence would have to be different from the theory of ours. Yet, the theory of symbols would be the same, for it would still be English that they spoke. Returning to earth, it would not matter a jot to the theory of symbols if competence among actual English speakers was entirely idiosyncratic.' (Devitt & Sterelny, 1989, p. 514).

This argument has caused a lot of discussion into which we do not want to go here for the moment, but we want to maintain that this functionalist approach, which is in fact nothing but a variant of the Turing Test, neglects a hitherto not enough noticed fact, namely, that 'there is something it is like to speak a natural language', which in turn means that in fact understanding and speaking a language (linguistic competence) is a quale. Moreover, functionalism at best does not believe that qualia can be fully functionalized, and because of the supervenience principle, according to Kim and others, in combination with certain intuitively plausible assumptions (causal exclusion principles), it implies a kind of epiphenomenalism, the view that every mental event is caused by a physical event in the brain but that mental events have no causal powers of their own. They are powerless to cause anything else, not even other mental events, and are therefore almost literally 'non-existent'. In our view, language and consciousness are intimately related, and both have a qualitative aspect. A further investigation into the nature of these qualia may, then, reveal something of the true nature of the mental and language. It seems to me that such an investigation has to concentrate on the embodiment of the mental and the linguistic.

Edelman on the Mind and Qualia

A first indication that such an approach is feasible can be found in the neurobiological model proposed by Gerald Edelman, more specifically, in his theory of neuronal group selection where he uses the concept of a Dynamic Core (Edelman & Tononi, 2000). Here, a distinction is made between primary consciousness and higher-order consciousness (Murphey & Brown, 2007, pp. 141–144). Primary consciousness involves the ability to construct 'mental scenes' with limited informational or symbolic content, such as may be found in an animal's ability to learn and adapt its behaviour and which is more or less immediate. Higher-order consciousness evolves later and is related to semantic capabilities, as is found in its fully developed form in human beings who have a fully fledged language with syntax and semantics. It therefore goes beyond primary consciousness as a 'remembered present' and extends into the past and the future. Essential properties of this kind of consciousness are unity, informativeness and privateness (accessible only to conscious agents).

The theory further provides a neurophysiological description of the likely mechanism at the basis of conscious awareness. It is argued that a conscious state is a dynamical process within the cerebral cortex in which functional interconnectedness is created by rapid, two-way (re-entrant), neural interactions. Such a functional integration is called a 'dynamic core'. This core is itself, then, a group of neurons—a neural state—and it is the sets of neurons involved in the (functional) relations between them that constitute the nature and content of consciousness at any given moment. Such dynamic core(s) may be sensitive to further bodily (sensory) inputs or inputs from other brain systems with which they interact, as well as with itself. This gives rise to extremely complicated patterns of interactions, but what is also important is that the whole system is embodied, especially with respect to the subjectivity of experience. The activity of the dynamic core leads to successive discriminatory states which entail sets of phenomenal experiences (qualia). Consciousness in this view arises from re-entrant interactions among neural populations, and the causal activity is produced by the dynamic core; so the qualia emerging from this core are caused by it, and according to Edelman, this does not necessarily lead to epiphenomenalism because these emerging qualia are informational structures even if they are not causal, and it may sometimes be useful to talk about these qualitative informational states 'as if' they are causal depending on the level of description. This does not, however, imply epiphenomenalism, since "There is, however, no need to conclude that C [qualia] is therefore meaningless and unnecessary. C states [qualia-states] are informational even if not causal. C states are the discriminations entailed by causal transactions among C '[neural states]'" (Edelman, 2003, p. 5523). The qualia, then, are the reflections of the permanent causal interactions in the (complex) dynamic core but are themselves not causally active: 'underlying each quale are distinct neuroanatomical structures and neural dynamics that together account for the specific and distinctive phenomenal property of that quale' (Edelman, 2003, p. 5523). As for the phenomenal experience of the quale itself, the theory holds that it is no problem—and that therefore the 'hard problem' and the problem of the 'explanatory gap' (Chalmers, 1995; Levine, 1983) are ill posed—because the need for a phenotype experiencing giving rise to the qualia is no hindrance for a scientific theory of consciousness, and, moreover, 'if the phenomenal part of conscious experience that constitutes its entailed distinctions is irreducible, so is the fact that physics has not explained why there is something rather than nothing' (Edelman, 2003, p. 5524). Let us notice right away that the re-entrant activity of the dynamic core which causes the qualia has a recursive nature.

Indeed, a main tenet of Edelman's thesis is the concept of re-entrant signalling between neuronal groups. He defines re-entry as the ongoing recursive dynamic interchange of signals that occurs in parallel between brain maps and which continuously interrelates these maps to each other in time and space. Re-entry depends for its operations on the intricate networks of massively parallel reciprocal connections within and between neuronal groups, which arise through processes of developmental and experiential selection. Edelman describes re-entry as 'a form of ongoing higher-order selection… that appears to be unique to animal brains' (Edelman, 1998, p. 46) and adds that 'there is no other object in the known universe so completely distinguished by reentrant circuitry as the human brain' (Edelman & Tononi, 2000, p. 49). So, in fact, the human brain is a massively recursive machine.

It should, then, come as no surprise in this view that human language indeed inherits this recursion. Indeed, Hauser, Chomsky and Fitch have suggested that recursion is a basic characteristic that distinguishes human language from all other forms of animal communication. They make a distinction between a faculty of language in the broad sense and a faculty of language in the narrow sense (Hauser, Chomsky, & Fitch, 2002, p. 1570–1571). The first of these faculties contains an internal computational system in combination with at least two other organism-internal systems which they call the 'sensory-motor' and the 'conceptual-intentional'. The second faculty is the abstract linguistic computational system by itself which is contained in the first faculty but which is independent of the other systems with which it nevertheless interacts. As such, then, the second faculty is a component of the first one, and the mechanisms that underlie the first one are a subset of the ones that underlie the other. It is in this process that they see a central place for recursion: 'We assume, putting aside the precise mechanisms, that a key component of FLN (faculty of language in the narrow sense) is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system…All approaches agree that a core property of FLN is recursion, attributed to narrow syntax in the conception just outlined' (Hauser et al., 2002, p. 1571).

However, a problem remains. A discussion of Edelman's neurobiological model has shown that at first sight, though qualia are real as informational states, in this approach they seem to be doomed to causal impotency. But if this is the case, one may say that they might as well not exist and that therefore the functionalist threat as present in the Martian argument has not been removed. If Chomsky is right, then in our view there is indeed 'something it is like to have a language', and the corresponding quale is a consequence of the recursive activity in the brain, but this is apparently not enough even though recursion is hardwired in the brain. So it would seem that we need a means of making explicit how causal activity is somehow inherent in the activity of qualia. We suggest that recursion can only be a support of this thesis if we go one step further by appealing to the concepts of downward causation and third-order emergence.

Consciousness and Causality

Edelman, again, notices that qualia emerge from their physical substrate and, as informational structures, reflect the causal transactions between the successive stages in the dynamic core. They constitute a ‘phenomenal transform’ (Edelman, 2003, p. 5523). While in themselves they are not causally active, they are nevertheless entities in their own right. If we want to make sense of the causal nature of qualia and phenomenal consciousness it is important, first, to notice that qualia *emerge* from their neural substrates. They are, therefore, a direct consequence of the embodiment of the mental and at the same time the expression of the dual nature of physically embodied information. The concept of emergence has a rich history (Murphey & Brown, 2007, pp. 78–80), and it comes in several variations. It was especially rife in the philosophy of biology throughout the twentieth century where it was proposed as an alternative to vitalist accounts of the origins of life. Gradually, the concept came to be applied to the mind–brain problem and especially the emergence of consciousness.

The concept itself is by no means easy to define, but there are some useful characteristics that it may be said to have: according to Kim (1998, pp. 226–229), emergentism implies the view, first, that all that exists in the space-time world are the basic particles recognised in physics and their aggregates. It further entails that when aggregates and their ↑ articles attain an appropriate level of structural complexity, genuinely new (higher order) properties emerge to characterise these structured systems and that these emergent properties are irreducible to, and unpredictable from, the lower level phenomena from which they emerge. Finally, once these higher level properties (e.g. mental states or qualia) emerge, they are able to manifest causal powers in their own right, thus affecting the lower level phenomena (‘downward causation’). Our approach intends to show that causal efficacy and downward causation are intimately related to emergence and, more specifically, to what Terrence Deacon (2007) has called ‘third-order emergence’.

Deacon has offered a novel account of emergence which sheds light on this. In general, the emergentist’s aim might be said to show how, within complex systems, new entities emerge that exhibit novel causal powers. Deacon starts from the assumption that in emergent phenomena, the physical laws governing the constituents of a system should not be violated and that an additional account should be given of the configurational regularities affecting constituent interactions. Furthermore, the relative autonomy and causal efficacy of such ‘holistic’ emergent phenomena should be accounted for. Deacon makes a threefold distinction between different kinds of emergent systems. In doing so, three interconnected hierarchical levels of emergence can be described. A first kind of emergence is ‘first-order emergence’ or ‘supervenient emergence’. This happens in systems in which relational properties determine the emergent higher-order properties, e.g. such as in the case of the liquidity of water as determined by the aggregation of water molecules (Deacon, 2007, pp. 97-98). A second kind of emergence, ‘second-order emergence’, is present in diachronic symmetry-breaking typically also found in living systems and mental processes: ‘in contrast, there is a self-differentiating feature to living and mental processes, which retains and undermines aspects of self-similarity. This characteristic breakdown of self-similarity or symmetry-breaking is now recognised in numerous kinds of complex phenomena, including systems far simpler than living systems. These complex emergent phenomena share this characteristic change of ensemble properties across time, and are often computationally unpredictable’ (Deacon, 2007, p. 99).

Third-order emergence involves, in addition, information or memory. It is the kind of higher-order regularity which can additionally exert a cumulative influence over the entire causal future of the system thus encompassing the evolution of second-order processes. According to Deacon, this involves an additional leap of recursive causality: ‘the relationship implicit in third-order phenomena demands a combination of multi-scale, historical, and semiotic analyses. Thus, living and cognitive processes require introducing concepts such as representation… information and function in order to capture the logic of the most salient emergent phenomena’ (Deacon, 2007, pp. 106–107). The level of third-order emergence, therefore, is ‘the point where physical causality constitutes significance’ (Murphey & Brown, 2007, p. 83). Among these third-order emergent properties, Deacon explicitly considers processes such as ‘mental experiences’, suggesting that the kind of causation involved here is downward causation. Among the processes emerging here, we surmise that linguistic capacity with a recursive basis also belongs. Indeed, in our view, qualia or phenomenal states can be described in terms of emergent processes of information which can have downward causation but arise only when the information is physically embodied.

Research by E. Thompson and F. Varela can be used to illustrate this (Thompson & Varela, 2001). In a framework describing the relationships between neural dynamics, consciousness and embodiment, they propose to investigate the neural correlates of consciousness. In doing so, they do not assume only a one-way (upward) causal relationship between neural systems and the contents of consciousness but also the possibility of a two-way (downward) causal relationship between embodied conscious states and neuronal activity (Thompson & Varela, 2001, p. 418). They suggest that as a result of emergence in complex system, it is to be expected that in such systems there is both upward causation and downward causation implied in the relationship between neural activity and conscious activity. Since specific cognitive acts demand the integration of distributed and interacting areas of the brain, the search for a neural correlate of consciousness must account for these large-scale integrations of brain activity.

The mechanism they suggest for this integration is the formation of dynamic links mediated by synchronisation of neuronal discharges over multiple frequency bands: ‘Given that the coupled dynamics of brain, body and environment exhibit self-organisation and emergent processes at multiple levels and that emergence involves both upward and downward causation, it seems legitimate to conjecture that downward causation occurs at multiple levels in these systems, including that of conscious cognitive acts in relation to local neural activity’ (Thompson & Varela, 2001, p. 421).

An important consequence of this is that these processes can be studied empirically. Thompson and Varela mention as examples of this the study of human epileptic activity and voluntary perceptual reversal. Human epileptic activity freely modifies the subject's mental competencies, but the converse would also seem to be the case. The subject can voluntarily affect his or her electrical condition that normally would lead to an epileptic seizure (Thompson & Varela, 2001, pp. 421, 422). Furthermore, models of the visual perception of multi-stable or ambiguous figures suggest that such perception is based on generic properties of coupled non-linear oscillators and their phase relationships which might show 'that different 'cognitive' interpretations of ambiguous figures initiated by the subject might shift the neuronal bias that defines the perceptual reversal' (Thompson & Varela, 2001, p. 421). Emergence and downward causation, therefore, can account for the causal activity of consciousness as embodied and that this activity does in fact involve not only an upward part but also a downward one. Our approach shows that embodiment is essential to consciousness and by extension also language, thereby partly rebutting the Martian argument since it clearly shows that the specific neural set-up of the mind is responsible for the specific result, i.e. human natural languages, and that recursion is present both on the neurophysiological and symbolic level. But at the same time, it seems that something else also has to be taken into account which points in a very different direction.

Recursion and Intentionality

The theory we have presented here is to a certain extent tributary to David Chalmers' so-called double-aspect theory of information: his hypothesis states that information has two fundamental aspects, a physical one and an experiential or phenomenal one (Chalmers, 1995, p. 216) and that, furthermore, phenomenal consciousness by virtue of its status is one aspect of information where the other aspect is found embodied in physical processing. Physical information is, as embodied information, then, 'a difference that makes a difference' (Chalmers, 1995, p. 216).

We have seen how qualia and conscious mental states can be causally active entities, whence they emerge and why the kind of downward causality they exhibit is not in contradiction to the physical closure of the world. The theory presented here is non-reductive. It allows for the emergence of mental states that are informational but not fully functionalizable since the information is embodied information.

Chalmers notes that there are certain structural analogies between the physical and the phenomenal aspects of information: 'there is a direct isomorphism between certain physically embodied information spaces and certain phenomenal (or experiential) information spaces… That is, we can find the same abstract information space embedded in physical processing and in conscious experience' (Chalmers, 1995, p. 216). Inevitably, at this moment, questions of computability of this kind of information show up.

One of the limits on the computability of a system is that it may be undecidable in a strict logical sense. Roger Penrose, for one, has suggested how quantum mechanics may help to explain how the human mind goes beyond Turing computability (Penrose, 1994). He does not talk explicitly about consciousness or language but, rather, about mathematical creativity and uses Gödel's famous incompleteness theorems to show how human beings can 'see' the truth of certain mathematical propositions even though within an axiomatic system these propositions are undecidable, i.e. neither the propositions themselves nor their negations can be proven.

