Prolegomena Paedagogica

INTRAMENTAL EVOLUTION AND ONTOGENY OF TODDLERESE

&&

FOUR SIMULATIONS

1st and 2nd volume of PhD. dissertation by

MSc. Bc. et Bc. Daniel Devatman Hromada

WHAT THIS TEXT IS NOT

This text is not a formal work in a mathematico-logical sense. Its aim is not to introduce *the Theory* of natural language, nor even a set of theorems, whose validity could be proven by blind application of rules of symbolic substition upon a pre-defined set of solid definitions and «self-evident » axioms. Assuming that

- indeed *incomplete is every formal system* whose explanatory power is at least equivalent to explanatory power of formal system of basic arithmetics (Gödel, 1931)
- explanatory power of any natural language system is at least as exhaustive as that of any conceivable arithmetic system

we consider the temptation to explain natural languages in strictly formal terms, as potentially counterproductive one.

Nor is this text a product of analytical approach to science. It shall not limit itself to study of a sub-problem of a problem which can be sometimes observable when one confronts the world with a particular terminology and methodology of a sub-branch of a highly specialized discipline. It does not, in Nietzschean terms, devote itself to the study of « the brain of the leech ».

In other terms, this text does not aim to attain knowledge – whatever it is - by reductionist act of focusing one's attention upon one sole boring fragment of "truth".

Knowing that truth is complex, sometimes contextually bound and more than often simply beyond reach of an individual observer, we do NOT pretend that ALL hypotheses presented on following pages are apodictically and universally true. For any hypothesis is just a piece of a bigger picture and it is this picture itself which is supposed to represent Reality – i.e. to be « true » - and not the pieces. On their own, theses and hypotheses are just indices helping the scientist to find his way on a path to such bigger picture.

Thus, even invalid hypothesis can serve the productive purpose if ever they succeed to transpose the scientist into realms where (s)he was never before.

And as we shall try to indicate and re-indicate during this whole text, it is indeed «by descending into & traversing through the valley of falsehoods» that the researcher can ultimately attain a perspective which is «higher» (i.e. more «optimal») than the original one.

This principle, we believe, applies both on a baby language learner as well as on evolution of scientist's knowledge and, possibly, on evolution of science in general.

WHAT THIS TEXT IS

This text is a tentative to elucidate «the mystery» of acquisition and development of linguistic competence in terms of evolutionary and complexity theory. Thus, it is principially a multidisciplinary scientific essay.

By being «scientific», its aim has to be either analytic or synthetic ; and since we have already that our main aim is not analytic, it follows that the goal of this essay is of synthetic nature. More concretely, the synthesis under question aims to involve following scientific disciplines : artificial intelligence and artificial life, cognitive psychology, developmental psycholinguistics, evolutionary computing, natural language processing, theory of complexity, and universal darwinism. Mindmap localizing the central Topic of this text within wider scientific context is presented on Figure 1 (c.f. list of Acronyms on page 355 if some abbreviations are unclear or ambigous).

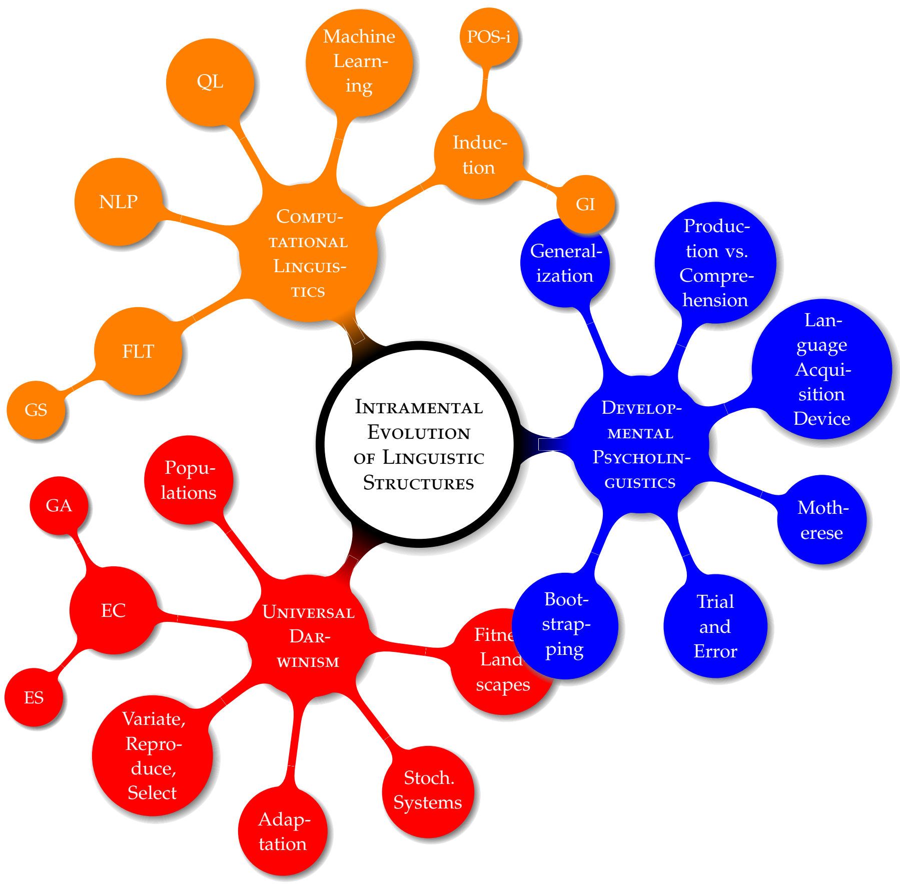


Figure 1: Central notions of this dissertation.

To demonstrate the validity of our perspective, this text shall three different proofs-of-thesis. The theoretical proof-of-thesis shall shall consist in making reference to & aligning with multiple theories scattered among different disciplines of cognitive sciences. Ideally, many seemingly unrelated phenomena could be thus brought under clef-devoute of one scientific paradigm. The observational / empiric proof-of-thesis shall aim to align the thesis with seemingly trivial observations of linguistic behaviour of a certain human subject. Finally, the computational / experimental proof-of-thesis shall hopefully illustrate that diverse problems of language acquisition are computationally solvable if ever they introduce an evolutionary component.

At last but not least, this text is also a dissertation work with which we aspire for the attribution of the title Philosophiae Doctor. For this reason, all chapters of this first volume contain a certain quantity of remarks which partially surpass the informatic, cognitive and/or psycholinguistic paradigm and point in direction of philosophy in general, and epistemology in particular.

HOW IS THE TEXT ORGANIZED

The text is composed of two volumes which, taken together, contain four parts. First volume consists of three parts, second volume (Hromada, 2016d) consists of only one. Each part is divided into chapters. Every chapter consists of introduction and conclusion preceding resp. following more specific subchapters which can fractally branch into sub-chapters , sub-sub-chapters etc. All such parts, chapters, subchapters etc. can be considered to be « non-terminal » nodes of structure presented by this text.

Basic structure of the text

The first part, labeled Theses, is a stem of whole text. It will introduce multiple theses at varying degrees of generality which shall be all - in one way or another - more directly addressed in subsequent sections. In order to weave the basic conceptual fabric, some definitions of terms like « evolution » and « language learning » shall be also offered. All variants of the thesis shall be briefly related to other cognitive sciences.

The second branch, labeled «Paradigms» is composed of chapters dedicated to Universal Darwinism, Developmental Psycholinguistics and Computational Linguistics. In these chapters, the theses presented in the first chapter shall be more deeply interpreted and contextualized in terms of respective disciplines.

The third branch, labeled «Observations» will describe multiple longtitudinal observations of one concrete human child. Subsequent interpretations in terms of the evolutionary theoretical framework shall follow.

The ultimate branch, called «Simulations» shall present multiple computational models addressing three problems related to language acquisition process. That is,

1. the problem of concept induction
2. the problem of induction of grammatical categories
3. the problem of induction of grammatical rules

Specific chapter will be dedicated to every problem in which existing solutions shall be described. Special focus shall be put on evolutionary solutions, if they exist. To every of four above-mentioned problems we shall try to offer our own unique evolutionary solution and subsequently we shall discuss its performance. PERL source codes shall be also attached and publish under mrGPL licence in order to facilitate reproducibility (Hromada, 2016e) of results by other scientists.

Text's nodes and their attributes

As a whole, the text hereby presented can be thus considered to be a tree with five major branches which bifurcates all the way to « terminal » (i.e. leaf) nodes. To all nodes of such « tree » shall be also attributed one among following types:

DEF	Definition	Intensive or extensive definition (or combination of both) of the term used throughout the book
TXT	Text	Longer piece of text, often dedicated to one specific hypothesis, topic, theory or model - this is a default node type
OBS	Observation	Transcription of an item from the observation journal
APH	Aphorism	A comment presenting author's stance in regards to topic raised in Text or Observation node. More subjective than TXT
SRC	Source code	Snippet of PERL source code

The type of the node is specified in its title. Preceding the title is a unique numeric identificator which can serve as an anchor for cross-references. Thus, a text dedicated to Piager's Genetic Epistemology which is contained in fifth secion of chapter eight, will be introduced with a following expression:

8.4.4. Genetic Epistemology (TXT)

The end of every node is marker by an expression containing node's numeric ID, title and the token END. An above-described node will thus be terminated with a following expression :

8.4.4. Genetic Epistemology END

Because the nodes can be embedded within each other, such syntax is needed to exclude any disambiguity. C.f. 1.0 and its relation to embedded nodes 1.0.1 and 1.0.2 for a concrete example of such embedding[1].

This is a self-referential margin-note

Margin-notes shall be also employed to facilitate even further the orientation within the text and cross-referencing between diverse parts of the text. Such a note shall be usually placed at the margin of the text whenever a new topic shall be addressed.

IN WHAT LANGUAGE IS THE TEXT WRITTEN?

This dissertation is written in a language which shares majority of its morphological, lexical and syntactic features with modern standard english[2]. Thus, majority of words are english words and majority of sentences can be easily parsed by a standard english-language parser.

But it has to be noted that this text is not written by a native english speaker. Written mainly in germany and deposed at french university by a child of slovak mother and czech father, inspired by compactness and eloquence of classic (i.e. latin, greek and sanskrit) treatises, and often aiming to denote very subtle distinctions and novel meanings: all this often lead to a *sapirwhorfian feeling* that communication of certain thoughts is inconsistent with certain well-established schemas and rules. If ever such situation occured, it was the communicative intention and not the rule which was prioritized: hence the origin of many seemingly agrammatical constructions present in this work.

Thus, asides multitudes unvoluntary and erroneous typos and asides multitudes of omitted and/or misplaced articles - a slavic speciality - this work also exposes the reader to a certain amount of errors which are, in fact, not "bugs" but "features". In certain cases, *italics* and **bold** were used to mark the moment whereby the author intentionally broke the existing schema - or invented a new one - in order to emphasize a certain aspect of the-intention-to-be-communicated.

1 Without this embedding, the arborescent structure of this Thesis would be reminiscent of Wittgenstein's Tractatus. But because this embedding is implemented, the structure resembles more a context-free (10.2.2) form of a valid XML document.

2 Within the context of this dissertation, standard english is principially understod in terms of set theory as *union* of british and american english. Given that it is defined as *union* and not *intersection* both (i.e. american as well as british) can be accepted as valid and used interchangeably in cases where two languages diverge (e.g. both british "optimi**s**e" as well as american "optimi**z**e" can be accepted).

Part I

THESES

In the distant future I see open fields for far more important researches. Psychology will be based on a new foundation that of the necessary acquirement of each mental power and capacity by gradation.

— Charles Darwin

In this part we shall posit and discuss multiple theses whose validity or invalidity we shall try to demonstrate in subsequent parts of this diseration.

After a brief discussion of Initial Thesis "mind evolves", the sense of the Hard Thesis "learning is a form of evolution" shall be more thoroughly criticized by exploring the conditions of its validity. The Soft Thesis "learning can be successfully simulated by means of evolutionary computation", the Softer Thesis "learning of natural language can be successfully simulated by means of evolutionary computation" and the Softest Thesis "learning of first language can be successfully simulated by means of evolutionary computation" shall be postulated next. At last, the Operational Thesis "learning of first language from its textual representations can be successfully simulated by means of evolutionary computation" shall turn out to be sufficiently concrete enough to become an object of computational simulations.

Definitions for terms mind, to evolve, evolution, brain, 2nd law of thermodynamics, evolutionary computation, natural language, first language and child shall be also provided. Asides all that, a so-called "alternative" hypothesis concerning the non-local storage of information in human brain shall be also introduced.

1 INITIAL THESIS

Mind evolves.