The phenomenon of logical incompleteness may show something about the subjectivity of subjective experience and phenomenal states. If Chalmers is right in saying that phenomenal properties constitute the internal (phenomenal) aspect of information, then a metalogical property such as incompleteness which is definitely a property of (complex) formalised mathematical systems is also a property of the information the system contains. Mental states and language are usually characterised as being 'intentional'. Indeed, intentionality or 'aboutness'—a term coined by the Austrian philosopher Franz Brentano—points towards the directedness upon the world of language and mental states, their content or their (possibly non-existing) reference. There is some discussion on whether qualia exhibit intentionality (Northoff, 2003). Some have argued that qualia are primary and that intentionality can be inferred from them. Others consider intentionality as a necessary condition for qualia: 'without qualia, which reflect the experience of perceptions and actions, a direction towards "observable and to-be effectuated events within the environment" would remain meaningless i.e. superfluous since it could no longer be experienced as such. Such a 'hollow' intentionality, i.e. intentionality without experience, remains naturally impossible…. qualia without intentionality would remain "empty"…' (Northoff, 2003, pp. 136, 137). However this may be, what we can say is that in as far as some complex physical systems, such as human brains, are able to refer to things they can be said to exhibit genuine intentionality and that therefore 'intentionality is at least a sufficient condition for mentality' (Kim, 1998, p. 23).

In view of the theory presented here, we suggest that there is a close relationship between intentionality and logical incompleteness. If a system is logically incomplete, it means that it has non-isomorphic models, also called non-standard models, which is, for instance, the case with the models of formalised arithmetic in Gödel's theorems. This comes down to the fact that the information in the axioms of the theory is in a sense insufficient to uniquely determine what they are 'about'. Incompleteness may, then, be seen as a logical analogon of intentionality in the following sense. In the correspondence theory of meaning and truth, the property of truth expresses the 'aboutness' of propositions in the sense that true propositions are 'true of' or 'about' true states of affairs. The property of truth, therefore, expresses the 'reference relation' for the propositions. If truth is non-computable or non-recursive, then so is intentionality. If Penrose is right, human beings are able to see the 'truth' of certain mathematical statements and are therefore able to perform non-computable operations which Penrose thinks are a consequence of non-computable elements in the physical laws which the brain exploits when it performs this task.

Conclusion

One need not agree with Penrose's idea of performing non-computable (non-recursive) tasks to see that the general phenomenon of incompleteness might epistemologically indicate that the 'aboutness' is a consequence of a tension between the inside ('information') of the system and the outside of the world, which is what the information is supposed to be about, a tension which in its turn is a consequence of the embodiment of the mind. The 'feeling of aboutness' might then, somewhat speculatively, be defined as the reflection of the difference between 'mind' and 'the world'. Be that as it may, the main aim of this paper was not so much to analyse the concept of intentionality as to investigate how a unified theory of the mind and language could be possible which can withstand the strong functionalist challenge of the Martian Argument and how in such a theory embodiment should be central. Our discussion, however, indicates that though recursion is a fundamental principle according to which human languages are built and function, there remains a genuine possibility of another level which we might call non-recursive and which has to be taken into account on the semantic side especially when one wants to give an account of linguistic intentionality.

References

Chalmers, D. (1995). Facing up the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

Deacon, T. (2007). Three levels of emergent phenomena. In N. Murphy & W. Stoeger (Eds.), *Evolution and emergence. Systems, organisms, persons* (pp. 88–110). Oxford: Oxford University Press.

Devitt, M., & Sterelny, K. (1989). Linguistics: What's wrong with "the Right View"? *Philosophical Perspectives. Philosophy of Mind and Action Theory, 3*, 498–531.

Edelman, G. (1998). Building a picture of the brain. *Daedalus, 127*(2), 37–69.

Edelman, G. (2003). Naturalizing consciousness. A theoretical framework. *Proceedings of the National Academy of Sciences, 100*(9), 5520–5524.

Edelman, G., & Tononi, G. (2000). *A universe of consciousness: How matter becomes imagination*. New York: Basic Books.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve. *Science, 298*(22), 1569–1579.

Kim, J. (1998). *Philosophy of mind*. Colorado: Westview.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly, 64*, 354–361.

Macdonald, C. (1989). *Mind-body identity theories*. London: Routledge.

Murphey, N., & Brown, W. (2007). *Did my neurons make me do it? Philosophical and neurobiological perspectives on responsibility and free will*. Oxford: Oxford University Press.

Northoff, G. (2003). *Philosophy of the brain*. Amsterdam/Philadelphia: John Benjamins.

Penrose, R. (1994). *Shadows of the mind*. Oxford: Oxford University Press.

Thompson, E., & Varela, F. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418–425.

Vergauwen, R. (2000). Realism, reference and logic. In F. Beets & E. Gillet (Eds.), *Logique en Perspective* (pp. 365–398). Brussels: Ousia.

Chapter 14
There Is No Recursion in Language

Pierre Frath

> *"Then Shem Macnamara had been very poor, only too ready for a free meal and a quiet sneer at the success of a fellow poet. Then, instead of expensive mouthwash, he had breathed on Hogg-Enderby, bafflingly (for no banquet would serve, because of the known redolence of **onions**, **onions**) **onions**.*
>
> *"Onions", said Hogg. He was frowned on in puzzlement. Cocktail onions, he offered. Well just imagine. Shem Macnamara deepened his frown. Something in that voice saying Onions? He did not take any onions."*
>
> "Enderby outside"
> Anthony Burgess (Penguin, 1982, p. 224).

Introduction

"There is no recursion in language". Such a radical and uncompromising title in a conference devoted to the very subject of recursion in language seems a bit provocative. Maybe it was, and I will try to explain why. But I have no regrets even if after listening to the speakers at the conference and discussing with participants, I admitted that there may be something in cognition which could be construed as recursion. This does not mean that I accept recursion as a property of the mind and language. I shall argue against this notion with the help of Wittgenstein's

P. Frath (✉)
Université de Reims Champagne-Ardenne, Reims, France
e-mail: Pierre.frath@aliceadsl.fr

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0_14,

philosophy (1961, 1963) and C.S. Peirce's semiotics. This paper will also be set against the background of recent linguistic work on reference in language by Georges Kleiber[1] and myself.[2]

I will first look at the naming process in terms of Wittgenstein's language games. I will then sum up and analyse the views developed at the Mons conference, and I shall finally try to account for what is meant by recursion.

Language Games

It is well known that Wittgenstein rejected metaphysical systems constructed on ontological hypotheses[3]; he favoured looking at the way language is used in what he called "language games"[4].

Let us consider the sentence template "Is there X in Y?". Will it work in the same way independently of the values given to X? Let us compare two short dialogues:

1. "*Is there any love in Tom and Susy's relationship*?
   - *Love*? *No way*. *They are always quarrelling*.
   - *Well*, *this is their way of expressing love*.
   - *Well*, *your notion of love is a bit weird*".
2. "*Is there any salt in the soup*?
   - *Well*, *I don't know*. *Let's taste it*.
   - *Yes*, *there is* (*or no there isn't*, *or yes there is but not enough*, etc.)".

There clearly are differences. The presence of salt in the soup is or is not a fact, and the speakers can easily agree by resorting to an action: tasting the soup. They may disagree about the quantity of salt in the soup, but they will certainly agree on what salt is. In the case of love, all will depend on what is meant by love. If there is disagreement, clarification will come in the form of another utterance which will develop an aspect of what the speaker means by love in this particular case. Peirce calls such linguistic developments *interpretants*[5]. They are constrained by the corpus of things that can be said about love and of which language is the repository.

---

[1] For an introduction to "dénomination", see Kleiber's *Nominales* (1994).

[2] See my papers and some of Kleiber's recent ones on http://www.res-per-nomen.org.

[3] See also Peirce, who thought that ontologies were only as strong as their foundations, i.e. not at all because all foundations are hypotheses (1868, Fr. tr. 1984: 197).

[4] See the beginning of his *Philosophical Investigations* (PI).

[5] On the notion of *interpretant*, see, for example, §5.473, 5.253 and 2.303, in Peirce's collected works. French translation in Peirce (1978).

Names and Objects

The difference is linked to the nature of the referent. Salt refers to an object whose existence in the real world is not doubted. As a consequence, *salt* can easily be understood in terms of its referent, its use by people, etc. *Love*, on the other hand, does not refer to a real-world object. If mankind were suddenly to disappear, there would still be salt in the universe (along with all other physical objects), but love and other such entities (hate, fun, intelligence, etc.) would disappear with the last human being. Let us call the first *real-world objects* and the others *anthropological objects*.

Names refer to real-world and anthropological objects in the same way. In both cases, objects can be talked about, and knowledge then consists in a discourse which expands such or such an aspect of the object. Yet there is a difference: all the meanings of a real-world object name (such as *salt*), however metaphorical, are related in some way to the object, whereas for anthropological object names (such as *love*), all the meanings refer to how the name is used in language and to a corpus of discourse that the speakers share.

The relationship between language and the world has puzzled philosophers ever since ancient times. For Plato, anthropological and real-world objects are related to a world of ideas and categories of which they are instances. A person feels love by virtue of the existence of *love* in a non-corporeal world; a blacksmith is able to mould a bronze sphere because bronze exists in the real world, albeit shapelessly, and because the geometric sphere exists somewhere in a world of ideas. As for categories (e.g. trees), they are endowed with some sort of Platonic existence independently of the actual objects they comprise (individual trees). They consist in a set of logical and/or psychological properties, which in turn determine whether or not an individual object belongs to a particular category (if it has a large vertical ligneous stem, branches and leaves, it is a tree). Discrepancies (are palm trees trees?) can be dealt with by modifying the set of properties. In both cases, the actual instance is explained in terms of reference to a Platonic entity, non-corporeal ideas for anthropological objects and categories for real-world objects. As a result, Platonic entities have some sort of causal value: their generic existence can be construed as the cause of particular existences. Such views have left a very deep mark in Western thought in many domains. Mathematicians, for example, often believe numbers have some sort of Platonic existence per se in the universe[6].

Names and their referents are so closely linked that we hardly make a difference between them. This is because names extract elements from our experience and give them a separate existence. What English speakers name a river is in French either a *fleuve* (a large river which flows into the sea) or a *rivière* (a river which flows into another river or a smaller river which flows into the sea): there are two separately named objects for a francophone, only one for an anglophone. In Russian, there are three words referring to what English speakers would call *fruit juice*. There is *sok*, which names juice made from fresh fruit such as apples, pears, oranges, and lemon.

[6] The alternative view is a nominalist one: integers are the names of sets.

There is also *mors*, which names juice made from fresh cranberries. And there is *kompot*, which names juice made from boiled dried fruit such as raisins, dried apples, pears, apricot and also juice made from fresh berries (strawberries, blueberries, blackberries, currant, etc.), either boiled or not[7]. Names select bits and pieces of our experience and give them existence. They do so differently in each language and quite randomly.

We live in a named universe. What is named is endowed with existence, whether in the real world (salt, trees, etc.) or in our human experience only (intelligence, love, Santa Claus, unicorns, etc.). Conversely, we believe that all objects have names and that any unnamed object is out there waiting to be named. Unnamed objects cannot be talked about, and no knowledge about them can be acquired and transmitted. In effect, they do not exist *for us* even if we surmise the universe is full of them. Language is our limit. As Wittgenstein says, "The world is my world: this is manifest in the fact that the limits of language mean the limits of my world"[8].

Names and Corpora

Once an object has a name, it can be used in discourse, and knowledge can accumulate in an ever increasing corpus. Names are the kernels of corpora sometimes dating back to very ancient times. As a consequence, any named object is never offered unveiled to our scrutiny, untouched by the corpus which contains the knowledge about it and which gives it meaning.

The question of consciousness, for example, has been the subject of innumerable texts dating back to antiquity. A number of words have been used to refer to it: soul, mind, *ego* and consciousness, each with their own corpora and language games. They all share the notion that there is some sort of non-corporeal entity inside humans (and maybe other animals as well). What many philosophers fail to see is that the very use of one of those words gives its referent some sort of separate reality, locking them into a dualistic quagmire[9]. When a philosopher studies consciousness, is he considering a real object with a separate existence or a linguistic and anthropological artefact? How can he draw the line?[10]

This means that the difference between real-world and anthropological objects is not so clear-cut. In fact, all objects of our experience are anthropological. They are at the centre of a corpus, and as such, they have semiotic existence. Only a subset of them also has real-world existence. Love is entirely anthropological. Salt on the

7 Many thanks to Olga Frath for her patient explanations.

8 *Tractatus Logico-Philosophicus* (TLP): 5.62.

9 This is why dualism is so resilient. Philosophers, such as Dennett, Chalmers, Popper, Crick, the Churchlands and many more, develop basically dualistic theories, even though some of them (e.g. Dennett) think they do not.

10 See Frath (2012).

other hand is very significant in most cultures as a real-world object, but it also has a powerful meaning at the symbolic level as shown by linguistic expressions such as "the salt of the earth" or "Cela ne manque pas de sel". Any named object, however mythical its real-world existence, can be relentlessly discussed[11], until we stop talking about it altogether and it disappears into linguistic oblivion[12]. As Wittgenstein says, truth is an anthropological entity. "It is what human beings *say* that is true and false; and they agree in the *language* they use. That is not agreement in opinions but in **form of life**"[13].

Are mathematical objects anthropological? Yes they are, because they only have existence within a theory, i.e. a highly constrained discourse, in other words a language game. They may sometimes describe the real world, but this is only a spin-off of their theoretical existence. Mathematicians are not basically interested in describing the world. Compare that with science, where categorising observed phenomena is fundamental. Psychologists, for example, name *schizophrenia* some set of observed behaviours, some other set *autism*. From time to time there is disagreement and the list is changed. Some categories may disappear altogether, as was the case with hysteria, a main subject of study at the time of Charcot, now considered too sexist to be discussed seriously. Categories allow psychologists to lump together observed phenomena, to classify and to compare them in an ever increasing corpus. The category is a Peircean interpretant.

Is There Recursion in Language?

Is there recursion in language, then? Clearly, this question resembles the *love* example above, and we shall examine the language game in the next section. Meanwhile, we shall look into the many houses of the mansion of recursion.

Chomsky's Recursion

Recursion vs. Iteration

When I prepared my presentation for the Mons conference, I assumed the recursion we were going to discuss was the mathematical self-calling function: a function is embedded in itself, and the embedded variables are calculated when a stopping

[11] As was, for example, Jacques Benvéniste's *water memory*, the notion that water may retain a "memory" of substances previously dissolved in it.

[12] As happened to *aether*, the nineteenth century notion that space was filled with some physical medium.

[13] PI§24. Italics is Wittgenstein's. Bold is mine.

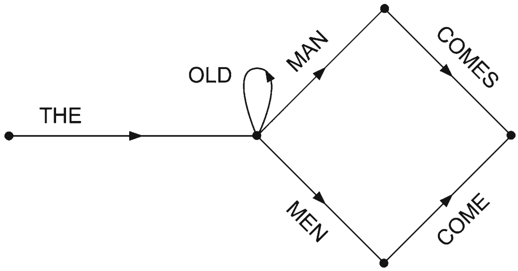


Fig. 14.1 Recursion as a loop in Chomsky's *Syntactic Structures*

condition halts the recursive process[14]. Clearly, there is no such function in language, hence my title. I also argued that Chomsky made a mistake when he introduced the notion in *Syntactic Structures* (1957: pp. 23–24). He thought of recursion as a loop, not as an embedded calculus, as is obvious from the picture above, taken from Chomsky (1957: p.19) (See Fig. 14.1).

Bickerton (2009) quite rightly points out that Chomsky's recursion is in fact iteration: the theoretical possibility of endlessly piling up adjectives in front of nouns is not recursion.

Recursion as an Expedient

I also argued that Chomsky brought in recursion in passing, as a sort of expedient. He admits that he resorted to recursion as a way to introduce infiniteness into an otherwise finite theory. Here is the passage:

> *We might arbitrarily decree that such processes of sentence formation in English as those we are discussing cannot be carried out more than n times, for some fixed n. This would of course make English a finite state language, as, for example, would a limitation of English sentences to length of less than a million words. Such arbitrary sentences would serve no useful purpose, however… If the processes have a limit, then the construction of a finite state grammar will not be literally out of the question, since it will be possible to list the sentences, and a list is essentially a trivial finite state grammar.* ***But this grammar will be so complex that it will be of little use or interest. In general, the assumption that languages are infinite is made in order to simplify the description of these languages. If a grammar does not have recursive devices*** *(closed loops, as in 8, in the finite state*

[14] For example, the **Fibonacci series** can be programmed recursively. By definition, the first two numbers in the Fibonacci sequence are 0 and 1, and each subsequent number is the sum of the previous two. In mathematical terms, the sequence F_n of Fibonacci numbers is defined by the recurrence relation:

$F = F_{n-1} + F_{n-2,}$ with seed values $F_0 = 0$, $F_1 = 1$ (adapted from Wikipedia).

Please note that recurrence can also be achieved by iteration, i.e. a loop. Iteration and recursion are very close and achieve similar results. Both iterative and recursive programmes necessitate a stopping condition to terminate the computing process. If not, computing goes on until the computer is switched off (in the case of iteration) or until the memory is full (in the case of recursion).

> *grammar*)[15], ***it will be prohibitively complex***. ***If it does have recursive devices of some sort***, ***it will produce infinitely many sentences***.

In *Syntactic Structures*, recursion is clearly a theoretical artefact (an anthropological object!) used to produce a grammar that is simple and interesting. It is not presented as a feature of language. There is no ontological claim at this stage in Chomsky's mind.

A Sociological Argument

I thought that these two arguments (recursion as iteration and recursion as an expedient) would help me rest my case quite easily. It did, but not for most participants. I had anticipated some reluctance, and this is why I had prepared a sociological argument. If languages are recursive, then linguists study a very interesting object indeed, and the status of linguistics is on a par with other more experimental and more formal sciences. Language is then a mathematical entity which can be transcribed into a computer programme, thus helping pave the way for artificial intelligence. And linguistics did indeed acquire a place between psychology and cybernetics, three domains which later joined to form a new science: cognitive science.

But could it be that there *is* something in language which can be interpreted as recursion in the light of some default philosophical notions? I shall presently analyse the other participants' papers and try to understand what they call recursion and why.

Recursion at the Mons Conference

It seems many authors take the existence of recursion as an established fact even if they disagree on some aspects of it. For **Corballis**, recursion did not appear suddenly in the human species but evolved gradually. **Freidin** explains how the recursive structure-building *Merge* operation was derived from transformational grammar in modern generative grammar and why it is now considered as the sole structure-building operation in the model. For **Fitch**, mathematical recursion is derived from linguistic operations, and he looks for other domains where such a derivative has taken place, i.e. in the visual and musical domains. He intends to show that recursion is a very general human feature. This would be an argument in favour of the claim made by Hauser, Chomsky, and Fitch (2002) that the essential difference between animal and human speech is recursion. Fitch indeed engaged in discussions with primatologists as to whether or not primate cognition is recursive. **Demolin** thought he recognised recursion in some primates' vocalisations, which would undermine Hauser, Chomsky and Fitch's claim. Conversely, **Lemasson and Zuberbühler** maintained they found complexity in non-human primate sound combinations but not recursion.

[15] See Fig. 14.2.

Some researchers led experiments to see if they could find evidence of recursion in the brain. **Gervain** found that newborns could distinguish between AAB and ABB vocal sequences, but she stopped short of calling this ability recursive. **Friederici**'s brain imaging experiments show that language is a highly complex process which involves many parts of the brain in a sort of swirling feedback loop. Her claim is that such a loop could be evidence of a brain basis for recursion. But the interrogation in the title of her presentation ("Is there a brain basis of recursion?") clearly means it is only a hypothesis.

Wautié, **Fortemps and Lowenthal** have designed computer simulations of recursive exercises based on nonverbal communication devices (NVCDs), which "may favour the emergence of new cerebral abilities" in people who have suffered brain injuries or strokes. In one of the experiments, a patient "is invited to construct, physically or mentally, a circuit performing" a sequence shown to the patient at the beginning of the experiment. Complex sequences are easier to reproduce with feedback loops.

Some researchers describe the possibilities of recursion in machine learning. **Dienes**, **Rohrmeier and Fu** explore artificial context-free grammars which feature left and right as well as centre embedding, and they think that such structures can be found in humans as well. **Cleeremans** examines recursion within the connectionist paradigm in neural networks. As for **Dubois**, he thinks that intelligence could emerge in machines using a fractal (therefore recursive) language.

Some participants have not directly addressed the question of recursion. **Fong** studies the computer implementation of unification-based derivations within the framework of the minimalist program; **Lefebvre** describes his research within the NVCD paradigm; **Paquier** addresses the issue of "cerebral neuroplasticity in brain-damaged children and adults from a clinical perspective".

Levinson takes a critical view of recursion, calling it "a lamp for the linguistic moths", "an obsession far from the centre of what linguistics should be focussed on". He claims that some languages allow embedding possibilities but never at more than two levels deep. If the reader now takes a look at the Anthony Burgess quotation put in as an epigraph at the beginning of this paper, he will have an example of a two-level-deep embedding. He will probably agree that the sentence is barely comprehensible. Of course Burgess's intention was not clarity there but wit, and this sentence is a fine example of Burgess's linguistic sense of humour. Embedding seems to be used here solely for its unusual character and the expressive power it provides. This suggests embedding is not a common feature of speech at all.

Why Recursion Then?

I could not agree more with Levinson: recursion distracts the linguist from more central tasks. But why has recursion gained such a following? I already mentioned sociological reasons; I shall now put forward more philosophical arguments.

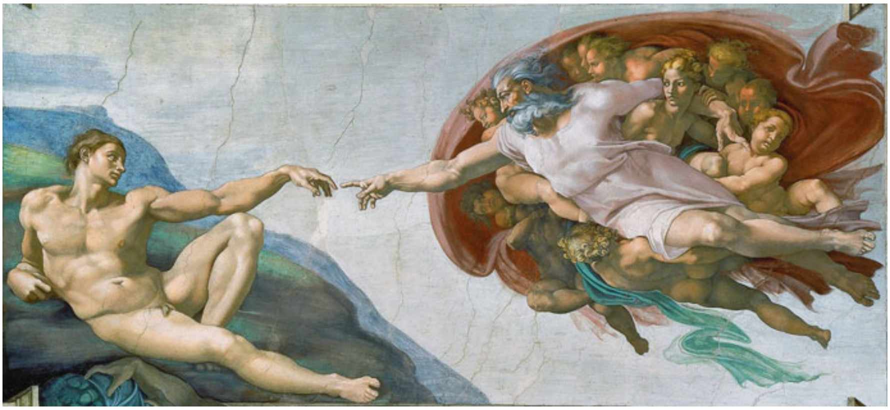


Fig. 14.2 Michelangelo's *Creation of Adam*

Recursion is clearly an anthropological object. Let us have a look at the language game:

– *Is there recursion in language*?
– *I don't know*. *What do you mean by recursion*?

I argued in Part II (Language Games) that a word like *salt* has anthropological *and* real-world existence. Is it the same with *recursion*? The Mons conference has clearly shown that there is no evidence for real-world existence. But there is agreement on meaning: recursion is generally thought of as a loop, possibly a feedback loop, sometimes with shallow embedding possibilities.

Is recursion a mathematical object? It certainly is within a mathematical theory. It is clearly not within the cognitive context, an entirely different domain. *Recursion* is a borrowed name, taken from a mathematical language game and used to name something else, to lump together a variety of objects and put them into a category. Why then not call these objects what they are, i.e. *loops* and *embeddings*? One reason is semiotic: once we have a name (*recursion*), we surmise there must be objects, and we are then keen to find them. Another reason is the causal effect attributed to categories (see Part 2). If recursion exists, we have a good "explanation" of what is so specific about human language (this is the gist of Hauser, Chomsky and Fitch's claim) and possibly transfer it to machines.

There is an ontological quest there, maybe some sort of theological desire to pinpoint our very nature, a need to find a primary cause to our human specificity. And if such a primary cause as recursion could be linked to some genetic configuration, we would then have a very powerful and naturalistic explanation of the essence of mankind.

Powerful indeed, but quite simplistic. Unless there is suspension of disbelief, a primary cause begs the question of what caused the primary cause. Michelangelo's "Creation of Adam" (see Fig. 14.2) shows how man acquired a soul, but we can only accept its truth at the expense of curiosity: why did God do such a thing?

What happened when His finger touched Adam's? What flowed into his body? (Was it recursion?).

Ontological quests are a natural feature of mankind. Yet they often take us down a blind alley. We should always keep in mind that language has a sort of demiurgic power: naming things, whether real world or not, gives them a separate reality of a kind. This metaphysical feature of language in effect makes it possible for us to speak meaningfully about our experience. Without the categorising capacity of language, we would not be able to consider objects together, to see what they share and how they differ. We would live in the confusion of a world of instances. The flipside of the metaphysical nature of language is what Wittgenstein calls "bewitchment"[16], i.e. the naïve belief that language can be taken at face value.

Conclusion

"We must do away with all *explanation*, and description alone must take its place", Wittgenstein says[17]. What he means is that it is a fallacy to explain human phenomena by linking them to an explanatory ontology. Language cannot be understood at a purely individual level, as a series of more or less brain-based functions and operators, as a code we encode and decode. Language cannot be explained away by other sciences. On the contrary, it is language which is the basis for all other sciences, and as such, it can express *all* our experience. "Language is perfect"[18]. Language is an anthropological given, a common treasure, a repository of knowledge, a corpus of what has existence for us, the horizon of our understanding and the frontier of our experience.

References

Bickerton, D. (2009). Recursion: Core of complexity or artifact of analysis. In T. Givón & S. Masayoshi (Eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution, vi* (pp. 531–544). Amsterdam: John Benjamins Publishing Company.

Chomsky, N. (1957). *Syntactic structures*. Berlin: Mouton de Gruyter. 2002.

Frath, P. (2012). La conscience dans la théorie linguistique. In P. Frath, V. Bourdier, E. Hilgert, K. Bréhaux, & J. Dunphy-Blomfield (Eds.), *Res-per-nomen III: La référence, la conscience et le sujet énonciateur/Reference, consciousness and the speaking subject*. Reims: Editions et Presses Universitaires de Reims.

Hauser, M., Chomsky, N., & Fitch, T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

[16] "Philosophy is a battle against the bewitchment of our intelligence by means of language" (PI§109).

[17] PI§109.

[18] PI§98, TLP§5.563, P§I1.

Kleiber, G. (1994). *Nominales. Essais de sémantique référentielle*. Paris: Armand Colin.
Peirce, C. S. (1978). *Ecrits sur le signe, rassemblés, traduits et commentées par Gérard Deledalle*. Paris: Seuil.
Peirce, C. S. (1984). "Quelques conséquences de quatre incapacités". In *Textes anticartésiens*, traduction de Joseph Chenu, Aubier. 1ère publication : *The Journal of Speculative Philosophy*, 1868.
Wittgenstein, L. (1961). *Tractatus logico-philosophicus suivi de Investigations Philosophiques*. NRF, Paris: Traduit par Pierre Klossowski.
Wittgenstein, L. (1963). *Philosophical investigations*. Translated by G.E.M Anscombe. Basil Blackwell, Oxford.

Part VI
Synthesis of the Main Discussion Sessions

Chapter 15
Synthesis of the Main Discussions

Francis Lowenthal

Preliminary Remarks

This book is made of papers written **after** and **inspired by** the discussions held during the Mons conference on "Language and Recursion." Therefore, these papers do **not** exactly correspond to the content of the presentations. This conference was an unusual one: the organizers had decided that the most precious thing was not the simple communication of new results but the exchange of ideas during discussions. It was thus decided, after the presentation of new results by a panel, to maximize the general discussion time. The aim was that during each of these discussion sessions, the participants would continue the same discussion about all already presented results. In order to achieve this aim, all participants were asked to attend all the sessions, and not only the session where they were supposed to speak. Astonishingly enough, in these times of fast and short communication, most participants accepted to "play the game" until the end of the conference and thus to "lose time," exchanging profound ideas.

The exchanges were very rich: some were devoted to a deep analysis of the results presented and others to the elaboration of new ideas about the relations between language and recursion, as mentioned in the preface of this book. It appeared relevant to present here a synthesis of these discussion sessions. It must be clear that difficult choices had to be made and that, for the sake of clarity, exchanges referring to different sections of this book had to be regrouped, to be "merged" and placed in the same section or subsection of this chapter. Inside each subsection, we respected the chronological order of the exchanges to show how the ideas evolved and converged, or not! It should also be understood that all interventions had to be

In this chapter, the name of the person who is speaking is in bold.