This is the Initial Thesis (IT) whose validity we hereby undertake to demonstrate. In order to do so, both terms of the statement are to be properly defined.

1.1 MIND (DEF)

Definition of substantive "mind"

An auto-organising set of structures and processes determining the characteristic behaviour of an individual.

END MIND 1.1

1.2 TO EVOLVE (DEF)

Oxford Dictionary definition:

1. Develop gradually
2. Develop in time as a result of natural selection
3. (chemistry) Give off gas or heat

Definition of verb "to evolve"

Etymological definition:

- 1640s: "to unfold, open out, expand," from Latin evolvere "to unroll," especially of books ; ... from ex- "out" + volvere "to roll".
- 1832: "to develop by natural processes to a higher state"

END TO EVOLVE 1.2

IT means that an auto-organising set of structures and processes determining the characteristic behaviour of an individual is endowed with propensity to gradually attain higher states of complexity. Hence, not only *structures stocked in and by mind*, but also the very processes which act in mind are to be understood as subjects to transformation.

A fact that the predicate "to evolve" is conjugated in indicative mode of 3rd person singular of present simple tense suggests that the statement tends to denote the-state-of-affairs independent of temporal context within which the *evolution of mind* occurs.

__IT__ seems to be a tautology

Thus, it can be reproached that IT is too general and potentially tautologic. Since it is difficult to see how such a statement could, *per se*, become an object of positivist endeavour, let's now discuss IT's less tautological variants.

END INITIAL THESIS 1

2 HARD THESIS

Hard Thesis (HT) is expressed as follows :

«learning is a form of evolution» .

The term evolution, as presented in HT, is to be understood in terms of generalized form of Darwin's theory, which is called Universal Darwinism (UD). In such framework, evolution can be defined as follows :

2.1 EVOLUTION (DEF)

Definition of substantive "evolution"

Evolution is a durative process emergent in any finite-resourced environment containing a population of information-encoding entities which :

1. Reproduce
2. Need resources for their reproduction
3. Vary because of inaccuracies inherent to reproduction process

END EVOLUTION 2.1

How evolution "works"

If ever there exists a causal relation between the information these entities encode (genotype) and the means how they exploit environment's resources (phenotype), the population will lead to gradual optimalization of its relations with the environment, i.e. discover ways which exploit resources more efficiently than before. In this sense shall be the next generations of individual-encoding entities better «adapted» to their common environment.

Evolution has many forms

It is important to realize that the notion of «evolution», as hereby defined, goes far beyond the traditional Darwinian theory which was concerned with just one instance of evolution, namely the biological one. Some phaenomena, which could be interpreted or even modelled as instances of systems whose functioning is constitent with the precepts postulated by Universal Darwinism, shall be in somewhat closer detail discussed in Chapter 8.

Logical necessity of evolution

For a Universal Darwinist, evolution is not an empirical but a logical necessity. It has to necessarily occur within any system fulfilling the above-mentioned conditions. Emergence of evolution in a system fulfilling above-mentioned conditions is independent from the concrete form of «natural laws » & physical constants which determine the particularities of such system.

The Hard Thesis states that psycho-pedagogical process of «learning» can be not only interpreted and simulated as an evolutionary process. The Hard Thesis states that learning IS functionally equivalent to evolutionary process. That on an ontological level, « learning » is an instance of an evolutionary process and therefore IS an evolutionary process. In UD-consistent sense.

HT is about ontological equivalence

2.2 LEARNING (DEF)

Definition of substantive / participle "learning"

Learning is a mind-transforming, information-processing, constructivist and embodied process.

END LEARNING 2.2

The attribute « mind-transforming » denotes the finality of learning - it means that both contents as well as processes which determine the characteristic behaviour of an individual agent can be modified by means of learning. Attribute « information-processing » denotes the modality of learning – it implies that learning always involves the processing of information – namely assimilation, accomodation, encoding, storage or decoding of information. The term « constructivist » suggests that learning is gradual and can potentially bootstrap itself. The term "embodied" suggests that learning could succeed only with big difficulties if it is not embedded into an individual monadic entity which keeps track -in one way or another- of its own trajectory.

Attributes of learning

The last term of HT is defined, in accordance with tradition, as follows :

2.3 FORM (DEF)

> « Form is the possibility of structure.» (Wittgenstein, 1922)

END FORM 2.3

Given the definitions 2.1, 2.2 and 2.3, the Hard Thesis - presented as a conjunction of terms « learning is a form of evolution » - can considered to be true *iff* following statements are true as well:

2.4 FIRST CONDITION OF HT'S VALIDITY (DEF)

Learning involves the reproduction of information-encoding entities.

END FIRST CONDITION OF HT'S VALIDITY 2.4

2.5 SECOND CONDITION OF HT'S VALIDITY (DEF)

These learning-enabling information-encoding entities consume resources in order to reproduce.

END SECOND CONDITION OF HT'S VALIDITY 2.5

2.6 THIRD CONDITION OF HT'S VALIDITY (DEF)

The process of reproduction of information-encoding entities can be influenced by stochastic phenomena which cause an unpredictable structural variation.

END THIRD CONDITION OF HT'S VALIDITY 2.6

2.7 FOURTH CONDITION OF HT'S VALIDITY (DEF)

The resources of environment, within which the learning occurs, are finite.

END FOURTH CONDITION OF HT'S VALIDITY 2.7

Hard Thesis, as proposed until now, defines learning in general and as such can be told to describe the form of « learning » of both human and artificial minds. In this general sense, it will be used in majority of the text which shall follow. For the rest of chapter 2, however, we shall discuss « learning » as related solely to humans.

The material substrate of human learning[1] is the brain and no positivist theory of learning thus cannot be considered to be adequate if it ignore brain's essential attributes. We list its essential attributes in a following definition.

2.8 BRAIN (DEF)

Human brain is a physical (i.e. four-dimensional) object of organic origin which consumes biochemical energy in order to process and/or store information in a non-local, highly parallel, and in certain extent also plastic, equipotent and holographic fashion robust to both endogenous and exogenous perturbations.

END BRAIN 2.8

The fact that the brain disposes of above-mentioned properties is usually explained in terms of « neural » connectionist theories whose validity is well demonstrated by multitudes of anatomical observa-

1 Cellular memory being an exception with which we cannot deal here.

tions and clinical experiments. And it is indeed true that when observed by a microscope, at a strictly material « level of abstraction » (LoA[2]), the brain is nothing else than a ball-sized walnut of wetware consisting of approximately one hundred miliard neural cells.

Subsequently, one when one adopts a more computational LoA, one easily comes to conclusion that the substrate of mutually interconnected neural cells can indeed yield a device capable of strongly parallelized computation. « Neural networks », « backpropagation », « stochastic gradient descent » – all these notions offer us useful conceptual tools which enable us to bridge the « objective » material reality of the brain with information-processing, i.e. « computational » faculties of the mind.

Of validity of connectionist Level of Abstraction

Ability to « learn» can be, of course, considered to be such computational faculty. And the Hard Thesis states that *above the « material » and « computational » LoA from which the brain can be intepreted, there exists also a scientifically sound « evolutionary » LoA at which learning can be functionally conceived as being both structurally as an instance of an evolutionary process* – i.e. a process involving reproduction, variation and selection of information-carrying entities. If HT is valid, the functions of one and same brain could be thus ideally intepreted by the prism of « material », « computational » and « evolutionary » LoAs at the same time.

An « evolutionary » LoA can be considered to be scientifically sound only if it does not contradict empiric knowledge – in the case of HT, it should not contradict the anatomical and clinical knowledge concerning the brain. Nor should it contradict the connectionist « computational » theory. What we already know about « brain » and « learning » should rather be consistent with the meaning of HT. Is it the case ?

It can be, if ever the conditions of HT validity (c.f. 2.4 – 2.7) would be found consistent with current neuroscientific knowledge. The last condition «The resources of environment, within which the learning occurs, are finite» seem not to pose a problem since both environment about which we speak here – the brain itself – and its material and energetic resources are finite : even in case of a most abnormal human being, a brain can simply not consume more than 25-30 % percent of one's energy. Hence, it is impossible, for a human brain as an energy-consuming system, to go beyond the upper bound of cca 500 kilocalories per day (Mink and Blumenschine, 1981). In this sense what holds for energy holds, mutatis mutandi, also for limits of nurturing chemical substances which the brain must metabolize to keep its vital functions in equilibrium. Their quantity is limited – even in case of a well-nurtured healthy individual are brain's material resources finite.

Of brain's consumption of resources.

2 C.f. Philosophy of Information (Floridi, 2011, pp. 46-58) for a more exhaustive definition of « Level of Abstraction ».

The third condition, i.e. «the process of replication of information-encoding entities can be influenced by stochastic phenomena which cause an unpredictable structural variation» also does not seem to be very problematic when we consider the fact that the replication does physically occur within its environment – i.e. in the brain – and that its environment is an energy&information-processing system. It is not problematic, because of 2nd law of thermodynamics.

2.9 2ND LAW OF THERMODYNAMICS (DEF)

« Every process occurring in nature proceeds in the sense in which the sum of the entropies of all bodies taking part in the process is increased. In the limit, i.e. for reversible processes, the sum of the entropies remains unchanged.» (Planck, 1926)

2ND LAW OF THERMODYNAMICS 2.9

Human brain, when understood as a physical system, is not an exception to this law. Nor are its components – lobes, neural circuits, neurons, axons, dendrites, receptors, proteins etc. Whenever and wherever is information processed, energy transforms its form and some residual heat is generated. Heat is energy with increased entropy – in its essence it is kinetic energy kicking the surrounding molecules in all directions. As such, it can induce unexpected «unpredictable structural variation» of brain tissue's molecular substratum. Thus, the very fact that the brain is an energy-consuming device implies a possibility of decay and loss of information encoded in brain's materia.

Of brain and heat

Heat aside, brain is also confronted with other sources of «unpredictable structural variation». From quantum phenomena, free radicals and different toxins contained in food and air to purely cognitive noise entering the brain through sensory channels – both brain's processes and structures are constantly confronted with both endo & exo-genous sources of «unpredictable structural variation». If a sort of replication of information-encoding entities would take place in the brain, it would be highly improbable that it would not be also subject to such variation. Thus, when it comes to human brain, we consider the third condition of HT's validity 2.6 as fulfilled.

Of intracerebral sources of variation

By its very definition, any activity of a material system involves consumption of energy and learning, understood as «information-processing mind-transforming constructionist process» 2.2 is not an exception. Thus, in case of a material system like brain, can the second condition of HT's validity, i.e. «learning involves information-encoding entities which consume resources in order to replicate» 2.5 be considered to be necessarily valid if first condition of HT's validity , i.e. «Learning involves the information-encoding entities which replicate» (2.4), is itself valid.

But now the thing gets complicated since, as far as we know, existence of such « reproduction of information-encoding structures » within the brain has not been, as of November 2014, demonstrated with sufficient certitude. At least not directly[3]. But note that such reproduction of information is at least indirectly implied at least since 1950s whence neuroanatomic observations, which were primarily concerned with effects of brain lesions upon the resulting behaviour of the brain, have demonstrated that information in brain is stored in a non-local fashion. As Karl Lashley, one amongst the biggest neuroscientists of 20th century who spent most of his life studying equipotentiality (i.e. the capacity of any part of functional area to solve a particular task), once put it: «The equivalence of different regions of the cortex for retention of memories points to multiple representations. Somehow, equivalent traces are established throughout the functional area.» (Lashley, 1950, pp. 28)

Of indirect evidence for intracerebral reproduction of information

There are at least two possible interpretations of such «non-local storage of information» based on "equivalent traces" and/or "multiple representations". The first one is « connectionist »:

2.10 CONNECTIONIST EXPLANATION OF NON-LOCALITY (TXT)

Information stored in the brain cannot be localized at one particular spatial locus because it is spatially distributed among multiple synapses of the neural network.

END CONNECTIONIST EXPLANATION OF NON-LOCALITY

In other terms, the connectionist interpretation states that a material representation of a cognitive structure S (or a cognitive function F) cannot be localized to this place « here », because it is also partially encoded « there » and « there » and « even there ». From « connectionist » perspective it is indeed this distribution, this decentralization of information among synaptic weights which gives to a neural network both its robust character as well as its capacity of generalization.

But there exists also a second interpretation of the fact that information in brain is not stored on a one specific place:

2.11 ALTERNATIVE EXPLANATION OF NON-LOCALITY (TXT)

Information stored in the brain cannot be localized at one particular spatial locus because it is materially encoded at multiple loci.