F. Lowenthal (✉)
Cognitive Sciences, University of Mons, Mons, Belgium
e-mail: francis.lowenthal@umons.ac.be

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*, 195
DOI 10.1007/978-1-4614-9414-0_15,

summarized. This synthesis is also biased because it presents the discussions through the filter of one single participant, the author of this chapter. He claims full responsibility for all the bad choices and all the mistakes he might have committed when writing this part of the book. He simply hopes that all readers will get here a glimpse of the constructive atmosphere we had during our debates.

Why Recursion and Language?

This discussion session started with an attempt to define what we actually mean when we say “recursion.” It quickly moved to the role of recursion in language, first, for humans and, later, for nonhuman primates. There were then several remarks about the relations between recursion and the brain, from the point of view of both the neuroscientist and the neuropsychologist. Finally, there were ontological reflections. In the present section, the author of this chapter tries to present these different aspects in an orderly fashion!

Attempts to Define the Concept “Recursion”

Fong suggested a definition from the point of view of the computer sciences specialist. In this domain, one speaks of solving a problem by tackling smaller versions of the same problem, so the parameters keep going down until they reach zero. This implies a recourse to self-reference which creates problems for many people. **Fitch** insisted that there must be a difference between hierarchical and recursive structures. As far as he is concerned, a structure is recursive if any part of it can regenerate the whole structure. He believes that there is such a recursion module “inside human language” because there is a need to generate a “discrete infinity.” **Levinson** mentioned that for him a structure is recursive if there are center embeddings and that, for our purpose, it is important to study them at the pragmatic level. Finally, **Corballis** said that the relevant distinguishing elements are the Theory of Mind (ToM) and the Mental Time Travel (MTT).

Recursion and Human Communication

Language and Chomsky’s Hierarchy

Fong stated that, according to Chomsky, English cannot be described by a finite-state machine (FSM). He therefore referred to Chomsky’s hierarchy. In order to describe the human language, we must be able to get infinity from a finite description, and recursion is the best way to do that. **Fitch** agreed that it is interesting to discuss Chomsky's hierarchy: according to him recursion crosscuts that hierarchy. It is thus relevant to examine through different species whether or not they can go beyond the level of FSM. He reminded us that the Fitch-Hauser paper mentioned in the preface of this book makes a clear distinction between FSM and context-free grammar. He nevertheless admitted that a test based on A^nB^n structures is not a good test for recursion. But, according to **Gervain**, Chomsky's 1957 hierarchy and recursion is only one important element: Chomsky's later evolution must also be taken into consideration, when this author wrote about the role of psychological reality. **Gervain** also mentioned Skinner's parallel between the infant who is learning the language and the linguist who is describing it. **Frath**, who disagreed, wondered what would happen to a computer if it wanted to learn a language. He claimed that this could not happen since a computer is just a tool. He added that Chomsky does not claim that there is recursion in the human language.

Is Full Recursion Really Needed for Human Language?

Lowenthal, considering that human beings never use infinite sentences, wondered if it is really necessary to envisage the full strength of recursion in the human language and thus potentially in the human brain. Fitch had spoken of "a discrete infinity" and told the group that it was easier and more parsimonious to use recursion to describe this discrete infinity: **Lowenthal** wondered whether that actually meant that full recursion is really needed or that recursion is simply a name and a description technique invented by humans. **Frath** replied that, according to him, the full strength of recursion is not needed: it is only necessary to have the possibility to pile up a little bit. Recursion can be used as a metaphor, but there is no evidence and there cannot be any evidence of recursion in the brain. It is not because it is named that it exists: humans speak a lot about consciousness, but it does not mean that consciousness has a separate existence. **Vergauwen** reacted hoping that Frath did not mean that reality does not exist. For **Friederici** it might be the case that we do not need the full power of recursion. Nevertheless, according to her, language cannot be a simple string of words; otherwise, there would be no explanation for the results concerning brain activations. **Fitch** also accepted that recursion is only an explanatory concept borrowed from mathematics. He said that it was not a good idea to search for "recursion in the brain" and then reify recursion as if it were an essential thing. Nevertheless according to him, recursion is more than a metaphor; it is a scientific explanatory concept.

Languages Without Recursion and People with Recursive Pragmatics

Corballis asked for comments on languages without recursion or without embeddings, such as the Pirahã's language. According to **Fitch**, Everett examined carefully the situation: Pirahãs have recursion in several cognitive domains and in particular in pragmatics. Nevertheless, they do not show any evidence of syntactic recursion (e.g., center embeddings). There seems to be no equivalent of the word "that" in their language. **Gervain** observed that even without a specific word, there could be embeddings. She considers that it is confusing to observe that there is recursion in many domains, except the syntactic domain. While **Freidin** claimed that the Pirahãs do not have a PSG, **Levinson** claimed that the only recursion which is relevant for human communication is precisely that which is found in the pragmatic dimension of communication. He added that historical and cultural traditions are inherited and are more important than a universal syntax.

Recursion, Merge, and Nonhuman Animals

Specialists of the Minimalist Program insisted on the importance of the Merge operator. Their remarks served as a starter for a new series of comments.

Friederici noticed that, according to Fitch, primates use for their communications combinations of two elements (e.g., "give apple"): she asked if this is a sort of Merge. And should this be the case, what do we need more than Merge to build a full structure? **Fitch** replied that some of the models used for syntax and for complex hierarchical structures are not to be found in other primates: the nonhuman primates do not use Merge. These complex structures can be found in bird songs or in whale songs, but in that case, the complex strings produced by the animals bear no meaning whatsoever. On the other hand, one could find meaningful communication in Kanzi, the bonobo, or in a dog barking, but in these cases there is no hierarchical structure. In human communication one finds meaning and hierarchy. **Friederici** contested Fitch's dichotomy saying that nonhuman primates can use more than one argument and mean something. They can also create a combination which means something new. She wonders if in these cases, one should not speak of simple sequences and not of Merge: since it has been said that Merge is the universal basis of language, we need another condition for the human language. **Friederici** suggested to say either that the human language uses a recursive Merge or that Merge is not the universal basis of human language since animals can do a form of Merge. **Fitch** contested: he does not accept the idea that Kanzi has Merge. Kanzi uses two items and they play together the role of a request, but that is all Kanzi can do. This bonobo cannot imbed this request in something else, which would be of a higher hierarchical level. Should we however decide to call Kanzi's action a form of Merge, then it would be a very bounded Merge. **Freidin** immediately objected and said that there cannot be such a thing as "bounded Merge": once one has "bounded Merge," "unbounded Merge" comes for free. To get a bounded form of Merge, one would have to place a constraint on it. According to **Freidin**, when Kanzi puts his hand out and communicates "give apple," "give" is not a verb and "apple" is not a noun phrase. This combination of actions has a meaning but there is no hidden hierarchical structure.

The discussion continued and **Fitch** claimed the contrary: according to him, Kanzi has Merge, but it is reasonable to ask what the structure of this form of Merge is because combining two things into one bigger whole that plays some functional role is the simplest thing an organism can do. **Friederici** explained that she thinks that the form of Merge we observe in Kanzi is not the form of Merge we use to describe human language. According to **Gervain** the problem is different. Merge is not only the action of putting two things together in order to have a new meaning; it is the action of putting two things together and giving a new label to this new object in a hierarchical way.

Fong asked if nonhuman primates can communicate or understand a message such as "pass the grape from person X to person Y": this would require a two-place predicate instead of a one-place predicate. He then observed that if these primates are combining this message in one structure, than they are doing Merge twice: it would be an example of nontrivial bounded Merge. **Fitch** confirmed that there are experimental data showing that chimpanzees are able to obey orders telling them to choose between two persons, one who must get the food and one who must not. Primates can do it and dogs can do it. A question remains open: are these animals merging into a larger structure? **Freidin** claimed that if this were the case, the animals would have to be merging in a linguistic structure, but **Fitch** explained that one can always put enough restrictions on something and then say that what the animal is doing is not the human version. If one wants to speak computationally, one has to define an operator such as "bounded Merge" or "Merge exponent n" and know with certainty which species are performing this operator. The problem is then an empirical one, not a problem of definition. During this discussion one should simply clarify the definition in order to be able to make the required experiments. The real problem will then be to find out how we can ask Kanzi if he has built a structure with a two-way binary tree.

Recursion and the Brain

Friederici had mentioned that Broca's area plays an important role. Her presentation and the way she described the differences between monkey brain, human adult brain, and child brain provoked another set of questions and remarks.

Corballis noted that Broca is analogous to F5 in primates: it is a part of their neural system, which is active when these primates are grasping. He thus wondered if the relation between grasping and an isolated movement in apes is similar to the relation between sentence and word in humans. Could this mean that the core of ape grammar lies in the grasping movement? What would then be the relation between grasping and their neural system? **Friederici** explained that F5 is involved in grasping but that there are simultaneously activations in the peripheral region. If one looks only at F5, it is more like articulation, so it should be associated to the premotor area rather than to Broca's area. If one considers simultaneously gesturing, grasping, and language, one must in fact come to sequencing and structuring. Broca's area is involved in certain types of predictive structures which are not necessarily hierarchical. Broca is always involved when the subject is learning a sequence and then has to predict what comes next. This is also true for intentional grasping. Broca's area seems to be involved in many activities, and the grammar system is sitting on top of them functionally. So it must be clear that Broca is not only involved in linguistic activities, it is also involved in other types of sequencing and more specifically in predictive sequencing. fMRI studies together with ERP studies show that when the brain makes a prediction and gets the wrong input, the Broca's area is very active.

Syntax or Semantics, Which Comes First?

Dubois mentioned researches with kindergarten teachers on the development of intelligence in 3- to 6-year-olds. One of the aspects observed was the verbal language they used. He remarked that their language dealt essentially with semantic aspects of language, but not with the syntactical aspects. The subjects used only words having a meaning for them, without taking care of the grammar. This seems to contradict Chomsky's theory which states that there is a universal innate syntactical aspect of language. **Freidin** explained that one cannot separate semantic and syntactical aspects. In any case, Chomsky was only trying to speak about the syntactic structure of language. This author did not say that there is no semantics; he just did not have anything to say about semantic structures. **Freidin** did not believe that children from 3 to 6 do not have any syntax. **Friederici** observed children of that age. She noticed that as early as 2;6 years of age, children have some syntax in German. Sometimes, they do not end a sentence because its meaning is very clear from the conditions, but one cannot say that they do not have the syntax. **Gervain** added that in perception, even children who cannot speak do discriminate syntactic patterns typical of their own language from those that do not appear in their native language. **Jacquet-Andrieu** objected that semantics is anterior to syntax in human beings: the child expresses a correct sentence only when he or she has understood the meaning of the verb which defines the roles of the participants in the conversation. **Freidin** noted that there must be a confusion. Chomsky was only talking about a system of knowledge, relating phonetic and logical forms, and observed how this system is put to use and applied. This has little to do with language acquisition in young children which is not a part of Chomsky's theory.

Recursion or Embeddings: Up to How Many Levels?

Levinson claimed that there is no unbounded recursion inside language—embedded clauses could be found up to level three but not beyond—but that there might be more levels in pragmatics. He thus suggested to bring everything down to counting languages. But Gervain, Fitch, and Friederici have presented results on brain activations. These results were based on structures similar to those mentioned by Levinson but did not belong to a counting language. The participants discussed this apparent divergence of opinion.

Counting Languages?

Levinson explained that he mentioned the counting languages but that he is also looking for homomorphisms between the structures in language and in this kind of formal languages: it will then be interesting to see to what kind of class the natural language belongs. One could examine different kinds of underlying structures and argue that they represent the nearest approximation to a particular natural language structure. **Fitch** replied that there must be recursion inside language. In any case, if there is recursion in pragmatics, as Levinson said, then there is recursion in language itself since pragmatics is part of language. **Levinson** reacted and said that Gervain and himself argued about this because there is recursion, but outside a language module. **Friederici** added to the discussion the following argument. She first considered the possibility of results based on counting on the basis of a first experiment: there were elements of type A and B, and the dependency between these elements was not hierarchically structured. Nevertheless, in a second experiment she took the necessary precautions to be certain that the dependency between those elements was bounded by a relation at a logical level such as voiced–unvoiced. Finally, she used actual German sentences, where the dependency between the subject and the verb is clearly installed in a hierarchical manner, and she showed that the brain region activations in the last experiment are similar to those observed for the previous ones.

More than Three Levels?

Freidin wondered how long it would take the participants to come up with a sentence that has four levels of embedding, while **Levinson** is saying that we can only get up to three levels. According to **Freidin**, human beings can easily produce sentences containing embeddings up to level four, five, or even six: they normally do not use such sentences, but they can create and understand them. **Frath** mentioned an example of such a sentence found in a book by Anthony Burgess: this author constructs a sentence at the end of which one has several occurrences of the same verb; these occurrences "pile up" because they all come from embedded clauses. He too added that humans normally do not create such sentences, except when they want to produce a special rhetorical special effect such as in "*J'ai vu l'homme qui a vu l'homme qui a vu l'homme qui a vu l'homme qui a vu l'ours*".[1] Although this process is not a natural feature of language, it can be used. **Frath** concluded that since

[1] Funny French expression. Its literal translation is "I saw the man who saw the man who saw the man who saw the man who saw the man who saw the bear." It is used to suggest hearsay evidence.

the beginning of the discussions, nothing has convinced him of the necessity of recursions inside language. Synthesizing what had been said, **Frath** agreed with Levinson that more than three levels of embedding is unreasonable in a corpus and that, as Jacquet-Andrieu mentioned, the apparent recursion in language is actually feedback. **Freidin** then offered a sentence with multiple embeddings: "I know they do believe that Mary will agree that John will be unhappy if Fred does not leave on time." When **Levinson** rejected this example, saying that it was peripheral, **Freidin** claimed that it was still recursion because there are clauses embedded in clauses: four embedded clauses into a main clause. **Fitch** objected that there is a confusion because here there is no center embedding.

Chomsky's Theory, Analysis of a Syntactic Structure, and Center Embeddings

Frath criticized Chomsky's theory because when one analyzes a syntactic structure, one has to know what sentence one wants to build in order to be able to start working on it: it is not the Merge function that does all the work. The same holds for syntactic structures such as NP + VP. Humans have to know where they are going beforehand, so since they know that, why bother with the process? **Freidin** replied that the knowledge Frath is referring to is in fact his knowledge of a given particular linguistic expression: it is its structure. He then must relate its phonetic form to its logical form. But, according to **Frath**, this is only valid once we are in Chomsky's theory. He thus asked what would happen should we not be inside this theory. For **Freidin**, what really matters is not Chomsky's theory, but the fact that linguistic expressions have a syntactic structure. If one does not accept this, one can account neither for structural ambiguity nor for displacement. Syntax is not "WYSIWYG": there are many covered elements. The derivations linguists use simply enable them to explain how one gets the representations: these derivations concern the people's knowledge, what they actually do when they pronounce a sentence. **Levinson** made clear that even with human languages, there are uncertainties about the best way to analyze some of these peripheral adjunctions: researchers do not always know whether they represent a subordination or not. For these reasons center embeddings are attractive: their status is always clear. **Fitch** agreed with Levinson on one main point: linguists are focused on structure. He then declared that people who study animals do not have that luxury: they do not have a microscope enabling them to look into animals' brains and see what structure they are building. These researchers are thus stuck with strings sets. This is why the Chomsky's hierarchy for formal language theory is a useful tool for them. This theory is not necessarily the only possible one nor is it necessarily the ideal theory for the study of language, but it is a useful tool to study animals. There remains nevertheless to figure out if A^nB^n is center embedded vs. cross serial in monkeys. **Jacquet-Andrieu** considered Chomsky's statement: "the mind is a system of knowledge." She believes that language is also a system of recognition. Recursion is thus necessary because humans want to look for past events. Language is also a production system; we need thus anticipation, which is parallel to recursion. **Friederici** replied that there might be a confusion about how people define what language is. It is clear that there is a difference between performance and competence, but Jacquet-Andrieu wants also to study the processing itself. It is a good idea but it is not the object of the present discussion which concerns which knowledge system language is. Moreover, **Friederici** observed that Freidin was talking about competence while Jacquet-Andrieu wants to examine performance. These two aspects should not be confused. **Jacquet-Andrieu** reacted by saying that it was clear in her mind: according to her there can be no competence without performance and vice versa.

Recursion and Pragmatics

Dubois noticed that one of Levinson's examples for recursion in pragmatics was a repetition of an adjective, which he considers of very limited interest. He thus wondered if Levinson had a more general definition of recursion in pragmatics. **Levinson** explicated that he suggests that recursion is not what researchers in the field of language are really interested in: they are in fact interested in complex structures in the human language. They also want to know whether these structures are specific to the language domain or not. Central embedding seems more interesting because it seems to give the required expressive power. He added that, according to him, most of the experiments by Friederici and Gervain are based on this kind of underlying structure. He thus claimed that recursion, as defined by Fong and as originally described by Chomsky, is not an interesting phenomenon in cognitive sciences. On the contrary, creating structured strings based on center embedding would be a worthwhile goal.

About the Minimalist Program

What Are the Aims of the Minimalist Program?

Freidin explained that the Minimalist Program simply means that the researchers want to have a minimal number of computation mechanisms. This program applies to two different aspects of language. The first one concerns the extent to which the computational system of human language is optimal and the other one concerns the extent to which the human language is a perfect system. According to the Minimalists, the human language looks optimal because there are very simple computational operations: all language transformation rules are abandoned since they belong to one specific language only, but case filter is considered as being a part of the universal grammar. The Minimalists have selected computational mechanisms which apply crosslinguistically to many languages. Various constraints that apply to all human languages explain the history of the program. Before the aim is completely reached, one must be more specific about what the language faculty is.

Merits of the Minimalist Program

Frath, trying to evaluate the merits of the Minimalist Program and of the Merge operator, wondered how this could account for the constraints observed by Sinclair for specific instances such as “set in”: this verb has usually a negative subject and is regularly placed in a clause which is final. He added that if one could account for these constraints with Merge, he might consider that this theory has something to offer. **Frath** added that the English speaker is unaware of the fact that one must have a negative quality for the subject of “set in,” but if one asks “*can you say happiness sat in*,” people would say no. **Freidin** then reacted and said that this simply means that the English speaker has some linguistic knowledge; knowing what is that linguistic knowledge and how it is substantiated is another question. In order to account for the constraints observed for “set in” by Sinclair, in terms of Chomsky’s theory, **Freidin** explained that since one wants to have a negative subject, one can assume that “set in” assigns a semantic function to its subject and that this semantic function has some negative aspect attached to it. This assumption belongs to the field of lexical semantics; it is idiosyncratic to “set in” and it is not a general property of verbs. In as much as the final clause constraint is concerned, it simply means that “set in” is intransitive.

Can a Minimalist Parser Work on a Large Corpus?

Gervain wondered what Fong’s parser provides when applied to a large corpus: does the model break down or diverge from the theory? **Fong** explained that his Minimalist Parser does not encode a certain amount of the information and can thus not be used on the general corpus. Another parser, the Principle Parameter Parser, contains about 30 different principles and has a much larger lexicon. This parser can be used on more constructions but not on a general text. This is due to the fact that the linguistic theory has not yet progressed to a point where one is able to explain what the different constructions existing in one particular language are. It is even more difficult if one considers all languages. It is thus possible to build a parser that can analyze the general corpus, but only part of this parser will be based on linguistic principles because the theory is still incomplete. In terms of the Minimalist Parser, it is interesting to observe that we have on one hand the theory and on the other this parser which can handle a few hundred parameters but remains incomplete. The Minimalist Program is a moving target: there is a lot more work to be done before a parser based on this program can be applied to the general corpus. **Levinson** completed Gervain’s question by remarking that Fong used another kind of formalism to implement grammar. He thus wondered if this is not the homunculus, creeping in and actually doing the parsing work for the human being, whereas in principle it should be Merge itself which would do “the job” while the parsing is done. According to **Fong** one must distinguish the implementation from the actual mechanism. Often a mechanism is available and can be implanted in different ways. In this case one uses a computer system which does not resemble the brain.

There are thus various ways in which one can implement Merge. One way consists in saying that there is a verb and a noun which can be merged: this can be encoded in various ways. As far as agreement is concerned, the simplest possible mechanism would be unification. All this has empirical consequences: this process might make the linguists aware that these mechanisms are the primitive ones and will reduce the number of computations needed. Should a feedback from linguists be needed, it is possible to implement these mechanisms and show where they break down. This will bring to attention the other mechanisms from other fields, like computer sciences, and could provide other approaches to the solution.

Language Acquisition and the Minimalist Program

Dienes wondered how language acquisition works in minimalism. There must be "language-specific" facts: he asked how children were learning these facts. As far as what has to be learned, **Freidin** expects that it will be possible to reduce what has to be learned to parameters, and parameters are a range of things. He clarified what he meant by saying that there are ways of saying things in French which cannot be used in English, etc. but this is not specific to minimalism.

Recursion and Philosophy

Vergauwen mentioned that every recursive function can be simulated by an iterative function and vice versa. This provoked some valuable reactions and the discussion took an unexpected philosophical dimension.

Does Recursion Have an Ontological Value?

Vergauwen, considering what is stated above, asked if there is an ontological point in saying that there is recursion. **Fitch** answered positively. Iteration in itself is ubiquitous in the animal kingdom but it is not enough to give rise to language. There must thus be something which marks the difference between human beings and other species. What is it? If we choose to say that it is only iteration, then we have to say what the special kind of iteration human beings have is. We can call it the way we want, but there is a difference and it is our job as scientists to explain this fact.

According to **Frath**, recursion has no ontological value. He justified this by his opposition to dualism. According to him, human beings are never isolated; they are part of a community. This community has ways of doing things and a language. These facts entail that concepts are not linked arbitrarily to a word: by giving a name, one gives a proof of existence. It is thus impossible to separate the body and mind. Human beings are tricked by language as it is and they should not go on an essentialist quest: it would be a mistake because concepts do not exist independently from the human beings using them. **Martins** responded by saying that when he refers to a materialistic view of the mind, it is only because humans can make some predictions. If a person is unable to behave like that, then that person must have some specific impairments in his or her mind. **Frath** reacted and said that this does not imply anything about the dualism mind-body. **Dienes** noted that if Frath is correct, the major role is left to the cultural environment: in that case one should wonder what the actual difference between the chimpanzee and the human being is. **Frath**, insisting on the role of a community, mentioned the "sentinel" bird warning the group of a predator's arrival. He also said that these animals have their way of life which does not lead them to our linguistic capacities. **Gervain** tried to conclude that should one give up all the existential claims, because one worries with the concept's actual existence, then it is not only recursion that goes down the drain but also many other elements, like phonemes. She added that recursion generates more discussions than other concepts like phonemes, so there must be a distinction between recursion and these other concepts. **Frath** ended this part of the discussion by noting that everything depends on what is meant by the word "exist." Some words can be used as noun and as adjective: the two categories overlap. Yet, nouns and adjectives are not used in the same way. Romans created the "zero category" to refer to the nonexistent articles: if recursion has this function in language, then, as explanatory concept, recursion has no ontological value.

The Principle of Parsimony

Vergauwen imagined the following situation: assuming that we have two theories, one based on recursion and describing completely what is going on in the brain and another one explaining the same observations but using only iteration. He then asked Fitch how he would choose between these two theories. **Fitch** replied that if the two theories explain exactly the same things, we can just say that they are equivalent and nothing else. **Vergauwen** wondered if Fitch would say that one is more true than the other or if he would simply choose the more elegant one. **Fitch** reacted by quoting the Ptolemaic and the Copernican systems: both are able to predict the movement of the planets, but one is simply more parsimonious. Parsimony must play a role in our decisions about scientific hypotheses and theories.

Recursive Cognition as a Basis for Language

Where does language come from and how did it originate? Some say that there was an important mutation in our ancestors which provoked a rewiring of the *Homo sapiens* brain. In that case, one individual could have gotten the language and transmitted it to the others. Corballis has a different point of view. Language is a complex ability and cannot have occurred in a single jump in a single individual. According to him, before the emergence of language, two precursors appeared: the

Theory of Mind (ToM) and the Mental Time Travel (MTT). On the basis of recent results, Corballis recently (July 2013) told the author of this chapter that he is now increasingly convinced that there is greater continuity between humans and animals than he previously believed. Maybe the essential difference boils down to language.

Anticipation and MTT

Dubois reacted to Corballis' 2011 claim that imagining future events could be a faculty that is unique to humans and that animals are not able to create future events. He described a cat's anticipatory behavior and concluded that this animal can imagine the future and thus that anticipation is also a property of animals. According to **Corballis**, animals can anticipate, but it is only short-term anticipation. Animals do not seem able to create and imagine a scene several weeks in advance. They can anticipate the presentation of food and thus reach a little bit into the future, but MTT requires the construction of new events that can happen at any point in the future. Humans always construct vast futures: they are constructing careers, planning parties, imagining what might happen at an interview with the boss, etc. There is a generative component in humans that is not present in the short-term animal anticipation. **Dubois** objected that he observed an experiment where birds perform successfully three successive tasks in a correct order. Birds can also build their nest before they have the eggs in it. There are thus many examples of anticipation on a rather long-time distance. However, **Corballis** did not accept this as evidence of MTT. He said that migration of birds and nest building are instinctive: a bird builds a nest without imagining first building a nest and then the eggs been placed in it. He admitted that this is controversial and he described a peculiar experiment concerning birds hiding food: if another bird is watching, the first one will remove the food he just got and hide it somewhere else. One could argue that these birds are imagining a future where their food will be stolen, or one can simply say that their behavior was a conditioned response. **Fitch** objected that many people in the field view this experiment as a strong evidence of MTT because they observed that it does not matter whether the bird had experienced being stolen from or not: what matters is whether they themselves have been thieves or not. Only those who have stolen food actually move the food. It is thus not only an association but also a combination of their own recollection of stealing with this future-oriented attitude of caching. According to **Corballis** there is a very subtle distinction between acting for the future and imagining what that future will be. Animals do not invent multiple futures, but human beings can describe several possible futures and perhaps select one among them. **Corballis** insisted that in animals, at least in birds, the generative component is missing. It is thus hard to conceive whether an animal is actually thinking about the future or whether it is simply following a pattern that leads to a reward. There are also instinctive behaviors.

MTT and Neural Networks

Dubois mentioned the properties of a neural network. The neurons in the animal are the same as those in our brain, and one can construct small programs which have anticipation. He thus wondered why the animals who have the same neurons as human beings do not have anticipation: he believes that the ability of projecting into the future is a property of neurons. **Corballis** explained that MTT is not about a projection into the future; it is about constructing and imagining a future. He also added that these questions can only be solved on the basis of further neurophysiological evidence.

Can There Be MTT in Other Animals?

Lowenthal asked if dolphins, who also communicate acoustically, have at least a first-order ToM, as described by Corballis, and some kind of MTT. **Corballis** replied that it is obvious that dolphins communicate, but it is not certain that dolphin A communicates with dolphin B, on the understanding that dolphin B understands what is in dolphin A's mind. There is not a certitude that they have the flexibility to communicate in that fashion, but they certainly communicate. **Corballis** also mentioned that there is a sort of ToM in other animals as far as empathy is concerned, but he said that an animal would need one more level of ToM to be able to really pick up another animal's mental state. A human being has the ability to go up that extra level of ToM, while an animal does not. **Corballis** also added that he has no evidence that dolphins have MTT in the sense that they would be able to bring to mind a future event.

MTT, ToM, and the Complexity of Human Language

Gervain wondered, from a linguist point of view, if Corballis' theory of language coevolution with other precursors predicts the actual complexity of language. Language is actually good for communicating about and for imagining a future or a past event, and it is also useful for the ToM. But why did the grammar have to become so complex? She mentioned the ease with which humans can communicate with one another if they simply know a few words of a shared language, without knowing all the subtleties of its grammar. **Corballis** replied that human lives are exceptionally complex. Humans live in groups of 140 people, which is a complex network to start with. They do complex things and are manual creatures. They move a lot so they do not have a fixed territory. There is thus complexity in all the things humans want to describe. When they move in time, and thus in place, they introduce more complexity. The complexity of human life is rapidly building up. He said that he is inclined to think that human communication started with pantomimes but, since human lives grew complex, humans had to invent some conventions and thus a grammar. Once one starts to break events down, one must have ways to combine them. The more conventionalized ways of combining humans require, the more they need abstract grammars everybody understands. In creating such grammars, they create the relations and the symbols they need. **Freidin** remarked that grammars can create ambiguities. He gave as simple example the sentence "John thinks that he is clever." This sentence has two possible meanings: either John thinks that he-himself is clever or John thinks that he-another is clever. He claimed that one cannot say "John thinks that he-himself is clever" in English. **Lowenthal** mentioned that in some languages, such as Iranian, it is possible to make an explicit difference between he-himself and he-another. **Corballis** replied that this ambiguity problem was not what he was discussing. He simply wanted to argue that grammars with this kind of problem could be the result of simpler systems rather than the result of a mutation that happened 100,000 years ago.