END ALTERNATIVE EXPLANATION OF NON-LOCALITY

3 In 8.6 we shall see some theories interpreting certain neural phenomena not only as «reinforcement» but also as «reproduction of information».

From this other perspective, brain stores the material representation of a cognitive structure S (or a cognitive function F) in multiple alternative places and|or in multiple forms. A trivial example illustrating the *essential* difference between two approaches is presented on Figure 2 which visualises "connexionist" and "alternative" representations of corpus containing fours tokens "MABA" and one token "MAPA".

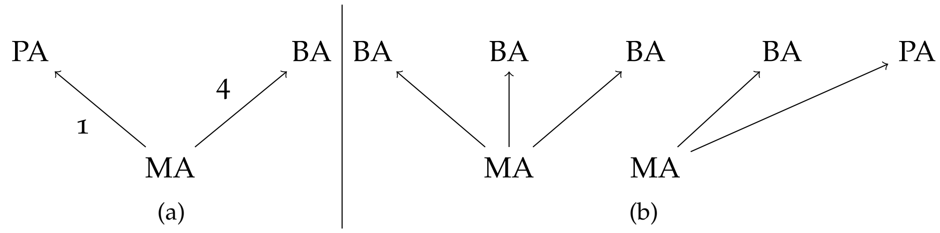


Figure 2: Distinction between "connectionist" (a) and "alternative" (b) representations of the same data. It is evident that the latter allows for more structural variation than the former.

Far for being mutually exclusive with the first intepretation of brain's non-locality, such «alternative» representation has one advantage and one disadvantage. Given that each particular locus encodes a particular instance of structure S (or function F), i.e. S_1 S_2, S_3, *any individual instance can be modified while leaving all others branches intact*. Every instance is thus an independent individual with an individual history: that is a non-negligeable advantage. The disadvantage is that in order to get encoded, such "alternative" representations need more space than "compressed" connectionist representations which superpose distinct instances one atop another in order to yield one ultimate representation.

Note that such diverse individual instances could be well confounded by an external observer who – if (s)he had not been equipped with fine-grained resolution imagery apparatus - could easily believe to witness only the activation of one and only neural circuit S. But the closer inspection shall reveal -so we speculate - that *the same stimulus and the same response is to be followed, respectively preceded, by activation of distinct neural loci*. Such observation could be potentially interpreted as an empiric evidence of «alternative» interpretation of non-local encoding of information in the brain.

Of utility of redundancy in organic systems.

It is true that from certain point of view, such «alternative» way of storage of information in multiple cerebral loci could be considered as redundant. But redundancy does not necessarily mean suboptimality. In a body of a multicellular organism, for example, is the complete genetic code stored practically in nucleus of every single cell (erythrocytes and trombocytes of higher vertebrates excepted). And it is verily this very fact that every cell contains the schema for the whole, which gives, among other properties, to such an organism a somewhat «miraculous» capacity to regenerate itself. This being said, it can be further speculated whether the «miraculous» property of brain called «plasticity» - i.e. the fact that the brain can, in some extent, restore the original knowledge even if some part of brain was damaged or even fully lesioned – can be also explained, mutatis mutandi, in terms of redundant storage of information at multiple loci.

Now back to question discussing the possibility of reproduction of information-carrying structures within the brain. If we accept that the «alternative» hypothesis concerning the brain's faculty to store information non-locally is at least partially valid, we may subsequently pose a question: «but how comes, that multiple individual instances of information S are stored at distinct loci L_1, L_2, L_3 ?». A possible answer : «because sometimes, somehow[4], information from L_1 is copied into L_2» could pave the way to experiments whose objective shall be to verify the 1st condition of HT's validity (2.4) demanding that learning should somehow involve the reproduction of information-encoding entities.

Note that for the purposes of level of abstraction at which Hard Thesis is postulated, it is secondary whether the replication of information-encoding structure is materially realized as a creation of new material synapses, or synchronization of firings of neural circuits, or modification of oscillatory properties of certain fields, or something completely different. The only thing, we believe, which is currently needed to offer ultimate neuroscientific evidence for the statement «learning is a form of evolution » is to directly observe spontaneous intracerebral reproduction of one concrete chunk of information ; from one locus to another. More formally, such a « reproduction » could be considered as taking place if, at spatiotemporal locus $T_1 L_1$, one would observe the emergence of « child » representation R_1 which is at least partial isomorph with « parent » representation R_0 which has been already observed at spatiotemporal locus $T_0 L_0$ and is still observable at some spatial locus in time T_1. Such ensemble of observations would indicate that at least some part of information E was copied from L_0 to L_1 in a way which leaves practically intact the original representation R_0.

But until such neuroscientific evidence is given, the first condition of HT's validity cannot be considered as sound on empirical grounds. This logically implies the consequence that the whole Hard Thesis must be -given the current state of neuroscientific knowledge- considered as nothing else than a speculative conjecture. The only thing which we can do to make this dissertation less speculative, and hence more scientific, is to soften the Thesis by reducing the scope of the domain upon which it applies.

4 For example during phases of «dreaming» or other activities of "repeating" and "rehearsing".

3 SOFT THESIS

Soft Thesis (ST) is expressed as follows :

> « learning can be successfully simulated
> by means of evolutionary computation »

ST simply postulates a sort of explanatory adequacy between « learning » and « evolutionary computation ». It does not, as HT does, express the statement about ontological position of « learning », it does not state what learning « is ». It simply states that the behavior of a system whose functioning is in agreement with principles of « evolutionary computation » could ressemble to behavior of a system which is considered to be "learning".

ST postulates explanatory and not ontological adequacy

3.1 EVOLUTIONARY COMPUTATION (DEF)

« Evolutionary computation uses computational models of evolutionary processes as key elements in the design and implementation of computerbased problem solving systems.» (Spears et al., 1993)

END EVOLUTIONARY COMPUTATION 3.1

Evolutionary computation (EC) can be thus considered to be a subdiscipline of informatics. This does not mean that the principles of EC should be relevant only to realm of silicon-based computers. It is so, because informatics aims to yield a general theoretical framework for description of information-processing systems, that is, a theory which could be ideally applied on both silicon-based (e.g. computers) and neuron-based (e.g. brain) computational devices[1].

Of EC and material substrate

In practice, however, are hypotheses related to informatic science best studied and most applied in relation to silicon-based universal Turing machines. *Voici* reasons why it is so:

Advantages of EC simulations in silico

- Minimal ethical concerns : it is considered ethically completely acceptable to program one's computer ; it is less so to do that with one's neighbor, or his intestinal flora.
- Full initial control : a programmer can control practically all initial states of his informatic model as well as the initial form of rules according to system shall subsequently behave.

1 And potentially to other types of computational devices. As of 2014, particularly promising seem to be devices developped in the discipline of biomolecular computing. Note that the very essence of these devices (e.g. DNA-computers) is particularly favorable to problem-solving by means of evolutionary computation.

- Reduced cost : construction, execution and evaluation of a model in silico is generally much less resource-demanding than construction, execution and evaluation of such model in vitro or in vivo.

Advantages of performing evolutionary simulations in silico

Since EC is a subdiscipline of informatics, it follows that above-described utility of silicon-based machines for informatics would be also appreciable in the domain of evolutionary computing. In fact, especially due to moral and security concerns, in silico seems to be the only way how living evolution can be empricially studied on a time scale directly percievable and interpretable by practically any human observer able to run a program on a computer. For this reason, when Soft Thesis relates the EC to « learning », it is principially a silicon-based computer which is supposed to be the subject of the « learning » process. With exception of Part iii - where we shall mainly discuss learning process as instantiated in human children - shall be, in the rest of this dissertation, computer understood as an entity capable of learning.

In 8.7 we shall discuss EC in somewhat closer detail. There, we shall also introduce the most important EC paradigms like «genetic algorithms» (8.7.1), «evolutionary strategies» (8.7.2) and «genetic programming» (8.7.3). But the particularities of these diverse approaches are not of a great interest for the subject which interests us in this chapter, that is : to elucidate the of meaning of the Soft Thesis. In order to do so, the term « successfully simulated » should be defined.

3.2 SUCCESSFUL SIMULATION (DEF)

A process P can be said to be « successfully simulated » by a system S *iff* the way, how outputs oS_1 , oS_2, . . . oS_n of the system S (given the inputs iS_1 iS_2... iS_N) are generated is isomorph, at certain Level of Abstraction, to the way how process P reacts to stimuli iP_1,iP_2 ... iP_N when generating outputs oP_1, oP_2 ... oP_n. Morphism $iP_X \rightarrow iS_X$ can be understood as representational mapping of inputs from the domain of the process P (i.e. « reality ») into the domain of the simulation S.

END SUCCESSFUL SIMULATION 3.2

Of stimuli and input

In less formal terms, a simulating system can be told to perform « successful simulation» if and only if tends to react to sequences of its inputs in the same way as does the process-which-is-simulated react to sequences of stimuli with which it is confronted. Note that in order to distinguish the two, we use the term "stimulus" when we speak about the data entering the original physical process-which-is-simulated and we use the term "input" when we speak about the data which enters the simulation. In light of this definition, the Soft Thesis practically postulates that by implementing the precepts of Evolutionary Computation (Section 8.7), one can construct computational models which shall gradually transform inputs into outputs in a way that would be, for an external observer, indistinguishable from the mappings gradually produced by the process of « learning ».

Let it be underscored that the above-mentioned definition speaks not only about simulating the outputs (results) of the process; it speaks also about the manner by means of which such results are obtained. It demands not only external but also internal adequacy between the simulation and the process which is being simulated. That is, NOT ONLY should the simulation yield the outputs which are the most accurate - i.e. ressemble the most the observable behaviours of the system - BUT ALSO should execute the input -> output mapping in a similar way. In case of tentatives aiming to simulate human cognitive processes, we find it useful to speak about such "internal adequacy" in terms of cognitive plausibility.

Morphisms among morphisms

3.3 COGNITIVE PLAUSIBILITY (DEF)

« We label as "cognitively plausible" a model which tends to address some basic function/skill of human cognitive system not only by simulating, in a sort of "black-box apparatus", the mapping of inputs (stimuli, corpus data etc.) upon outputs (observed behaviors, results etc.), but also tends to faithfully represent – at least when interpreted from a certain LoA- the way how the respective function/skill is accomplished by a real human mind.» (Hromada, 2014b)

COGNITIVE PLAUSIBILITY 3.3

We believe that it is often pertinent to ask the question "is computational model M of process P cognitive plausible?". In case of process of "learning" and its computational "machine learning" (ML) counterparts, an analysis through the prism of "cognitive plausibility" could potentially yield surprising results: while many ML models perform more than well in a task which was previously the domain of exclusively human learning, they are far from being cognitively plausible. Extent in which the model successfully simulates the real process (i.e. its performance) and an extent in which the model does it in a way similiar to human mind (i.e. its cognitive plausibility) demarcate two independent axes which are not to be confounded. Engineers interested only in attaining the best results (i.e. the most adequate outputs, given the inputs) can often ignore the manner by means of which a natural system solves a given problem. On the other hand, researchers aiming to understand the functioning of the natural system are often more ready to accept lesser performance of their model if ever it seems to exhibit the same properties and faculties as the natural system. Only in rare cases do such engineering, i.e. result-oriented, and scientific, i.e. knowledge-oriented, axes converge.

Of researchers and engineers

END SOFT THESIS 3

4 SOFTER THESIS

An important question was left unanswered during our discussion of the Soft Thesis. That is : what shall be the object of learning which is supposed to be successfully simulable by means of Evolutionary Computation ? What shall be the nature of stimuli ip1 ip2 ... ipN entering the learning process we aim to simulate ?

To concretely address this question, we are, once again, obliged to soften the Thesis somewhat more, thus obtaining the Softer Thesis which can be expressed as follows:

«learning of natural language can be successfully simulated by means of evolutionary computation»

Contrary to ST, which relates EC to a very broadly defined notion of « learning », does the S^2T specify the object of «learning» which is supposed to be EC-simulable. It is learning of natural languages.

4.1 NATURAL LANGUAGE (DEF)

Natural language is a system composed of prosodic, phonetic, phonologic, morpohologic, syntactic, semantic and pragmatic structures and principles which allows human beings to encode messages in a way that is comprehensible to other human beings.

END NATURAL LANGUAGE 4.1

Further definitions related to natural language, notably those *-ic* terms, shall be presented in the 9.2 for they are not inevitably needed for eludication of S^2T's meaning. What we consider of bigger importance here is to introduce the reasons which have motivated us to study evolutionary computation in relation to learning of natural languages.

4.2 WHY NATURAL LANGUAGE ? (APH)

Of essence of humanity

Among all the faculties which distinguish man from other animals is the mastery of language potentially the most salient one. This was already well-known for the ancients among which Aristotle, for example, defined man as ζῶον λόγον ἔχον, «an animal which word has». Centuries later, Wittgenstein (1953) had indicated that whole philosophy, and potentially even more, can be understood as a realisation of some sort of perenial «language game»...

In the meantime, on the very frontier between «natural» and «human» science, emerged linguistics : the science whose object of study is language, understood as a system, and whose objective is to understand principles governing such a system. During the century which followed after de Saussure (1916) presented linguistics as a mainly positivist study of diverse forms of linguistic structures, linguistics had refined its methodology and terminology in a way such broad and deep that currently -as of 2014- among all other sciences studying one specific domain of human activity, linguistics has practically no equal in both quantity and quality of scientific knowledge which has been already accumulated.

Of linguistics

Thus, one reason why we have chosen to focus on the natural language is purely pragmatic one: natural languages are well-studied. For us it principially means that we are not obliged to «reinvent the wheel» and can instead use the already existing methodology and terminology, refer to past observations and experiments and potentially exploit the established corpora.

1st reason

Notably the discipline of developmental psycholinguistics, with its focus on the process of «language development» (Section 9.1) as well as an increasingly popular discipline of Natural Language Processing (NLP, Section 10.3), located on the border between linguistics and computer science, seem to be of particular importance in regards to potential proof of validity of S^2T.

The second reason for focusing our interest on natural language is related to the role which natural language seems to play in development of every healthy human individual. This role is considered to be non-negligeable by those who consider the language to be the very fundament of human society ; and is considered to be vital by those who know that on its own, i.e. without society's protective matrix, a human individual – and especially a human child – simply could not survive and/or develop full capacities of a self-realized member of homo sapiens sapiens species. Simply stated, language is a phenomen present in all cultures and as such can be considered to be the anthropological constant *par excellence*.

2nd reason

By having already mentioned philosophy, anthropology, and linguistics, we consider it important to underline that the topic of natural language seems to be recurrent in all cognitive sciences.

Neuroscience, for example, had fully established itself as an empiric science the very day when Broca (1861) realized that the damage of brain's inferior frontal lobe of the dominant hemisphere leads to problems in production of language (he was later followed by Wernicke (1874) who noticed that the damage of superior temporal gyrus leads to troubles in language comprehension).

Language plays also important role in both psychotherapy and psychology. In both Freundian and Jungian psychanalysis, in Rogerian person-centered psychotherapy, in Frankl's logotherapy or individual

Role of language in psychology

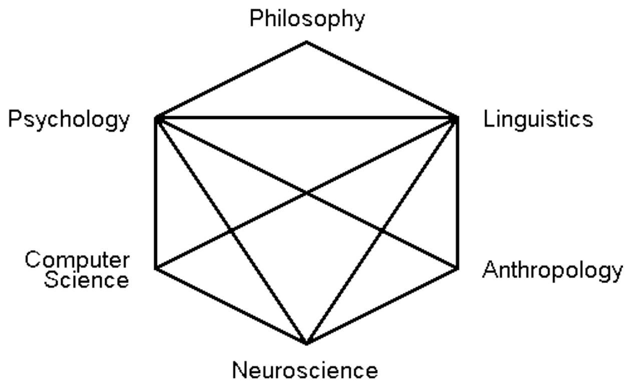


Figure 3: Cognitive Hexagram

psychology Adler (1976) and possibly in many other psychotherapeutic systems, language is considered to be therapeutic tool of utmost importance. What is more, in a very sound psychological "theory of multiple intelligences", as articulated by Gardner (1985a), is man's faculty to understand and produce linguistic utterances important enough so that it merits to obtain the label of «verbal-linguistic» intelligence. Along with six other intelligences, this «linguistic intelligence» is considered to be the basic computational module of human cognitive system. Also within a theory coming from a different (russian) tradition, that of Vygotsky (1987), is language considered to be a crucial component of man's psyche: in Vygotski's framework, in fact, is the *thinking itself* understood as a so-called *inner speech*.

The centroid of the hexagram

All this arguments lead us to belief that natural language is a topic which is localized very close to the centroid of the hexagram delimiting the object of study of all cognitive sciences (depicted on Figure 3). In one way or another, explicitely or implicitely, all cognitive sciences deal with natural language.

On their own, these two reasons, «language is well-studied» and «language is central» would yield, we believe, sufficient an answer to the question «Why does S^2T relate evolutionary computing with learning of natural language and not, for example, with learning of deer-hunting or learning of swimming?"

3rd reason

But there is another, AI-related, reason for which we consider the study of language learning to be of particular importance in relation to evolutionary computing and/or computer science. More concretely, similarily to Turing (1950), who saw in language a means how to address the question «Can machines think ?» in an answerable way, we see in natural language a potentially first solid bridge between the realms of artificial and human beings.

END WHY NATURAL LANGUAGE? 4.2

END SOFTER THESIS 4

5 SOFTEST THESIS

The Softest Thesis (S^3T) is expressed as follows :

«Ontogeny of toddlerese can be successfully simulated by means of evolutionary computation.»

In this definition, the term "ontogeny" is used in the sense practically synonymous to "learning", the sole difference between the two being our intention to mark the notion that toddlerese is not only passively learnt, but that it emerges and is actively constructed. When it comes to toddlerese itself, it is hereby defined as:

5.1 TODDLERESE (DEF)

Toddlerese is a transitory protovariant of a natural language which is transferred from minds of human adults into the mind of a child by means of repetitive exchange of sequences of contextualized symbols.

END TODDLERESE 5.1

Thus, the term "toddlerese" has a meaning similiar to meaning of terms widely-used terms like "first language" or "mother language". But contrary to these terms -which are used to denote not only the language which develops but also, and mainly, the end-state language resulting from such development- the term "toddlerese" is conceived to denote only a certain transitory state, or a sequence of states in development of such "first language". In other terms, mother language stays develops in man's mind for the rest of (her | his) life but toddlerese language L_T gradually disappears, or at least gets latent, in parallel with child's cognitive and physiological development away from the toddler state. The term "protovariant" is used to mark even more both temporariness as well as its function of a base for a full-fledged language which shall unfold from L_T in mid-childhood and later.

Toddlerese is a transitory language

More concretely, we define -for the purpose of this Thesis- toddlerese as language L_T emergent from child's interactions with the world within the temporal interval (0,2;6) years, id est between birth and two and half years of age[1].

Age range of toddlerese

1 In order to facilitate bridging between computer science and developmental psycholinguistics, we shall not use the decimal notation, but a year;month;week notation to speak about child's age (e.g. 2;3;1 when speaking about child which is two years, 3 months and one week old)

Thus, the term "toddlerese" has a meaning similiar to meaning of terms widely-used terms like "first language" or "mother language". But contrary to these terms -which are used to denote not only the language which develops but also, and mainly, the end-state language resulting from such development- the term "toddlerese" is to denote only a certain transitory state, or a sequence of states in development of such "first language". In other terms, mother language stays active in man's mind for the rest of (her|his) life but toddlerese language L_T gradually disappears, or at least gets latent, in parallel with child's cognitive and physiological development away from the toddler state. The term "protovariant" is used to mark even more both temporariness as well as its function of a base for a full-fledged language which shall unfold from L_T in mid-childhood and later.

Of repetitivity and reproduction

Another important notion included in the definition of « toddlerese » is repetitivity. Repetition of symbol S can be understood as a sort of « reproduction » along the temporal axis and in following chapters we shall often interpret phenomena, which repeat themselves, not only as reactivation of the original schema, but rather in terms of activity of multiple schemas which are being reproduced. We repeat ; we restate ; we reiterate: at a certain LoA, repetition can be understood as a form of reproduction.

But the most important terms of definition **??** are those of «transfer» and «exchange». Initially, these terms seem to denote divergent concepts: the term «transfer» carries with itself the conotation of somewhat unidirectional movement from the origin (mind of the parent) to the destination (mind of the baby) while the term «exchange» denotes a bidirectional process whereby neither of interactors plays the dominant role and both dispose of faculty to partially influence or fully transform the behaviour of the other. But they can be reconciled through the metaphor of a «mirror».

The mirror metaphor

At first sight, mirror is a completely passive device simply reflecting the objects which project (transfer) their shapes on its surface. But by the very fact that «mirror mirrors», it has also the power to influence the behaviour of the one who is looking in it and thus *entrer en échange avec l'autrui*. It is important to realize that since mirrors can be constructed differently, they can mirror things differently – the image they offer in exchange is thus not only dependent upon the-object-they reflect, but also determined by the material and the way how mirror was physically forged[2].

Child mirrors its parents

Something similar holds, mutatis mutandi, when it comes to transfer of linguistic competence from the parent to the child. By means of diverse neural mechanisms (e.g. «mirror neurons» (Rizzolatti et al., 2008)) does child's plastic brain assimilate information from its environment. We count among the objects of such assimilation also

2 By interpreting «tabula rasa» hypothesis as a particular case of the mirror metaphor hereby introduced, one could partially align the empirist and nativist doctrines.

structures explicitely expressed or implicitely encoded in sequences which child observes and which are, most often, generated by less-plastic and more-crystalized minds of her parents. The child somehow «parses» such information, processes, understands it and acts accordingly. This action is subsequently projected into external environment by diverse means – most prominent of which are undoubtably child's vocal tract and child's facial expressions – and by these means is the very environment transformed. Minds of parents including.

We precise that by introducing the metaphor of the mirror we do not, of course, want to state that child is just a receptive information-assimilating entity passively reflecting its external environment. Such a statement would be completely contradictory to the fact of ceaseless activity which every healthy child continuously demonstrates. This fact of childs activity being in fact so salient, we propose to integrate it in the very definition of what the term «child » means:

5.2 CHILD (DEF)

Child plays.

END CHILD 5.2

It is by game that child mirrors the world; by playing the game which is pure activity without finality. Child sees around (her | him)self the world in movement, then understands that (s)he can also move and thus (s)he moves. Child's way of mirroring is thus principially mirroring by playful action and it is by playful action that the child exerts influence in and upon its environment[3].

Pages which shall follow, and notably the Part iii, shall furnish further illustrations of what we mean by «playful action» in regards to both language learning and evolutionary *tâtonnement*. Other computational language games shall also be introduced, mostly in form of programs able to induce sets of classes (10.4.7) or transcription rules (**??**) from diverse textual corpora. All programs shall apply the principles inherent to « evolutionary computing » in order to furnish some data validating (or falsifying) the hypothesis S^3T.

On the other hand, none of the programs will be able to account for phonetic or pragmatic layers of languages under study. For this reason we are obliged to delimit, for the last time, the scope of our Thesis.

END SOFTEST THESIS 5

3 Notions of «game» and «playfulness» are not the same for adults and children. Adults often consider as hazardous activities which children consider as a game and vice versa, children often consider as serious the sandbox activities which are not at all perceived as such by adults. The transfer of adequate categories «game» and «serious» is an important goal of socialisation and possibly learning in general.

6 OPERATIONAL THESIS

The Operational Thesis (OT) is defined as follows:

> «Learning of toddlerese from its textual representations can be successfully simulated by means of evolutionary computation.»

OT is thus very similar to the softest thesis, the only difference being the specification of the modality of representation of inputs in confrontation with which the toddlerese is supposed to be learnable, in simulation, by means of evolutionary computation. It is precised that such learnable modality is «textual».

6.1 TEXT (DEF)

Sequence of discrete graphemic symbols representing morphosyntactic and semantic contents of natural language utterances.

END TEXT 6.1

This definition principially states that text encodes only subset of information which a normal « hearable » utterance contains. That is : semantic information related to its meaning and sense, and morphpsyntactic information related to its grammatical composition. C.f. sections 9.2.3 and 9.2.4 for discussion of «morphosyntax» and «semantics» respectively.

By specifying the modality of data with which it shall operate, OT has drastically reduced the scope of applicability of the softest thesis. More concretely, by defining « text » as the modality of representation with which we shall confront our computational models, we have left aside the phonetic, phonologic, prosodic and pragmatic aspects of language. That is, aspects of language which have been -during practically all human history- crucial whenever the « speaker » intended to pass information to the « hearer ». It is only during few centuries that the communication by means of text became prominent and only within last decades it became dominant, mainly because of increasing role of computers in our lives. This is at least partially so because computers are essentially machine built for processing of sequences of discrete symbols and that's what a text is – a sequence of discrete symbols. Contrary to flux of spoken language, which is also a sequence, but composed of units whose boundaries are often unclear and whose features overlap.

Text does not have phonetic, prosodic and pragmatic layers.