More About the Big Mutation or About Many Small Mutations ?

Freidin recalled that when Corballis talked about externalization of language, he mentioned thought and said that thought is generative and has a recursive structure. According to **Freidin** this is not the answer to what Chomsky is saying. One could say that there was a mutation in which humans got the recursive Merge and which helped them thus to organize their thoughts. It seems difficult to say that these thoughts existed before recursive Merge existed. **Freidin** commented that we know what language and thoughts humans have now, but there is no clear information about the way it was for their ancestors before they had the externalization of language. It could thus be that the recursive Merge triggered the whole process. **Corballis** explained that he does not disagree with Chomsky as far as the description of grammatical structures is concerned. He nevertheless disagrees with the idea that the recursive Merge could have come about as the result of a single mutation rather than as the result of a series of small changes going back to million years. These changes could be related to the growth of the brain and increase of social structure. It is possible that it started with the first pantomimic communication systems and gained complexity over the course of time. **Freidin** objected that the single mutation did just one thing: it gave humans the recursive Merge. Everything else, in the field of language, followed. He clearly disagrees with Corballis' idea that the recursive Merge had to be broken down in little bits that would have come by with time and that would have combined to give the present recursive Merge.

Freidin then objected to the Corballis' point about the ToM. He observed that people have different knowledge of the world, different beliefs, and different ways of reasoning. They also have different emotional responses. This creates a rich cognitive stew out of which the participants construct their thoughts, which are then translated into language. Language becomes then a form of sound wave from the speaker to the receiver. Once the utterance is out of the speaker's mind, it does not have that rich cognitive stew anymore. It is not until the receiver takes it into his mind that this utterance reacquires a rich cognitive stew as support. But then, everything happens in the receiver's mind and not in the speaker's mind. He concluded that a shared stream of thoughts is suspect in normal conversation. **Corballis** responded that it varies over time and environment. He mentioned teenagers: they share a lot of knowledge in silence.

Fitch wondered if mutations of large effect are more common than mutations of small effects. Everybody believes that only mutations of small effect were common, until about 20 years ago when researchers started to discover how evolution and development work together. Presently, people speak about genes: they know that one can get a single mutation by changing a single amino acid in one of these high-level master control genes. This apparent small change has a big effect on the phenotype.

ToM and the Affective Dimension

Jacquet-Andrieu asked Corballis how he explains the affective incidence on the construction of the ToM. What is the role of emotions and affection? How is it encoded? **Corballis** explicated that the ToM probably came out of empathy, which is an emotional response. This emotional component of the ToM exists but is not the only part of the ToM. One can usually know what other people are thinking independently of any emotion. It is possible that the ToM may have emerged from something that one sees in nonhuman primates, such as empathy, which is the understanding of the emotional state of another individual. **Corballis** agreed that animal behavior can be directed partly by emotions. **Jacquet-Andrieu** claimed that the ToM implies a communication with another person. It is an understanding between two individuals, not only an instinctive understanding but also an affective understanding. She claimed that this fact is very important because people with autism have no ToM. **Corballis** admitted that the ToM is partly a theory of the emotional mind, understanding the emotional state of others and perhaps responding emotionally to that. But the ToM is also important when one plays chess: one has then a different kind of ToM, unemotional.

Recursion, Memory, Time, and MTT Together with ToM

Lefebvre mentioned, from the point of view of the neuropsychologist, that there must be an episodic memory and a semantic memory, because otherwise one would lose the spatiotemporal environment when remembering through the semantic memory. Is it possible to explain some links between recursion and this process? **Corballis** said that he believed that episodic memory cannot exist by itself, because if one makes an episodic memory of something that happened in a different city, one must have some built-in semantic knowledge. Thus, it does not seem that episodic memory is completely independent of semantic memory. This holds also for plans concerning future events. An argument in favor of these hypotheses is that in semantic dementia, episodic memory seems intact. **Lowenthal** argued that episodic memory seems to be strongly linked to the ToM: when a person thinks about going to Paris next month, that person is also thinking about what is going to happen to him and also about what Paris is going to do to him. **Corballis** replied that in this case, it is partly the theory of that person's own mind.

Lefebvre also asked if there are links between time, perception, and recursion. **Corballis** explained that, according to him, our MTT is pretty inaccurate. Human beings do usually not know exactly when an event occurred in the past, unless they have explicitly put a date on it. The time line is more or less fuzzy. Thus, MTT involves the sense of time, not so much the sense of time going by as when one watches his clock, but the sense of time in the broader sense: one can situate events and know that they happened last year or the year before. This sense of time is different from the actual perception.

Communication in Nonhuman Animals

Human language is so complex than it cannot be the result of a single mutation, or can it? Some researchers believe that it can only be the result of a long evolution and that a good manner to gather information about what possibly happened is to look at potential language precursors in contemporary primates. Zuberbühler and Lemasson (2014) describe observations they made concerning acoustic flexibility in Campbell's monkeys. Demolin described recent findings in the vocalizations of Muriqui monkeys showing that these monkeys build part of their communication on utterances recombining a set of limited elements: Muriqui monkeys can produce syntactically well-formed strings, which include recursion, but these utterances should not be compared to what is found in human communication. The results of these two groups of researchers suggest that the emergence of recursion in primate communication might not only be the consequence of genetic factors but also of social and ecological factors. This contradicts assertions made by Fitch concerning the uniqueness of human recursion as the result of a single mutation. All this resulted in a very deep debate among the participants.

About Campbell's Monkeys Verbalizations

Do Campbell's Monkeys Produce Hierarchically Organized Verbalizations?

Fitch asked how the researcher can demonstrate the distinction between a structural combination vs. a simple flat serial concatenation. With humans, one can rely on meaning, but this is not the case for nonhuman primates or for birds. **Lemasson**

agreed that with animals, one must deal with the context present at the moment of the calling: it is not clear if animals can also communicate about another space or about another time.[2] Nevertheless, he mentioned evidences of meanings that are related neither to predation nor to emotion. Such evidences have been observed when the monkeys want to start traveling: the male calls the females and they all know, even if they do not see him, that they have to travel and in which direction. **Lemasson** claimed that, although it is difficult to assess the meaning as one would do with humans, there are evidences in favor of hierarchical combinations with meaning.

Conscience as Main Difference Between Men and Nonhuman Primates?

Dubois wondered what makes humans able to understand and speak a natural language while animals do not seem to have this possibility. There is no big difference in DNA between man and ape. What is lacking in apes? Conscience? **Dubois** also noted that when a researcher tries to teach human language to apes, he or she fails. He suggested as major cause for this difference the fact that humans are the only ones to have a conscience. **Lemasson** replied that, according to him, the main difference is the fact that humans can talk about space and time. The fact that humans have an infinite way to create new meanings is another difference. This last faculty is clearly nonexistent in nonhuman primates or only in very rare occasions. **Lemasson** refused to enter the conscience debate because it is not his expertise. Anyway, he believes that the real reasons belong to the field of cognitive abilities. He also mentioned that there are differences between monkeys and apes. Apes are able to deal with much higher members of communicative signals: they can do more vocal signals than monkeys. They can also perform a lot of gestures that they combine with sounds, so they deal with more modalities of communication. Two parameters must be taken into consideration: the cognitive ability and the social needs. A comparison among apes of species that have very simple social structures like gibbons and others that have more complex social structures like chimpanzees shows that they have also very different repertoire sizes. His conclusion is thus that the difference between man and primate communication systems is a matter of cognitive ability and aptitude to innovate.

Gestures and Verbalizations

Corballis wondered if the monkeys observed by Lemasson produced manual gestures associated with the sounds: did the gestures he mentioned occur independently or were they combined with the vocalizations? Did they have any meaning? According to **Lemasson** it depends on what one means by gestures. The monkeys observed did not exhibit hand gestures, but they use other visual signals such as very

[2] This could be associated to a lack of MTT. Martin-Ordas et al. (2013) claim that chimpanzees and orangutans might have a form of MTT, at least unconsciously.

limited postural gesture and some face mimics. They always associate these "gestures" with sounds, but they use them very rarely. **Lemasson** also mentioned a new study of pointing with red-capped Mangabeys. These monkeys do not use visual signals, probably because they are forest animals. Since they are as good as chimpanzees in learning pointing gestures, a question arises: are they avoiding the visual signals because they cannot do them or because their habitat is not adequate for that kind of communication?

Vocal Innovations

Lemasson had described a vocal innovation created by the captive monkeys he observed: these monkeys created a new alarm code that the wild ones do not produce. It signals those, among humans, they are scared of. **Gervain** wondered if this call can be derived from any of the calls they produce in the wild. **Lemasson** replied that it is clearly a derivation from another alarm call, but it is longer and has a higher peak frequency. Both calls are different but derive from the same basic structure.

Are There Different "Campbell's Languages"?

Gervain also asked what happened if a member of one captive group was transferred into another distant group: would these monkeys still understand each other or are there different "languages" in monkey communication? **Lemasson** replied that it is difficult to answer that question. One individual was indeed replaced in the group. But this completely changed the relations inside the group: after the replacement of the individual, the social network was completely disturbed, the previous binds totally disappeared from the group and new binds appeared. Communication was totally disturbed. But he can relate this question to another experience. He exposed the monkeys to two different types of acoustic signals. Some signals were understood by the animals. They do not behave in the same fashion for, on the one hand, a foreign signal or a non-signal or, on the other hand, for a current signal or an old-fashioned one not present in the group anymore. This seems to indicate that, at least at the group level, there are some social signatures.

Are These Vocalizations Based on Verbal Reflexes?

Vergauwen wondered if there is a difference between the vocalizations of Campbell's monkeys and some kind of verbal reflexes. According to **Lemasson** some researchers might say that these verbalizations are only emotional reflexes because they look as if they were motor cortex controlled, at least these are the data from neurobiological evidences. Nevertheless, the behavior evidences show that the theory of emotional reflex cannot explain everything. There are studies showing that these monkeys can actually target one specific individual. They also behave differently according to the context. The acoustic structure they produce cannot be explained only by degrees of emotion: signals with different meaning are acoustically very different from one another. But the speed of call delivery is probably of reflex-emotion type.

About Muriqui Monkeys Verbalizations

Muriqui Monkeys Have a Low-Level Recursion

Demolin explained that Muriqui monkeys generate typical utterances. They can use discrete elements in their vocalizations. There is a prosodic pattern used by the whole group. These utterances are not stereotyped and they respect a rule (which can be considered as a formal grammar). This grammar cannot be generated by a simple class of finite-state grammars. The data enable **Demolin** to conclude that there is a low level of recursion in Muriqui monkeys. The grammar must thus be a context-sensitive grammar but not a context-free grammar. This yields a regular language. This fact does never occur in humans. **Fitch** wondered how Demolin determined that there is a real structure in the verbalizations of the Muriqui: it could be just a string of p's and t's. How did he establish that there is a subjacent hierarchical structure? **Demolin** said he used the prosodic structure as independent indicator, but **Fitch** asked how it is possible to know that this idea is not simply in the researcher's head, but well in the Muriqui's head. **Demolin** gave examples of appropriate utterances: one Muriqui vocalizes and another responds. Every animal is capable of producing utterances of this type and there are patterns which are repeated. The strings produced have the same duration. He admitted that other analyses are possible. **Fitch** specified that he did not hypothesize that the Muriqui produces random sequences. Nevertheless, he thinks that these utterances could be produced by a finite-state generator. They would thus not be random, but the order in it would be a serial "flat" one, and not one that has a higher-order structure. **Demolin** admitted that he cannot prove with his data that what he actually observed is a hierarchical structure. Nevertheless, there is a form of central embedding. But he accepted that this is not enough: it is not possible to discuss clearly the hierarchy without having anything to say about semantics. His researches leave him at a level where he can say with certitude that there is a structure, but he cannot prove that there is a hierarchy.

Muriqui Verbalizations, Merge, and Human Language

Freidin remarked that there is at least a minimal hierarchical relation between the utterances and the basic constituents. Furthermore there is some structure inside the sequences, but it is clear that it is not unbounded Merge.

Demolin stated that one cannot compare Muriqui communication with human language. The Muriqui communication is a result of their evolution after they diverged from humans. The question that must be examined now is if one can find similar results in other nonhuman apes. **Dubois** mentioned that the sounds of Muriqui communication made him think of the click languages. He believes that the evolution of language started with clicks only. He wondered if there could be any relation with Muriqui communication. **Demolin** replied that clicks are probably not the first sounds produced by humans. The Muriqui's communication system lacks several phonetic characteristics that human languages have: there are no consonants in Muriqui communication.

Corballis asked if all Muriqui utterances were on the same breath. **Demolin** acknowledged and said that there is no ingressive vocalization in the Muriqui communication. As far as the single breath is concerned he believes that these utterances are made in such a way that it seems to be one big planning in terms of breathing. The Muriquis do not retake any breath between parts of the productions, which is very different from the chimpanzees and the bonobos.

Nonverbal Approaches, Language Acquisition, and Recursion

The discussion following the session on Nonverbal Communication Devices lasted more than 1 h. Some parts of the discussion concerned more details about the experimental procedures used in researches based on NVCD-like approaches; other parts were clarifications concerning the results. Since this information served as basis for deep discussions, it appeared relevant to mention them briefly here before presenting the synthesis of the main part of the discussions.