But the fact that practically no prosodic[1], phonetic or pragmatic information shall be involved in our computational simulations does not mean that these simulations will not be concerned by natural language. On the contrary – it is evident, from experience of every reader, that text indeed is a «communication system which human beings use to express information in a way comprehensible to other human beings» (4.1). In other terms : if message is clear and if productive linguistic competence of the writer overlaps with the receptive linguistic competence of the writer, message shall make it possible that the reader shall understand the writer. In this sense, text can be considered as a valid and functional modality of representation of natural language.

Text is a form of natural language

However, the question « Whether text can be also considered as a modality of representation sufficient for learning of language, and most notably first language ? », is still an opened one. While some existing computational models indicate that at least for certain subproblems of language learning, like POS-induction (10.5) and grammar induction (10.6), the answer can be «yes», empiric observations of first language learning of human children also suggest that prosody and phonology play crucial role (9.2.1) and to ignore them would mean to miss out the crucial component of the language learning process.

But since children which are deaf, and thus without any access whatsoever to prosody or phonology, are able to learn the sign language -and since the sign language ressembles, in the sense that it is visual and sequential, to text - the operational reduction of language to text is potentially not a completely unreasonable one.

Thus, an operational definition **language** → **text** shall be principially used in sections dedicated to computational simulations of language learning. In other sections, however, this reduction shall not be applied and language will be most often discussed in its full extent, i.e. involving its phonetic, prosodic and pragmatic facets.

END OPERATIONAL THESIS 6

1 One can argue that exclamation (!) or question (?) signs add certain prosody to text since they can possibly represent increasing or decreasing tone or accent. This is, however, discutable because prosodical cues are present « along » whole utterance while the interpunction signs are normally located only at sentence's final position.

7 SUMMA I

In this section we had introduced multiple theses which we consider as valid. These theses were discussed in deductive order, i.e. from the most general to the most specific one.

Discussion started with the initial thesis « mind evolves » and definition of mind as « auto-organising set of structures and processes». Because such thesis is so general that one may suspect that it is in fact a tautological statement-of-faith than a verifiable hypothesis, a so-called Hard Thesis was subsequently introduced, stating that « learning is a form of evolution ».

Learning was principially defined as an information-processing constructionist process and it was further precised that the term « evolution » is meant in Darwin-consistent sense, i.e. as an adaptive process based on reproduction, variation of selection of information-carrying structures. What was not yet explicitely said, however, is that both evolution and learning share an important feature : they involve trials and errors.

7.1 TRIAL AND ERROR (DEF)

Most fundamental heuristics based on repetitive confrontation of system's activity with external and internal constraints and demands.

END TRIAL AND ERROR 7

It is generally believed that in learning, trial events are related to other trial events only in a serial, vertical manner - one trial follows another one in time. On the other hand, in an evolutionary process, trials are related to other trials not only in serial (i.e. one generation folows another) but also in parallel (i.e. generation consists of multiple individuals) manner.

The principal sense of the Hard Thesis is to state that such distinction is illusionary and that learning process almost always involves a sort of horizontality, a sort of population of parallely co-existing structures which underlay and determine the observable manifestation of individual "trial". What's more, HT postulates that as in evolution, so in learning are such individual structures endowed with the faculty to reproduce the information which they encode into another locus. It was further postulated that

1. if ever a stochastic phenomenon can cause variation of information content of an individual entity E generated by the reproduction process

2. if ever the information encoded by entity E influences the amount of resources consumed during the reproduction
3. and if ever such multi-iterative reproduction occurs within the environment having only finite amount of resources

then, with logical necessity, a sort of adaptation of entities to their environment shall follow.

After proposing four conditions under which HT can be considered as plausible, it was further discussed whether human brain could be potentially considered as such "environment" for a sort of intracerebral evolutionary process. The brain was primarily defined as a finite physical object storing information in a non-local way. As a physical system, brain is subordinated to laws of physics like 2nd law of thermodynamics: brain generates heat and heat can, with non-zero probability, cause variation of its own material content. Such variation of materia could subsequently result itself in the information of information which the brain encodes. Thus, the very fact that brain is a finite physical system implies that third and fourth conditions of HT's validity - when related to learning faculty of human brain - are to be considered as fulfilled.

Much more problematic are conditions 1 and 2 of HT's validity relating to the question "does brain contain information-encoding structures able to reproduce?". Since reproduction of information-encoding entities has not yet been directly and irefutably observed within the brain, conservative scientists are often reluctant to answer such question in affirmative. On the other hand, an "alternative" (2.11) explanation of well-observed phenomenon of non-local storage of information implies, that a process resembling reproduction, a process copying information to multiple loci could, indeed, take place within the region of brain's wetware. It was also suggested that in natural *ensembles* like organisms, species or even societies, redundancy of information makes often systems more robust against unpredictable perturbations and it was suggested that same "robustness through redundancy" principle holds, mutatis mutandi, also for human mind.

Unfortunately, the questions raised by HT are too wide to be adressed, in extent they merit, in a limited scope of this dissertation. For this reason, the Hard Thesis is reduced into the soft form which states that learning can be simulated by means of evolutionary computation. ST thus does not postulate the ontological adequacy between nature of evolutionary and learning processes - it simply postulates that computational models of the former can successfully simulate the latter. The notion of successful simulation was defined in terms of isomorphism between input-to-output mapping of the simulation and stimulus-to-reaction mapping of the process-which-is-simulated. The need to create not only externally but also internally adequate computational models of human faculties was also discussed. By introducing the notion of cognitive plausibility, we have proposed to focus not only on result but also on the path which leads to attainment of the result (Section 3.3). Thus, when considering the realm of machines, ST postulates that there exist at least certain class of problems -usually solved by means of traditional "machine learning" techniques- which could be also solved by means of evolutionary computation with similar or better results. And whose manner of functioning ressembles the manner of functioning of the system which is simulated.

A so-called Softer Thesis have subsequently precised that learning of natural languages is such problem. Natural language was definedin a most liberal way as "communication system which human beings use to express information to other human beings" (Section 4.1). Natural languages were chosen as the topic of our interest for three principal reasons: Primo, natural languages are well-studied. Secundo, natural languages are thematized, in one way or another, by all cognitive sciences. Tertio, the canonical (Turing's) method to answer AI's central question "Can machines think?" is principially a test evaluating machine's mastery in simulation of understanding and production of natural language utterances and discourses.

Since the expression "learning of natural language" can cover too many phenomenon, the S^2T is further transformed into the Softest Thesis (S^2T) which speaks only about the "learning of first language". First language is defined as a communication system transferred from the mind of the parent into the mind of a child by means of repetitive exchange of sequences of symbols. Serial - in contrast with parallel, sequental and repetitive nature of first language was discussed and the apparent contradiction between unidirectional "transfer" and bidirectional "exchange" was subsequently reconciled by means of the "mirror metaphor". Human child was defined in terms of its most distinctive propensity, i.e. propensity to "play", to execute activity which lacks the absolute finality.

The last thesis which have been presented is the Operational one. This specifies that the modality of representation, with which the "first language learning" evolutionary computation algorithms will be confronted, shall be textual. Given that text does not include practically any phonetic, prosodic or pragmatic layer, the complexity of the first language learning from text could be substantially reduced. The question whether such reduction is not too strict was also addressed.

By positing 6 theses of varying degree of universality - i.e. Initial, Hard, Soft, Softer, Softest and Operational - we have delimited the level at which the rest of this dissertation shall operate. By defining terms like *evolution, learning, form, brain, 2nd law of thermodynamics, evolutionary computation, successful simulation, cognitive plausibility, natural language, first language, child and trial & error* we have demarcated the basic form of a prism - a theory- through which one could see that the theses we posited hereby are, indeed, valid.

This theoretical prism shall be polished in the following part.

Part II

PARADIGMS

Ideas are never static but develop across time and context, constantly cross-fertilizing with other currents of thought.

— Edwin F. Bryant

This part shall start the crossover of three seemingly unrelated scientific paradigms.

In its initial chapter devoted to Universal Darwinism, we shall introduce scientific disciplines and their respective theories, which are either derived from - or at least consistent with - Darwinian Theory of evolution, understood as gradual development of populations of information-carrying structures. Thus, not only biological evolution shall be discussed, but also evolutionary and genetic epistemology and psychology, memetics, neural darwinism and different branches and sub-branches of evolutionary computation.

In the subsequent chapter, devoted to Developmental Psycholinguistics, we shall introduce the fascinating field of study of acquisition of first language by human children. After definition of few necessary notions we shall bring to reader's attention towards few widely accepted facts and do a brief historical overview of most important language-acquisition theories. More concretely: associanist, behaviorist, nativist, constructivist and sociopragmatic theories shall be mentioned and thematised.

The last chapter of this part shall invite the reader into the realm of Computational Linguistics and Natural Language Processing. After brief introduction into Formal Language Theory and its Grammar Systems Theory variant, the discussion shall be focused on computational problems of concept construction, part-of-speech induction and grammatical inference. Some state-of-the-art computational models aiming to solve these problems shall be described in closer detail in order to pave the theoretical path towards future evolutionary models of first language acquisition.

8 UNIVERSAL DARWINISM

Universal Darwinism (UD) is a scientific paradigm regrouping diverse scientific theories extending the Darwinian theory of evolution and natural selection (Darwin, 1859) beyond the domain of biology. It can be understood as a generalized theoretical framework aiming to explain the emergence of many complex phenomena in terms of interaction of three basic processes:

1. variation
2. selection
3. retention

According to UD paradigm, interaction of these three components yields « universal algorithm valid not only in biology, but in all domains of knowledge where we can extract informational entities – replicators, which are able to reproduce themselves with variations and which are subjects to selection» (Kvasnicka and Pospichal, 2007).

This generic algorithm is nothing else than traditional Evolutionary Theory (ET) which, when when considered as substrate-neutral, can be applied to such a vaste number of scientific fields that it has been compared to a kind of « universal acid » which «« eats through just about every traditional concept, and leaves in its wake a revolutionized world-view, with most of the old landmarks still recognizable, but transformed in fundamental ways» (Dennett, 1995).

UD is a source of both theoretical inspiration and practical precepts for many scientific disciplines, technological methods or artistic endeavours. The most prominent include:

1. biology
2. evolutionary art, e. psychology, e. music, e.linguistics, e.ethics, e.economics, e.anthropology, e.epistemology, e.computation
3. sociobiology (Wilson, 2000)
4. memetics (Blackmore, 2000)
5. quantum darwinism, neural darwinism, psycho darwinism
6. artificial life

et caetera. We shall now discuss some of them.

8.1 biological evolution

Evolutionary Theory was born when young Charles Darwin realised that the « gradation and diversity of structure» (Darwin and Bettany, 1890), which he had encountered among mockingbirds of Galapagos islands, could be explained by natural tendency of species to « adapt to changing world ». Parallely to Darwin's work which was gradually clarifying the terms of variability and its close relation to environment-originated selective pressures, Gregor Mendel was assessing statistical distributions of colours of flowers of his garden peas in Brno in order to finally converge to fundamental principles of heredity . But it was only in 1953 when the double-helix structure of the material substrate of heredity of biological species – the DNA molecule – was described in article of (Watson et al., 1953).

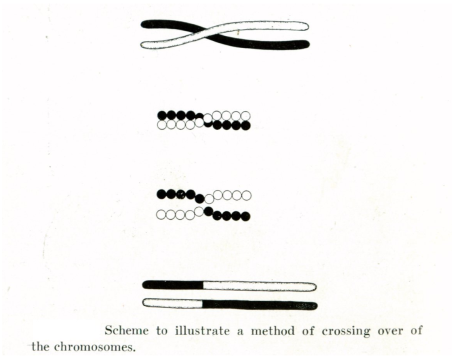
In simple terms : In the DNA molecule, information is encoded as a sequence of nucleotides. Every nucleotide can contain one of four nucleobases, it thus ideally carry 2 bits of information. Continuous sequence of three nucleotids gives a « triplet » which, when interpreted by a intracellular « ribosome » machinery, can be « translated » into an amino-acid. Sequences of amino-acides yield proteins which interact one with another in biochemical cascades. The result is a living organism with its particular phenotype aiming to reproduce its genetic code.

If, in the given time T there are two organisms A and B whose genetic code differs in such an extent that their phenotype differs, and if ever the phenotype of organism A augments probability of A's survival and reproduction in the external world W, while the B's phenotype diminishes such probability , we say that the A is better adapted to world W than B, or more formally that fitness(A) > fitness(B). Evolutionary Theory postulates that in case that there is a lack of resources in world W, descendants of the organism B shall be gradually, after multiple generations, substituted by descendants of a more fit organism « A ». This is so because during every act of reproduction, the material reason for having a more fit phenotype - the DNA molecule – is transferred from parent to offspring and the whole process is cumulative across generations.