Experimental Procedures Used for NVCD-Like Approaches

The experimental procedure used is described in details by the researchers. During a first step, each subject gets the material and plays freely with it. During the second step, the subject is asked to solve a very simple problem involving only an elementary construction with the device used: at this step the subject is confronted with the technical constraints of the material. During the third step the subject is required to make a construction based on a sketch and then to solve a logical problem. In the case of the Dynamical Mazes (DM), this step is devoted to the construction of a network (based on a sketch) immediately followed by the exploration of this network in order to discover its regularities. Finally, the subject is asked to make predictions about the order in which the exits are used, taking into account the evolution of the inner states of the network: first, short-term predictions (next exit), then middle-size predictions (two or three next exits), and finally long-term predictions (all next exits). It must be noted that long-term predictions are impossible for subjects who do not have a fairly good notion of number. These predictions can be done in two different ways: Lefebvre has observed that 6-year-olds only look at the list of exits; he has also noted that young university students were combining inner states and produced exits (the "if… then…" attitude), while young adults who did not study after the end of compulsory schooling (18 years of age in Belgium) behaved like 6-year-olds.

It must be mentioned that the research does not occur at the same pace with handicapped, young healthy, and adult healthy subjects. Nevertheless, the approach chosen is always a spiral approach, as defined by Bruner, and each exercise can be solved at each of the three levels of representation described by this author: by performing actions, by thinking in terms of images, or, finally, by using symbols to code the whole problems.

Details Concerning the Results Obtained

Lowenthal described two paradigmatic examples of research with handicapped subjects: the case of Saïd, a child who lost the ability to communicate in a structured way after a meningoencephalitis, and the case of Yannick the adolescent with fragile X syndrome and an IQ of 43. In both cases the subjects acquired after a certain time a form of structured communication: Saïd was able to communicate nonverbally but in a very expressive and structured way, using symbolic gestures he had created and written words; Yannick discovered, through DM, the meaning of the "if… then…" and was able to behave accordingly when reading very short messages despite his mental retardation.

Freidin asked for more details concerning the results with normal young subjects. **Lowenthal** explained that in this case there was always a control group. The young healthy subjects of the experimental group (EG) progressed in reading and in verbal communication. As far as verbal productions are concerned, there are significant differences (in favor of the EG) for three variables: the Mean Length of Utterances (MLU), the number of specific words used vs. the number of periphrases, and the number of relative clauses. He also explained that these differences are more important when the subjects are asked to describe a situation they invent (the invented end of a video which was shown incomplete to the subjects). **Lefebvre** added that 6-year-olds using the DM made significant progresses (when compared to subjects using the Pegboard or to the Control Group) in abstract visual-spatial ability but did not discover, through the DM, the "if… then...."

Influence of DM Exercises on the Brain

Lefebvre's fMRI observations of young university students and of young adults who did not study after the end of high school show different activations for the EG after DM manipulations, but only for the group of "educated" subjects. In fact, at the posttest, there is more activation in the basal ganglia for that group. **Friederici** mentioned that the circuit between basal ganglia and Broca's area is supposed to do the procedural knowledge work. She noticed that the major effect observed concerns verbs. She also remarked that, in the literature, the basal ganglia are often active whenever the subject has to do some learning. She thus wonders, with Fitch, whether the learning implied was of a very specific type or it is more a general type of learning. She thinks that the difference is not important for the therapy activities as long as there are progresses with this type of learning, but scientifically it would be interesting to see whether this is a general learning effect or a particular learning effect. To answer this question about the specificity of the training, it would be relevant to use some other training as well and then examine whether there is the same basal ganglia activation or not.

What About the Control Group?

Martins wondered what served as control group for the experiments with handicapped subjects. **Lowenthal** explained that it was not possible to have a control group for these, since all of them had very specific and very different impairments. This might be considered as a bias, but the number of subjects having undergone such a therapy counterbalances this lack of control group.

The participants discussed what could be a proper control for this kind of studies. **Friederici** said that the results presented show that the specific training used with NVCD-like approaches has positive effects. It remains to examine whether these results are really due to this specific training or could another training, not based on recursive structures, but based on simple sequencing or anything else, yield to the same results.

Lowenthal mentioned once more the studies with handicapped subjects: each of these studies is a case study and it seems nearly impossible to have the same type of validity as that offered by an experimental group vs. a control group. One can only observe that all previous training failed to provide positive results and that a long NVCD-like therapy helped the patients. One must nevertheless admit that it could be maturation, or simply the frequent presence of the researcher that did the job: it is possible that another training would have helped, but none was available! We can thus only conclude, for these studies, that something was done with the subjects and that the patient made progresses, but we cannot claim with certitude that there is a causal relation. One could probably say that the NVCDs played the role of precursor or that of trigger and that after an initial training, the rest came more normally. In the case of healthy subjects, the presence of a control group enables the authors to claim that the results observed are due to the training. Nevertheless they do not yet know exactly what this training is triggering. It is thus not possible to claim that this training is the only possible one.

Friederici insisted that her remarks are not meant as a critique. She tries to figure out the neurocognitive basis of this type of research. The actions the subjects have to do, when they manipulate the material, are important. These actions imply some type of action sequencing, and this type of activity involves some brain areas that are also involved in language processing. It is thus interesting, not only for therapy but also for basic research, to understand which brain areas are involved in these processes. Moreover, some of the subjects had brain lesions in the frontal cortex, including Broca's area. This fact must be examined since the literature shows that Broca's area is involved in action sequencing, in language, and in visual-spatial processing of sequences.

What About Other Tasks?

Lowenthal mentioned that he envisages to use other exercises with less structure in it but where the three Brunerian levels of representation are involved and see what type of possible progresses occur. **Fitch** suggested to use several types of tasks as it is done in the animal literature. There, several types of sequential learning tasks are used. Some are very easy for animals, such as pure sequence learning, learning a sequence of numbers, and learning a sequence of objects. Monkeys can learn to do this very quickly. Phylogenetically, it is the simplest thing humans can do. A second group of tasks could be based on sequences of movements, such as a kind of navigation task: the subject has some overall idea of where he is and must simply keep in mind some kind of vector pointing back. This is also just a sequence of learning. A labyrinth task could be a paradigmatic example. The third group of tasks could be based on hierarchical sequence learning where the subject has to perform some complex motor action that has at least two parts, one of those parts having to be fully completed before the second part can start. This could be called hierarchy learning. Finally the most challenging group of tasks would be truly recursive tasks where each iteration through the task requires the subject to remember past structures. He thus suggested to train the subjects through these four groups of tasks and then start to examine what is the most influential one.

NVCDs and Recursion

Fong and **Dienes** noted that the sequences of exits produced by DM networks can be produced by a finite-state machine: the loops of these networks do thus not really correspond to recursion as it is meant by linguists. **Fong** said that when some people talk about recursion, they just mean simple loops. Finite-state machines have simple loops and can be very complicated. The networks built with DM elements can be extremely complex. Nevertheless they do not have the flavor of things involving a hierarchy or bracketing. It would be nice to have different kinds of puzzles that would involve bracketing: puzzles that only work out if the elements are properly bracketed. It would then be relevant to test those games against the DM puzzles. **Lowenthal** replied that one of DM pieces, not presented during the conference, is a real counter which can be used to construct more complex networks. This enables the experimenter to reach the level of real recursion in the following sense: one can create a network producing a pattern (sequence of exits) which can be produced by something which is calling itself at a lower level of complexity until it goes down to zero. **Lowenthal** then explained that Cohors-Fresenborg, the creator of the DM puzzles, has devised a computer language which is based on bracketing. The class of problems that can be introduced via this language, which is clearly recursive, has been proven to be equivalent to the class of possible DM networks, taking into account the counter. He also said that while Cohors-Fresenborg's purpose was to ask a subject to create a network producing the solution for a given verbal problem, his personal approach was different: he asked the subject to examine a given network and describe the class of problem for which it could produce a solution.

Dienes suggested to create exercises where the main network would be a "meta-network": each of its exits would be another network called by the "meta-network." This would enable the "meta-network" to call itself. Considering the Fortemps, Lowenthal, and Wautié (2014) paper, he notes that this approach is now technically feasible. It will thus be possible from now on to test subjects' reactions to different types of networks.

Suggestions for Future Researches

Gervain suggested to consider the reverse approach: one could train people using linguistic structures and observe if they perform better with NVCD-type exercises. **Lowenthal** replied that it could be done with healthy subjects but that it seems difficult with aphasic patients and with mentally handicapped subjects. Considering the researches with young healthy subjects, he noted that using the language to train and observe the progresses of their thinking would be subject to the same biases as those mentioned by Piaget's critics. Furthermore, this type of training misses one important point: the fact that NVCD-like exercises can be solved at each of the Brunerian levels of representation. **Freidin** asked Gervain what type of language training she was thinking of. **Gervain** replied that asking children to do something they cannot do would be nonsense, but she believes that some verbal training can occur at very specific ages. The point she is trying to make is a theoretical one, because if the NVCD-type remediation works only in one direction, it might involve a brain structure which is different from those concerned if it goes both ways.

Dienes mentioned the mathematics educator Zoltan Dienes and his principle of variability of embodiments: in order to get an understanding of a structure, it is necessary to embody it in as many different ways as possible. He thus suggested to try to embody the same recursive structure in different NVCDs in order to observe possible transfers to language, and better improvements. He then suggested to use the new software described in Fortemps et al. (2014) to create three types of DM exercises producing the same pattern of exits but different in structure: without loops, with loops but no real recursion, and, finally, with full recursion. It would be interesting to compare the results of subjects confronted to these different approaches. He added that, as far as therapy is concerned, if recursion is an important part of the training, this training should be organized in such a way that the person receiving therapy would be able to understand the difference between a recursive and a non-recursive situation. It might be relevant to train the subject on both situations and make him observe the differences. By doing so, it might be possible that these subjects become aware of the structural differences between the two situations. In any case, according to **Dienes**, if recursion has something to do with language, one would also expect to observe more embedded clauses in the subject's verbal productions. Should that not be the case, then whatever training is done, it is in fact different from the recursion observed in language.

Are NVCD-Type Approaches a Form of Implicit Learning?

Gervain wondered how much the NVCD-type approaches differ from usual implicit sequence learning as it is used in other domains of cognitive psychology. Is it the fact that the participant has to construct the network? Or is it not so different from other types of sequence learning? According to **Lowenthal** the difference is partly due to the fact that each exercise can be solved at each Brunerian level of representation: enactive, iconic, and symbolic. There are several observations showing that different subjects are working at different levels of representation. He also thinks that the fact that subjects are constantly asked to formulate hypotheses, to test them and possibly to validate them by actually going through the network, and thus going one or two levels down in representation, is more explicit than implicit.

What Is Really Acquired Through NVCD-Type Approaches?

Friederici wondered how a training on structures can help subjects to learn to use the proper terminology, i.e., proper words instead of periphrases. **Lowenthal** reminded her that these results were mainly observed through Pegboard training. He wondered if the fact that subjects were trained to give names to sequences, in order to create procedures, did not favor their naming ability.

Gervain asked if other cognitive functions were checked, such as executive functions and general intelligence. Improvements in those areas as well would require a specific explanation. **Lefebvre** replied that two groups of results have been observed: some concerning language and others concerning the visual-spatial ability. He claimed that the results obtained do not show any possible link between these two groups of results. **Lowenthal** added that the results obtained with psychiatric patients on attention, anticipation, and inhibition do not seem to be related to language. He also recalled the discovery of the "If… Then…" by a mentally retarded subject, without any relation to a language improvement.

Words or Concepts

Gervain wondered, as far as the healthy young subjects are concerned, if Lowenthal did not observe a conceptual improvement instead of a linguistic one. According to her, a Piagetian could argue that if the subjects improve on "if… then…," it is not that they do not know how to use it linguistically, but that they just do not understand what "if… then…" is. This implies that these subjects only improve on conceptual knowledge: they always had the linguistic means to express what they now do, but they could not think of the words to say it because they lacked the concept. This remark boils down to say that a set of NVCD-type manipulations favors a semantic or conceptual improvement but not a syntactic one. **Gervain** mentioned that in a certain sense the fact that one does not say something can mean two things: either one does not know how to say it linguistically or one does not know how to think that. According to **Lowenthal** one first develops a concept and then one gets the words to express it. The words acquired before the concept are irrelevant since one cannot use them meaningfully. The converse idea has been a critic of Piaget's experiments, because he was using verbal communication, thus language, to test the hypothesis that language is not necessary for the development of thought. Linda Siegel created nonverbal tasks, equivalent to verbally presented Piagetian tasks, showing that the concept could be introduced in a nonverbal way. In the context of NVCD-type approaches, existing concepts are used in order to provoke the use of words. The concept is necessary to communicate, otherwise one communicates in a vacuum, but the words, as seen with aphasic patients, are not: they appear later, after the use of NVCDs.

According to **Fitch** there are many circumstances where a child learns a word, and can say that word, without knowing what it means. There are situations where a child can correctly answer a question, such as "What is the capital of France?" without actually knowing what "capital" means. Since this can always occur in a language-acquiring child, researchers must always wonder how they can know if this child actually understands what he or she said. It is not easy to know whether the child has the concept first or the language first. One should find a way to distinguish between these two possibilities.

Lowenthal claimed that if a subject has the word but does not know what it means, or what it refers to, or in which context it is used, then this subject has nothing. In his experiments, he has behavioral observations showing that the associated communication is meaningful for the subject. But that is not enough evidence for **Fitch**: one could have some phonology, some associations, some syntactic information, but nothing else. **Freidin** retorted that Fitch's remark does not work for the "if… then…": if a child uses the "if… then…" structure, he or she must know what it means. But **Vergauwen** remarked that there is a real possibility that the situation could be similar to Weizenbaum's ELIZA program. In this program a pseudo psychologist, ELIZA, answers to questions: the answers are apparently adequate but in fact they are randomly selected from a stock of possible answers. It is clear that the program does not understand anything. In the case of human beings the situation is different. There is a basic assumption: human beings do understand the questions and their utterances have a purpose. These utterances are thus meaningful. So, in fact, human beings need the concept, as Lowenthal said, despite the fact that it is possible to conceive an automatic sequencing machine which answers questions apparently adequately. **Fitch** then reacted by saying that he believes that when adults use words, they know the concept, with the possible exception of recursion. According to him, one uses words because one knows what they mean. He was only suggesting that during the child's acquisition of language, one cannot be so sure: a child might initially use a word without a full understanding of its meaning. This child is probably grasping toward some notion of meaning, but its notion of meaning is not the same as the adult notion of meaning.