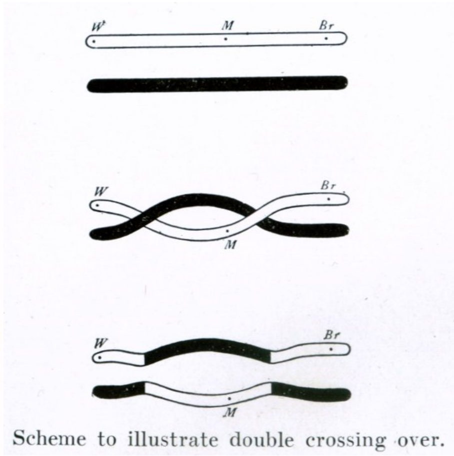
It can, however, happen, that the world W changes. Or a random (stochastic) event – a gamma ray, the presence of a free radical - can occur which would tamper A's genetic code. Such an event – called « mutation » - shall result, in majority of cases, in decrease of A's fitness. Rarely, however, can mutations also increase it.

Another event which can transform the genetic sequence is called « crossover ». It can be formalised as an operator which substitutes one part of genetic code of the organism A with corresponding sequence of organism B, and vice versa, the part of B with the corresponding part of A. It is indeed especially the crossover operation,

Figure 4: One-point and two-point crossovers. Figures reproced from Morgan (1916).



(a)



(b)

first described by in the article (Morgan, 1916), which is responsible for « mixing of properties » in case of a child organism issued from two parent organisms. In more concrete terms : the genetic code of such « diploid » organisms is always stored in X pairs of chromosomes. Each chromosome in the pair is issued from either father or mother organism which, during the process of meiosis, divide their normally diploid cells into haploid gamete cellls (i.e. sperms in case of father and eggs in case of mother). It is especially during the first meiotic phase that crossover occurs, the content of DNA sequence of two grand-parents being mixed and mapped during crossover operation into the chromosome contained in the gamete which, if lucky, shell fuse with the gamete of another parent in the act of fecondation.

Resulting « zygote » is again diploid, contains mix of fragments of genetic code originally present in the cells of all four grand-parents of the nascent organism. Zygote subsequently exponentially divides into growing number of cells which differentiate from each other according to instructions contained in the genetic code which are triggered by biochemical signals coming from cell's both internal and external environment. If the genetic code shall endow the organism with properties that will allow it to survive in its environment until its own reproduction, approximately half of the genetic information contained in its DNA shall be transfered to the offspring organism. If not, the information as such shall disappear from the population with death of the last individual who carries it.

END BIOLOGICAL EVOLUTION 8.1

8.2 EVOLUTIONARY PSYCHOLOGY

We have already quoted Darwin's statement that asserted that psychology in the distant future shall "be based upon a new foundation of the necessary acquirement of each mental power and capacity by gradation". While two possible intepretations of this Darwin's idea exist, the discipline Evolutionary Psychology (EP) focuses only on the first one. It aims to explain diverse faculties of human soul & mind in terms of selective pressures which moulded the modular architecture of human brain during millions of years of its phylogenetic history. Its central premises state :« The brain's adaptive mechanisms were shaped by natural and sexual selection. Different neural mechanisms are specialized for solving problems in humanity's evolutionary past.» (Cosmides and Tooby, 1997).

In more concrete terms, Evolutionary Psychology explains quite successfully phaenomena as diverse as emergence of cooperation and altruistic behaviour (Hamilton, 1963) ; male promiscuity and parental investment (Trivers, 1972) or even the obesity of current anglo-saxxon population (Barrett, 2007). All this and much more is explained as a result of adaptation of homo sapiens sapiens (and its biological ancestors) to dynamism of its ever-changing ecological and social niche.

Thus, in the long run, EP tends to explain and integrate all innate faculties of human mind in the evolutionary framework. The problem with EP, however, is that in its grandious aim to « assemble out of the disjointed, fragmentary, and mutually contradictory human disciplines a single, logically integrated research framework for the psychological, social, and behavioral sciences » (Cosmides and Tooby, 1997), it can sometimes happen that EP posits as innate, and thus explainable in terms of biological natural selection, cognitive faculties which are not innate but acquired. Thus it may be more often than rarely the case that whenever it comes to the famous "nature vs. nurture" (Galton, 1875) controversy, evolutionary psychologists tend to defend the nativist cause even there, where it means to commit a epistemological fallacy to do so[1].

And what makes things even worse for the discipline of Evolutionary Psychology as is currently performed is, that the forementioned Darwin's precognition has, asides the nativist & biological one, also another intepretation.

Id est, when Darwin spoke about mental powers and capacities acquired by gradation, one cannot exclude that he was speaking not only about gradation in phylogeny of species, but also ontogeny of an individual. END EVOLUTIONARY PSYCHOLOGY 8.2

---

1 If ever we accept the notion of falsifiability as an important criterion of accpetation or rejection of the scientific hypothesis (Popper, 1972), many hypotheses issued from EP would have to be rejected because, since being based in the distant past which is almost impossible to access, they are less falsifiable than hypotheses explaining the same phaenomena in terms of empiric data observable in the present.

8.3 MEMETICS

Theory of memes or memetics is, in certain sense, a counter-reaction to Evolutionary Psychology's aims to explain human mental and cognitive faculties in terms of innate propensities. Similiarly to EP, memetics is also issued from the discipline of sociobiology which was supposed to be « The extension of population biology and evolutionary theory to social organization» (Wilson, 2000). But contrary to both EP and sociobiology, memetics does not aim to explain diverse cultural, psychological or social phenomena solely in terms of evolution operating upon biochemical DNA-encoded genes, but also in terms of evolution being realised on the plane of more abstract information-carrying replicators which Dawkins (1976) named « memes ».

The basic definition of the classical memetic theory is: « Meme is a replicator which replicates from brain to brain by means of imitation» (Blackmore, 2000). These replicators are somehow represented in the host brain as some kind of « cognitive structure » and if ever externalised by the host organism – no matter whether in form a word, song, behavioral schema or an artefact – they can get copied into other host organism endowed with the device to integrate such structures[2]. Similary to genes which often network themselves into mutually supporting auto-catalytic networks (Kauffman, 1995) , memes can also form more complex memetic complexes, « memplexes », in order to augment the probability of their survival in time. Memes can thus do informational crossovers with one another (syncretic religions, new recepts from old ingredients or DJ mixes can be nice examples of such memetic crossover) or they can simply mutate, either because of the noise present during the imitation (replication) process, or due to other decay factors related to the ways how active memes are ultimately stored in brains or other information processing devices.

Memes coalesce in auto-catalytic memplexes

Memetic theory postulates that the cumulative evolutionary process applied upon such reproduction of information-carrying stuctures BETWEEN minds shall ultimately lead to emergence of such complex phaenomena as culture, religion or language. It can be thus considered to be mainly the theory of inter-mental reproduction of information. In complementarity with such a view, this dissertation claims the existence of reproduction of information WITHIN the individual mind. Thus, the theory hereby presented can be labeled as a theory of intra-mental memetics. END MEMETICS 8.3

Of inter- and intra-mental memetics

2 In neurobiological terms, the faculty to imitate and hence to integrate memes from external environment is often associated to «mirror neurons»

8.4 EVOLUTIONARY EPISTEMOLOGY

Epistemology is a philosophical discipline concerned with the source, nature, scope , existence and divesity of forms of knowledge. Evolutionary epistemology (EE) is a paradigm which aims to explain these by applying the evolutionary framework. But under one EE label, at least two distinct topics are, in fact, addressed :

1. EE_1 which aims to explain the biological evolution of cognitive and mental faculties in humans and animals

2. EE_2 postulates that knowledge itself evolves by selection and variation

EE_1can be thus considered as sub-discipline of EPSection 8.2 and as such, is subject to EP-directed criticism. EE_2, however, is closer to memetics since it postulates the existence of a second replicator, i.e. of an information-carrying structure which is not materially encoded by a DNA molecule.

The distinction between EE_1 and EE_2 can also be characterised in terms of « phylogeny » and « ontogeny ». Given the definition of phylogeny as

8.4.1 PHYLOGENY

Process which shapes the form of species.

END PHYLOGENY 8.4.1

and contrasting it to ontogeny defined as

8.4.2 ONTOGENY

Processus which shapes the form of an individual.

END ONTOGENY 8.4.2

we find it important to reiterate that while EE_1 is more concerned with knowledge as a result of phylogenetic moulding of DNA, EE_2 implies the moulding of non-DNA replicators in both phylogeny and ontogeny. Thus, the notion of EE_2 can be subsequently analysed into two sub-notions :

- $EE_{2\text{-}1}$ Knowledge can emerge by variation&selection of ideas shared by a group of mutually interacting individuals (Popper, 1972)

- $EE_{2\text{-}2}$ Knowledge can emerge by variation&selection of cognitive structures within one individuum

This distinction is homologous to distinction between inter- and intra- mental memetics, as discussed in Section 8.3.

It is worth noting that while a so-called recapitulation theory stating that « ontogeny recapitulates phylogeny » (Haeckel, 1879) is considered to be discredited by many biologists and embryologists ; it is still held as valid by many reseachers in human and cognitive sciences. In anthropology, for example, some scientists observe a « strong parallelism between cognitive development of a child and ... stages suggested in the archeological record» (Foiter, 2002). Also in relation to pedagogy it was observed that « education is a repetition of civilization in little» (Spencer, 1894).

8.4.3 INDIVIDUAL CREATIVITY

In fact, the evolutionary epistemology was born with the tentative of D.T. Campbell to explain both creative thinking and scientific discovery in terms of « blind variation and selective retention» (Campbell, 1960) of thoughts.

Creativity as intrapsychic evolution

Departing from introspective works of mathematician Henri Poincare who stated « To create consists precisely in not making useless combinations and in making those which are useful and which are only a small minority. Invention is discernment, choice...Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart...What is the cause that,among the thousand products of our unconscious activity, some are called to pass the threshold, while others remain below?» (Poincaré, 1908), Campbell suggests that what we call creative thought can be described as a Darwinian process whereby the previously acquired knowledge blindly varies in unconscious mind of the creative thinker and that only some such structures are subsequently selectively retained.

The theory which interprets the creative process as an evolutionary one has been subsequently developped by Dan Simonton who answers his rhetorical question "How do human beings create variations?" with a UD-constitent answer: « One perfectly good Darwinian explanation would be that the variations themselves arise from a cognitive variation-selection process that occurs within the individual brain.» (Simonton, 1999).

END INDIVIDUAL CREATIVITY 8.4.3

8.4.4 GENETIC EPISTEMOLOGY

« The fundamental hypothesis of genetic epistemology is that there is a parallelism between the progress made in ... organization of knowledge and the corresponding formative psychological processes. Well, now, if that is our hypothesis, what will be our field of study? Of course the most fruitful, most obvious field of study would be reconstituting human history: the history of human thinking in prehistoric man. Unfortunately, we are not very well informed about the psychology of Neanderthal man or about the psychology of Homo siniensis of Teilhard de Chardin. Since this field of biogenesis is not available to us, we shall do as biologists do and turn to ontogenesis. Nothing could be more accessible to study than the ontogenesis of these notions. There are children all around us.» (Piaget, 1974)

When understood only superficially, Piaget's developmental theory of knowledge, which he himself called Genetic Epistemology (GE) may seem to be utterly non-Darwinian. Its concern is not the phylogeny of human species, it is not even concerned with biochemical genes. In fact, during practically all his fecond life-lasting research, Piaget had focused solely on the study of ontogeny of diverse cognitive faculties in human children.

Thus, Piaget uses the term « genetic » to refer to a more general notion of « heredity » defined as structure's tendency to guard its identity through time. These structures, which he called « schemas » can be defined as « a basic set of experiences and knowledge that has been gained through personal experiences that define how things should be and act in the person's environment. As the child interacts with their world and acquires more experiences these schemes are modified to make sense, or used to make sense of the new experience.» (Bee and Boyd, 2000)

There are basicly two ways how such schemes can be modified. Either they « assimilate » data from external environment. Or, if ever such assimilation is not possible because it is simply not possible that child's cognitive system matches the perceived external datum with the internal pre-existing category, the process of « accomodation » takes place which transforms the internal category to match the external datum.

Ultimately, the set of schemes gets so out-dated or so altered by past modifications that they are not useful anymore. Whenever such «equilibriation » occur, old set of schemas is rejected, the child tends to « start fresh with a more up-to-date model» (Bee and Boyd, 2000), thus attaining new substage or stage of its development. In the Piagetian system – which is based on very precise yet exhaustive observations of dozens of children including his own – the order of stages is fixed and it is very difficult, or even fully impossible, for evolving psyche to attain pre-operational stage 2 or concrete operational stage 3 if it had not even mastered all that is to master during the sensorimotor stage 1.

1. sensorimotor stage - repetitive but playful manipulation of objects without goal

2. egocentric stage - imitation of behavioral schemas of others without understanding of why it is done
3. cooperative stage - coordination of one's activity with one's environment
4. autonomous stage - understanding of procedures which allow to change rules governing one's environment

Given that that the GE paradigm involves

- heredity – schemes tend to keep their identity in time
- variation – schemes are altered by the environment-driven assimilation or accomodation[3]
- selective pressures – only those schemas which are most well adapted to environment and/or form most functionally fit complexes with other schemas shall pass through the equilibriation milestone

it can be briefly stated that Piaget's GE could be aligned with ET and UD. And what more, it may be the case that notion of Piagetian stages is consisted with the notion of attractor or locally optimal states whose emergence is, according to complex system theory (Kauffman, 1995; Flake, 1998) inevitable in a system as complex as child's brain, mind and psyche definitely is.

END GENETIC EPISTEMOLOGY 8.4.4

END EVOLUTIONARY EPISTEMOLOGY 8.4

8.5 EVOLUTIONARY LINGUISTICS

Analogically to Evolutionary Epistemology, objects of interest of EL subdivide it at least into two branches:

- EL_1 the study of origin and development of faculties related to comprehension and production of linguistic signal by homo sapiens sapiens and its ancestors
- EL_2 the study of historical development of diverse languages

Distinction between EL_1 *and* EL_2

EL_1 can be thus considered to be closely related to Evolutionary

3 Note that in terms of EC, one can relate the Piagetian notion of assimilation to an operator of local variation which attracts the cognitive system to locally optimal agreement with its environment, while accomodation suggests an interpretation in term of more global variation operators (like cross-over), potentially allowing the CS to adapt to its physical and social environments in a more globally optimal way.

Psychology (Section 8.2) and discuss phylogenetic evolutionary phenomena taking place during hundreds of thousands of years while EL_2 can be said to take place in the historical time (order of ten thousand years and less) and is thus closely related to disciplines like anthropology, culturology, comparative grammar and memetics. In simple terms, EL_2 is dedicated to study of linguistic ethnogeny.

8.5.1 ETHNOGENY (DEF)

Processus which shapes the form of a human community.

END ETHNOGENY 8.5

EL_2's central tenet that "language changes in time" is far from being new. Socrates have believed that « ...the primeval words (πρώτα ονόματα) have already been buried by people who wanted to embellish them by adding and removing letters to make them sound better, and disfiguring them totally, either for aesthetic considerations or as a result of the passage of time...» (Plato, 80BC) and the best of Plato's students was well aware that change can be expressed in terms of

Language changes

- insertion
- deletion
- transposition
- substitution

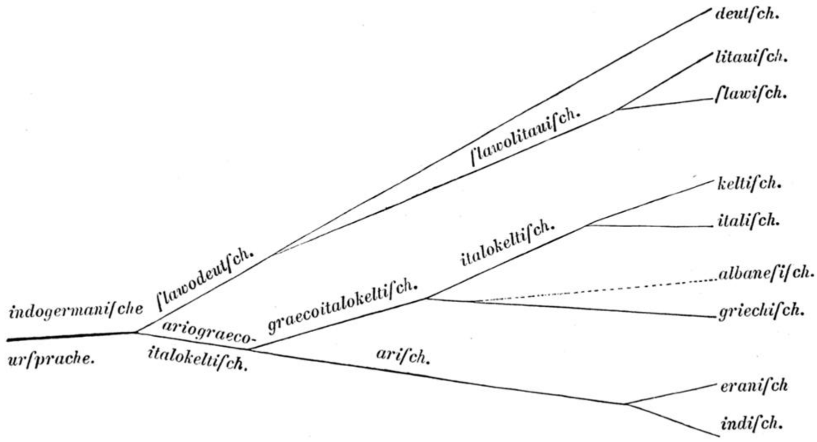
Aristotle (42BC). Ancient syntacticians like Apollonius Dyscolus could subsequently apply such notions to describe particular linguistic phenomena (Householder, 1981).

It was, however only centuries later when men of science had realized that language change is far from being a linear degression of the primordial ideal, as the ancients have mostly believed. On the contrary: sir William Jones's discovery that sanskrit is similar to Greek, Celtic, Gothic and Latin languages and that they all « sprung from some common source, which perhaps no longer exists» (Jones, 1788). Subsequent realization that these similarities make it possible to cluster languages into hierarchical taxonomies combined with the trivial fact that languages exchange their internal contents (e.g. word-borrowing), all this has led to evermore stronger belief that *languages can be studied as living entities*. Darwin himself was well aware of the parallelism between biology and linguistics:

Language evolves

« The formation of different languages and of distinct species, and the proof that both have developer through a gradual process, are curiously parallel...We find in distinct languags striking homologies due to community of descent, and analogies due to a similar process of formation.» (Darwin, 1859)

Figure 5: Schleicher's Stammbaum of family of Indo-European languages. Reproced from Schleicher (1873).



Practically in the same as Darwin was preparing his opus which was to change biology forever, was, on the linguistic side, the existence of such parallelism articulated by Schleicher (1873) in his "tree" *Stammbaum* theory of Indo-European languages .

Language tree theory

By publishing his theory, Schleicher had in fact triggered a completely new form of evolution - that is, evolution of linguistic theories. In a dozen years that followed was the influx of articles related to Stammbaumtheorie so high that *Société linguistique de Paris* had decided, in 1866, to refuse any articles on the subject. Which is somewhat a pity because many theories which emerged during that period, for example "the wave theory" (Schmidt, 1872) taking into consideration not only temporal but also spatial (i.e. geographic) aspects of language spread, were indeed preminiscent to diffusion models which became prominent in biology only a century later.

The fossile absence problem

One of the reasons for Societe's "ban" was the fact that languages, contrary to "biological species" do not left fossile traces after them and therefore any endeavour to understand their distant past or even origin is only speculative and inconsistent with empiric method of science. EL simply does not go well with the principle of scientific parsimony, the omnipresent Occam's razor. Notwithstanding this critique which stays, we believe, valid today as ever[4], allowed the advent of computers to EL_2 to catch the second breath.

Glottochronology

An often criticized but nonetheless very important step in making EL computer-positive was the introduction of "lexicostatistical" and *glottochronological* methodology originally based on cognate distance matrices (Swadesh, 1952). These numeric matrices, whose elements M_{ij} were denoting the number of cognates - i.e. the number of similarly sounding words having the same meaning - subsequently al-

4 C.f. the footnote in Section 8.2 or citation of Piaget in Section 8.4.4 for other reformulations of the same critique.

lowed to computationally "discover" and (fals | ver)ify hypothesie concerning kinship of existing or past languages. An article of Atkinson and Gray (2005), from which we reproduce a Table 1 a Table describing parallelism between biological and linguistic evolution, offers a satisfactory introduction to some EL_2's state-of-the-art computational models some of which pretend to unveil knowledge about ancestry of languages as far as the end of last ice age (Pagel et al., 2013).

BIOLOGICAL EVOLUTION	LINGUISTIC EVOLUTION
Discrete characters	Lexicon, syntax and phonology
Homologies	Cognates
Mutation	Innovation
Horizontal gene transfer	Borrowing
Hybrid plants	Creole languages

Table 1: Conceptual parallels between biological and linguistic evolution. Table partially reproduced from Atkinson and Gray (2005).

Section 8.7.5 shall discuss a so-called "Evolutionary Language Game" computational model. Since ELG addresses - and some may say that also answers - the question "How may a coordinated system of sound-meaning mappings evolve *ex nihilo* in a community of mutually interacting agents?", it can be posited at the very border between EL_1 and EL_2. According to Pinker, who is one of the most famous proponents of so-called "nativist" theory in developmental psycholinguistics (c.f. item 10.2) models like ELG « suggest ways of connecting the evolution of language to other topic of in human evolution, allowing each to constrain the others» (Pinker, 2000).

But there is another question related to evolution of language which has not yet been sufficiently resolved by ELG nor any other EL theory[5]. That is: "Why are languages subject to some types of changes and not to others?". Why indeed is history of languages so full of insertions (e.g. "osm" in czech and "osem" in slovak), deletions (e.g. "mravenec" in czech and "mravec" in slovak), substitutions (e.g. all instances of what is a diftong "ie" in slovak are pronounced in czech as a long vowel í) and metathetic transpositions (e.g. "hmla" in slovak and "mlha" in czech)?

Why are some changes more fit than others?

Our answer to this question is as follows : because the changes observable in ethnogeny of diverse languages, dialects and accents are, at their origin, triggered by "variation operators" inherent to every fundamental unit of any linguistic community which is, of course,

Cognitive constraints in language evolution

5 We set aside a so-called neo-grammarian school of historical and comparativist philology who believed that language change can be described in terms of sequences of universally applicable "laws which suffer no exceptions". We set them aside because we are strongly persuade that evolution not only does suffer "exceptions" but, in fact, endorses them in order to be fully operational.

an individual human mind. Stated more simply: the reasons why language forms develop in the way they develop are in great extent *cognitive*.

Part iii shall present somewhat more concrete an evidence of activity of such operators of intramental variation which potentially influence the process of language production in human children .

END EVOLUTIONARY LINGUISTICS 8.5

8.6 NEURAL AND MENTAL DARWINISM

It was already an evolutionary biologist John Maynard-Smith who have remarked that « there is a similarity between the dynamics of genetic selection and the operant conditioning paradigm of Skinner» (Maynard Smith, 1986)[6]. But it was only the book *Neural Darwinism: The Theory of Neural Group Selection* of Nobel-prize winner Edelman (1987) who had, as first in history of science, described in a fine-grained detail how a process similar to evolution could be potentially instantiated within the human brain.

Basic tenets of neural darwinism

Stated in one sentence, Edelman's theory postulates that « complex adaptations in the brain arise through process similar to natural selection» (Fernando et al., 2012). Stated in a more fine-grained detail, the theory shows how epigenetically influenced interactions of "cell adhesion molecules" and "substrate adhesion molecules" can lead to generation of so-called primary repertoire. Synapses within diverse groups of this repertoire are subsequently, during postnatal ontogenesis, "differentially amplified" into a secondary repertoire by a process which is, according to Edelman, functionally equivalent to the process of selection as known in evolutionary theory. Edelman also believes that well-known processes like cell proliferation, cell migration, cell death, neurite branching or synaptic pruning are potentially also governed by analogic selective processes.