Gervain came back to the "if… then…" structure. According to her, there is developmental evidence that shows that children understand "if… then…" only when the "if…" part is obviously true, such as "if it rains, then the ground is wet." But the same children cannot deal with counterfactual situations and do not understand what it means to say "if it rains then the ground is wet and it is not raining." So at a certain level of complexity, it becomes very difficult to check if the subjects really understand what they say. According to her, this shows that children might have difficulty understanding the concepts. **Freidin** disagreed: he stressed the fact that this discussion concerns children's productions, and not the concepts in general. It thus seems logical to accept that when a child produces an "if… then…," this child must have some idea of what it means, otherwise they would not be using it. **Lowenthal** mentioned Dumont's 1980 paper[3]: this French researcher confronted many young soldiers, during their military service, to different sentences containing an "if… then…" structure. The results show the importance of the formulation: many young adults do not understand this structure, and those who do not completely master it, do not use it spontaneously.

Neural Networks and Implicit Learning

This section includes remarks and questions made by the participants during several discussion sessions. These remarks can be subdivided into three main groups: what are SRNs capable of? Is a SRN a simulation of the human brain or does it give information about our brain? Can we use neurocognitive information concerning the human brain for SRNs? The discussion was vivid as the opposition between pro- and anti-connectionism became more and more palpable.

---

[3] Dumont, B., (1982) The influence of the formulation of multiple choice questionnaires on the answering behaviour in relation to so-called "logic problems." In Lowenthal, F., Vandamme, F. and Cordier, J. (Eds), *Language and language acquisition* : New-York, Plenum Press. 225–262.

What Can We Do with an SRN?

SRN and Innateness: Is Human Cognition Acquired or Not?

Cleeremans explained how the Simple Recurrent Network proposed by Elman can master long-distance, recursive contingencies by means of simple associative learning mechanisms that continuously attempt to predict the next element of a sequence. He claimed that such networks represent a viable alternative to the traditional view insofar as they are capable of exhibiting performances similar to that of human learners on a variety of linguistic and nonlinguistic tasks that involve long-distance contingencies. **Fitch** reacted by saying that it is now obvious that an SRN can duplicate what our brain is producing, but that it does not mean that it is a good model of how our brain function. According to him, connectionism cannot solve the deep questions we are interested in. The SRN can be used for simulations, but cannot provide us with a reply for the fundamental questions about how our ideas change. He noted that Cleeremans objected to Marcus' statements about innateness (innate format for representations, innate set of operations, innate ability to do variables, etc) on the basis of results obtained with SRNs. According to **Fitch**, in a SRN, we have an innate structure, with innate operations, both given by the programmer. The SRN can thus only perform certain kinds of operations and has a lot of built-in structure. The question "Can this object learn a finite-state grammar?" is a worthwhile question to ask, but this question has little to do with the much bigger questions about innateness vs. learning. **Fitch** asked what kind of constraints there are on the brain that humans have and chimps or birds lack. According to him, this kind of questions is way beyond what can be addressed with connectionist models. He thus asked: "What can connectionism do for us?"

Cleeremans replied that he has the intuition that, while some say that a great amount of our cognition is innately given, there could be alternatives. He also admitted that connectionist networks, like any other approach, are not free from initial parameters. It is thus important to figure out what is depending on the maturation processes, on genetic constraints, etc. There is a balance between how much we think such simple systems are able to learn by themselves and what are the minimum requirements needed to make them able to become developing mature structures. He then said that over the past 10 years, connectionism has been replaced by Bayesian approaches among others. However, one must note that McClelland is now emphasizing a distinction between emergent knowledge structures and descriptive models. The models presented now are concerned with abstraction and consciousness. **Cleeremans** mentioned Dienes' work showing that it is possible to create an SRN able to generalize knowledge acquired about sequential regularities in one particular domain to a completely different domain. According to **Cleeremans**, this is one way of dealing further with abstract knowledge. Nevertheless these networks are not able to perform true abstraction in the sense that the knowledge acquired remains in the network instead of becoming a set of representational objects.

Cleeremans also insisted about the importance of implicit learning as a defining feature of human cognition. He said that it is not clear to which extent this can be done by connectionist networks. In fact, the problem is actually to determine how one can go from first-order networks which get 90 % of the knowledge needed to be abstract, but fail to fully implement this abstraction, to second-order networks which do the rest of the job. It seems that language plays a fundamental role in the human ability to do that.

Dienes explained that he considers connectionism as a tool. It is an interesting tool because when people like Marcus say "I cannot do that," connectionism enables us to show that it is possible! It seems unlikely that SRNs can solve all the problems, but it seems worth pushing this concept as far as possible. Indeed, in an SRN, there is a stack which could possibly do some recursion. One should try to push these simple devices to see what they can do and compare that with what the circuits in the brain are doing when we see something structurally similar.

Can an SRN Master a Pattern Without Knowing Its Elements?

Fitch described an experiment where he trained pigeons on bigrams. The animals mastered the pattern, thus the abstract rule governing it. But they were not able to identify the bigrams. He thinks that an SRN could not do that: it could not learn the big structure before it had learned the bigrams. He thus wondered if the fact that animals are getting this larger scale pattern but are not mastering the bigrams would falsify an SRN as a general model for sequence learning. Or is another model needed for pigeons? He considers that the pigeons do not use the transition probabilities. **Dienes** reacted by saying that this situation is different since humans use the transition probabilities. **Fitch** replied that this was his whole point: on this matter humans are different from birds, but there is a "primitive animal" which is able to learn complex patterns based on bigrams but without knowing the bigrams. The question with humans is thus: do they just have the ability to learn the bigrams or do they have that ability plus another ability which enables them to learn an abstract structure? If one is confronted to an animal which does not use the bigrams at all, what will a real connectionist say about the SRN? According to **Dienes**, this is not a problem: we can change the assumptions we used or modify the format of the input in a connectionist model. This does not falsify the SRN in se and another SRN will probably provide us with a better account of the observed. Connectionism for him is just a language for formulating the model and should be abandoned.

Is our Brain an SRN?

Corballis asked, since our brain is an associative network, why should there be, in principle, a difference between what a brain can do and a neural network can do? Does the human brain have something extra that the researchers do not understand?

Vergauwen retorted that the main difference is the embodiment: the brain has also physical information which makes the difference. **Corballis** thus wondered if the difference was due to our body: our brain is not living as a pure abstract problem solving machine. It seems to him that neural networks are important as tools, but apparently it is not possible to go all the way with them. **Dienes** specified that "going all the way" is precisely what he wants to do, but it is true that one needs the body. It is obvious that telling the body what to do and getting information back from it are essential activities for our brain. There is thus something neural networks are doing that explains what the brain is doing, but one does not know yet how to model that "something." It is therefore important to explore presently simple connectionists networks; other researchers will complicate them and make them more and more neuron like. This is not acceptable for **Fitch**: he claimed that it is a mistake to say that the brain is essentially a network: the human brain is an incredibly complex organ with many types of fibers running one way and not the other. The brain, according to **Fitch**, is not just a bunch of nodes where everything is connected to everything else, because each of these neurons is a living organism that develops its own relationships. **Dienes** reacted and said that in a neural network it is not the case that everything is connected to everything else. **Fitch** replied that the difference between a standard connectionist network where you have rich interconnections and any adult brain circuit is tremendous. It is thus not possible to say that the brain is only an associative network. **Cleeremans** disagreed: according to him, there is only one learning mechanism in the brain which is copied in neural networks. The brain must thus be like an associative learning machine. The human brain is also very plastic: learning must thus be an important part of the brain activity. **Cleereman**s did not observe any extra ingredient in the human brain beyond the fact that it has incredibly sophisticated networks of incredibly sophisticated organisms and that the number of neurons is huge.

SRNs and Cognitive Neuroscience

Cleeremans explained that there is now a tendency to merge to some degree the early connectionist ideas with Cognitive Neuroscience: researchers try now to build networks informed of the different functions of the brain circuits. **Vergauwen** noted that in order to do so the researchers must start from the assumption that simulation equals duplication, but this is not necessarily the case. According to **Cleeremans** this is Soar's argument. He added that one thing only is missing: the embodiment, i.e., the body.

Friederici recalled that there are differences between different parts of the brain depending on the place and nature of the cells. She thus asked if it is Cleeremans' and others idea to learn more about the biological system and then make better neural networks. **Cleeremans** agreed saying that present neural networks are an oversimplification. They constitute nevertheless a good tool to try to understand what type of computation occurs in the brain. Since there are two systems of representation for memory in the brain, one must create two different systems of computations in a neural network. **Dienes** claimed that researchers concerned by neural networks do not ignore what is done in Cognitive Neuroscience, but they need two strategies: on the one hand they must use simplified neural networks and see what happens locally and on the other hand they must put simple elements together in order to create more complex versions and observe advanced processes.

SRNs and Ambiguity

Freidin, working on syntax, wondered how neural networks can handle structural ambiguity. He mentioned the expression "a review of a book by two professors" and remarked that it can be interpreted in two ways: a review made by two persons or a review of a book which is written by two persons. He also noted that this is a case where one has a noun phrase inside a noun phrase, thus a recursive structure. **Vergauwen** contested the last point: he does not believe that Freidin's phrase involves recursion but considers that we have there a semantic property: "aboutness." **Dienes** told Freidin that connectionism does not deal with ambiguity nor with syntactic structures, at least not yet.

SRN, Language Acquisition, and Consciousness

Dubois claimed that if a computer succeeded to learn a human language, it would automatically have a conscience. **Vergauwen** wondered how one could go from a simulation of a natural language to the automatic duplication of consciousness. **Dubois** replied that, according to him, conscious intelligent systems can understand a natural language and process it. Thus, if a machine is able to process a natural language, this machine must be a conscious system.

Visual Recursion

In order to test the dependence of human language on recursive abilities, Martins used a new method to assess the ability to represent recursive structures in a nonlinguistic (visual) domain. His visual recursion task (VRT) is based on the observations of the first steps of visual fractal generation. Martins then asks his subjects to make inferences about subsequent iterations.

Gervain observed that Martins' VRT does not enable him to represent certain types of rules such as linguistic rules, counting rules, or putting the words in the backward order. These actions can be performed by people. So the VRT does not seem to measure completely the recursion ability in human beings. **Martins**

explained that he decided to build his VRT in order to be able to represent stimuli as a set of recursive generative rules in a way that is independent of language. This task can thus be applied to animals, to children, and to people with no language. VRT enables the researcher to study if people can perform the required actions independent of language. Results indicate that there is a relation with the verbal working memory: aphasics are worse than healthy subjects when performing this task. This raises another question: if one forces the subjects to answer in different conditions while masking all verbal aspects, can these subjects actually do this task in such a way that they do not recruit verbal computations? In other words, is it possible to perform recursive computations in a way that is independent of language? **Fitch** asserted that after the Hauser, Chomsky, and Fitch paper (2002), there were several debates and claims concerning the exact role of recursion in human cognitive activities: e.g., Jackendoff thought he had proven that there is recursion in music because he could do a recursive analysis of various pieces of music. The real problem is not whether Jackendoff can do a recursive analysis, or whether Bach had recursive processes in his mind when he wrote his pieces of music, but whether subjects actually perceive that structure. Jackendoff and Pinker said that it is obvious that there is recursion in vision. Martins' aim is to prove, or disprove, this statement.

Gervain wondered if Martins' task is not something people can do but would not naturally do. **Martins** replied that fractals are everywhere. In fact the VRT consists, in predicting what is going to happen next, independently of the level of embedding. The limitation on the level of embeddings mentioned by Levinson is thus not relevant. The next question will be to examine how this ability relates to other known cognitive abilities.

Lefebvre noted that, according to Martins, "fractals are everywhere" and that according to Dubois fractals are recursion. Should one conclude that recursion is a general ability, i.e., is recursion the same in all models, or do we have a specific recursion for each cognitive domain? **Martins** answered this question in two phases. He first said that one finds fractals everywhere. There is a recursive structure in them. There is nevertheless a difference between this fact and the fact that a subject is able to perceive this structure. There are thus two different levels: the existence of a structure and the recognition of this structure. As far as the uniqueness (or not) of the type of recursive structure human beings use, **Martins** mentioned a new research partly devoted to build other tasks similar to the VRT, but in other domains, such as the auditory domain. The purpose of the research is to compare the circuits that are recruited. **Dubois** reacted and said that there are fractal patterns which can be generated by several different recursive algorithms. There is not only one way: the same pattern could be generated by a deterministic or a random process.

Frath claimed, on the basis of an experiment he made, that as human beings, we might have recursion in our mind, but it could also be that we have an impression of regularity and that we choose what is more in conformity with that pattern. **Martins** commented that in vision one might be able to represent recursive processes at a pure intuitive level, and thus navigate through recursive structures, in order to make predictions. **Frath** objected and said that this might have nothing to do with recursion. **Martins** defended his VRT saying that if somebody performs it intuitively, then this person might be able to use recursion in other cognitive domains. He also said that nonhuman animals might also be able to use his VRT and thus perform some kind of recursive precursor.

Corballis wondered if Martins asked his subjects what they were doing. Did you note whether they said “Hey it is a fractal !” or not? **Martins** explained that, in a pilot study, he asked subjects which strategy they used more frequently. Subjects who said they did it more intuitively also said “Oh I just tried to look at the general characteristic in the images and I tried to find something similar,” while subjects who reported they used more analytical strategies also said “Oh yes because the small ones are build on the same manner as the bigger ones.” Both strategies are possible and performances are not predicted by the strategy used.

Concluding Remarks

During these discussion sessions, the participants had the opportunity to gather more facts and new information about their common field of interest: language. Specialists from different domains and from different schools of thoughts participated actively to the debates. The discussions helped them to clarify their divergences and even sometimes to agree on some points or to understand why they disagreed. Pro- and anti-recursion in language, believers in innate factors and in acquired factors, etc., had an opportunity to compare their views and reach some kind of synthesis. New research approaches were suggested and new experiments were prepared. The first results of these experiments constitute the chapters of this book—to be continued!

References

Fortemps, P., Lowenthal, F., & Wautié, V. (2014). Computer simulations of recursive exercises for a non-verbal communication device. In F. Lowenthal & L. Lefebvre (Eds.), *Language and recursion*. New York: Springer.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve ? *Science, 298*, 1569–1579.

Martin-Ordas, G., Berntsen, D., & Call, C. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology, 23*, 1–4.

Zuberbühler, K., & Lemasson, A. (2014). Primate communication: Meaning from strings of calls. In F. Lowenthal & L. Lefebvre (Eds.), *Language and recursion*. New York: Springer.

Index

F. Lowenthal and L. Lefebvre (eds.), *Language and Recursion*,
DOI 10.1007/978-1-4614-9414-0, © Springer Science+Business Media New York 2014

S

T

U

V

W

Z