From neural to mental

It is not possible for us to explain Edelman's *tour de fource* in the limited scope of this section and it would be, in fact, an act of scientific dishonesty to do so since as computational linguists, we do not feel competent to express any definite statement about truth or falsity in such expert domain as neurology definitely is. But we nonetheless consider as important to emphasize that Edelman is definitely not alone in his view of things. Thus, for example, important authorities of continental neurological tradition were not afraid to state that « the thesis we wish to defend...[is] that the production and storage of mental representations, including their chaining into meaningful propositions and the development of reasoning, can also be intepreted, by

6 Skinner's behaviorist theory of verbal behaviour will be more closely discussed in 9.4.1

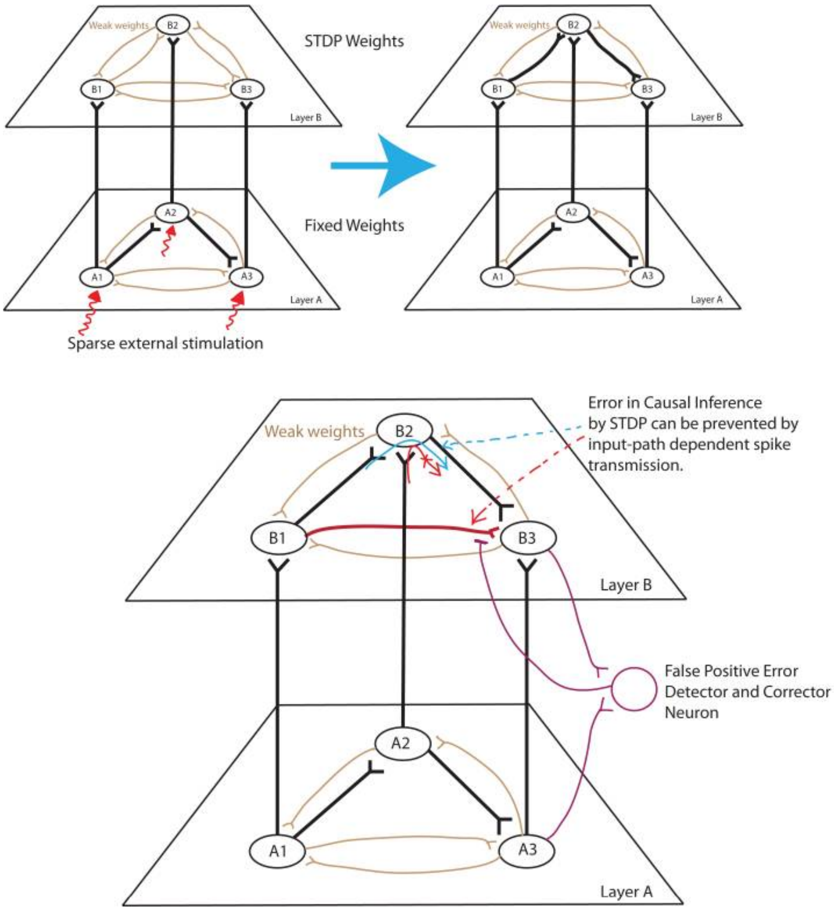


Figure 6: Possible mechanism of replication of patterns of synaptic connections between neuronal groups. Reproduced from Fernando et al. (2012).

analogy, in variation-selection (Darwinian) terms within psychological time-scales.» (Dehaene and Changeux, 1989)

It has to be noted, however both Edelman's "neural" and Dehaene's and Changeux's "mental" Darwinism describe processes fundamentally based on variation and selection, but not on replication, of information-encoding neural groups. It is a well-known fact that neural cells do not reproduce and the possibility that the reproduction of neurons would yield a material basis of existence of intracerebral replicators is thus a priori excluded. As is pointed by (Fernando et al., 2012) this fact in itself, however, does not mean that mental or neural darwinism are not evolutionary. They are evolutionary because one can postulate a sort of evolution for any system whose global development is governed by famous Price's theorem (Price et al., 1970) which is - so the authors argue - also the case for development of neuronal group structures.

Variation and selection of neuronal groups

Same authors also suggest possible process of replication of information between neuronal groups . This process - fundamentamen-

Replication among neuronal groups

tally based upon a well-known form of Hebbian-learning[7] called "spike-timing dependent plasticity" (STDP) and the existence of a neural "topographic" map between the original *replicans* (circuit A) and the following *replicandum* (circuit B) - can be described as follows: « If a neuronal circuit exists in layer A and is externally stimulated to make its neurons spike, then due to a topographic map from layer A to layer B, neurons in layer B will experience similar spike pattern statistics as in layer A. If there is STDP in layer B between weakly connected neurons then this layer becomes a kind of causal inference machine that observes the spike input from layer A and tries to produce a circuit with the same connectivity, or at least that is capable of generating the same pattern of correlations.» (Fernando et al., 2012). Whole process is visualised on Figure 6.

Neural darwinism still speculative

While we strongly believe that such a mechanism does indeed operate in human cortex, we reiterate what was already stated in 2.11 : under current state of knowledge is existence of neural replicators not indisputably demonstrated and stays speculative. But in regards to overall objectives of this dissertation does this speculative nature of intracerebral replicators NOT pose any hindrance. This being so because our aim is to apply use evolutionary theory to explain linguistic phenomena. And linguistic phenomena are principially intangible, mental, high-order phenomena which are potentially irreducible to tangible and physical phenomena labelled as "neural".

Bridging the explanatory gap

On the other hand, it may be the case that a sort of theory intramental evolution allow us to bridge the "explanatory gap" between tangible and intangible, neural and mental. Thus, for example, whenever we shall emit hypothesis like "canonical babbling is a sort of replicatory process" (9.2.2), we hereby tacitly imply that neural mechanisms, as the one presented on Figure 6, are to be sought-for in Broca's area of one year old infants.

END NEURAL DARWINISM 8.6

8.7 EVOLUTIONARY COMPUTATION

Universal Algorithm

Evolution can be thought of as a universal, generic algorithm. But our growing knowledge of evolution serves not only descriptive and explanatory purposes. It is becoming normative. Thus, not only can « evolutionary theory » serve us to explain diverse phenomena around us, it can be also exploited for finding solutions to diverse problems. Many researchers in informatics have already realized that diverse of "evolutionary recepts" offer useful heuristics making it possible to discover (quasi)-optimal *ways out* of wide range of concrete practical isses.

7 Principle of Hebbian learning shall be more closely discussed in Section 9.4.1.

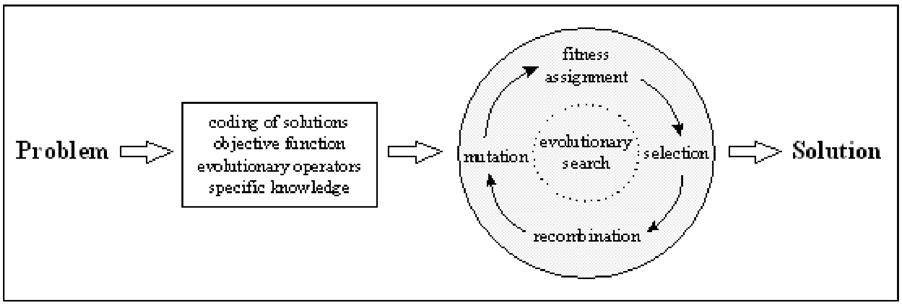


Figure 7: Basic genetic algorithm schema. Reproduced from Pohlheim (1996)

Evolutionary computing (3.1) approaches differ from classical optimization methods in following aspects :

- using a population of potential solutions in their search
- using probabilistic, rather than deterministic, transition rules »
- using «fitness» instead of function derivatives Kennedy et al. (2001)

First computational models which have the above-mentioned attributes were named « evolutionary strategies » by Rechenberg (1971), «genetic algorithms» by Holland (1975) and « evolutionary programming » by Fogel et al. (1966). These paradigms, along with the «genetic programming » paradigm later introduced by Koza (1992) constitute the most important sub-branches of «evolutionary computation» Sekaj (2005) branch of computer and informatic science.

8.7.1 genetic algorithms

Basic principle of « genetic algorithms » is illustrated on Figure 7. GAs iteratively produce populations of data stuctures. Each individual data structure is a possible solution, population of every generation is thus a set of diverse solutions. Every individual solution is encoded as a vector of values (also called « chromosome » or « genome ») which can either vary or be copied verbatim from one to generation to the other. Designer choice related to the way how the problem solutions are encoded in chromosomal vectors, e.g. the type (Boolean ? Integer ? Float ? Set?) of different elements of the vector is also a crucial one and can often determine whether the algorithm shall succeed or fail.

In every generation – i.e. in every iteration of the algorithmic cycle represented by the circle on Figure 7 - all N individuals in the population are evaluated by the fitness function. Every individual thus obtains the « fitness » value, which subsequently governs the « selection » procedure choosing a subset of individuals from the current generation as those, whose genetic information shall reproduce into next generations. More on fitness functions in 8.7.1.

Another important design decision which every programmer of GAs have to do, is to choose the selection operator. An operator which is widely used, and which we shall also implement in all future EC simulations (c.f. 10.4.7, **??** & volume 2) is the «fitness proportionate selection». This operator, also called « roulette wheel operator » normalizes the fitness f_i of individual i into the probability p_i of its survival by means of a formula :

$$p_i] = f_i / \sum_{j=1}^{N} f_j$$

where N is the number of individuals in the population. Once these probabilities are calculated to different individuals, one can use them to guide the process of selection of individuals which shall be reproduced into the next generation. Minimal PERL source code for such fitness proportional selection operator is:

Fitness Proportional Selection (SRC)

```
sub fitness_to_proba {
        my @weights = @_;
        my @dist    = ();
        my $total   = 0;
        local $_;
        foreach (@weights) {
            $total += $_;
        }
        for my $weight (@weights) {
            push @dist, $weight/$total;
        }
        return @dist;
}
sub weighted_rand {
        my @dist = @_;
        while (1) {
            my $rand = rand;
            my $i=0;
            for my $w (@dist) {
                        return $i if ($rand -= $w) < 0;
                        $i++;
            }
        }
}
```

END FITNESS PROPORTIONAL SELECTION (SRC) 8.7.1.0

Another widely used selection operator is a so-called tournament selection based on repeated selection of the best individual of population's randomly chosen subset. The tournament selection operator offers multiple advantages: for example, by tuning the tournament size parameter one can easily adjust the selection pressures favorizing or defavorizing fit candidates. And it can also be used in parallel computation scenarios.

Tournament selection

Once the « most fit » candidates are selected by the selection operator, they are subsequently mutually recombined by means of « crossover » operators and/or modified by means of « mutation » operators. Many different types of selection, mutation and crossover operators exist. For the purpose of this work let's just note that the probabilities of occurrence of mutation or crossover have to be fairly low, otherwise no fitness-increasing information could be transferred among generations and whole system will tend to present non-converging chaotic behaviour (Nowak et al., 1999).

Values for variation operators

Another useful strategy, which guarantees that maximal fitness shall either increase or at least stay constant, is called elitism. In order to implement the strategy, one simply guards one (or more) individual(s) with highest fitness unchanged for next generation, thus protecting « the best ones » from variations which would, most probably, decrease rather than increase the fitness[8].

Elitism

Yet another widely used approach reinforces the selection pressure by removal of the weakest individuals. Both elitist « survival of the fittest » and the contrary « removal of the weakest » are often combined within the sequence of instructions which, alltogether, form a genetic algorithm.

The selection of the most fit individuals from the old generation, their subsequent replication and/or recombination and diversification yields a new generation. Because individuals with lower fitness have been either completely or at least partially discarded by the selection process, one can expect that the overall fitness of new generation shall be higher than the fitness of the old generation. With little bit of luck, one can also hope that the most fit individuals of the new generation shall be little bit more fitter than the most fit individuals discovered in the new generation – this can happen if ever a « benign » mutation have occured, i.e. a modification which had moved the individual from the lower point on the « fitness landscape » to somewhat higher state.

Drift towards higher fitness

END GENETIC ALGORITHMS 8.7.1

8 Note that in nature, elitism is often but not always the case. For it can happen that, due to stochastic factors, the most fit individuals die before they succeed to reproduce the information they encode. But, in such a case, are such individuals truly "the most fit"?

Fitness functions and fitness landscapes

Of functional core

The core component of every genetic algorithm is the objective «fitness function» able to attribute a cardinal value or ordinal rank to any individum in the population of potential solutions. In other terms, the fitness function yields the criterium according to which one candidate individum is evaluated as «more fit» a solution, in regards to the problem under study, than other potential solutions present in the population.

What is fitness function?

The choice of good fitness function determines, more than anything else, the success or failure of GA as a means to find the solution for the problem at hand. Ideally, the fitness function is a mathematical representation of the very essence of the problem which is to be solved. For purely mathematical problems, the choice of the fitness function is straightforward - fitness function is simply the function whose global optimum one wants to find. Also in many practical implementations - notably those of optimalization of physical components - the fitness function is also often evident: one can deduce it from well-established physical laws.

Fitness function as a design choice

But fitness functions for other problems are far from being certain. The "first language learning" which we aim to address in this dissertation belongs among such problem since it is not trivial to answer the question: "which model of language is better (i.e. more fit): X, Y or Z?". Such an answer is strongly determined by the theoretical point of view one adopts: an engineer prefering a sociopragmatic (9.4.4) or constructivist (9.4.3) theory of language acquisition, a model of language competence of 12-month old baby which generates utterances like "tato tek tete" would be considered to be more "fit" than model generating utterances like "father, had your colorless green ideas slept furiously?" (Chomsky, 1957). Rather contrary should be the case for an engineer who would decide to formalize his fitness function on the grounds of nativist (10.2) theories of language acquisition.

Fitness landscapes

The notion of fitness landscape, first introduced by Wright (1932) is a metaphor useful for understanding, discussion and comparison of diverse fitness functions. The landscape is depicted as a mountain range with peaks of varying height. The height at any point on the landscape corresponds to its fitness value; i.e. the higher the point, the greater the fitness of an individual represented by the given point of the landscape[9] In such a representation, the evolution of the organism to more and more « fit » forms can be depicted as a movement uphill, towards the most closest peak (i.e. local optimum) or towards the highest peak of the whole landscape (i.e. global optimum). Figure 8

9 Note that to find an optimal solution of the problem with N variables, one has to look for it in the N dimensional search space. This multi-dimensionality is what makes the search so difficult since the number of possible solutions grows exponentially with the number of dimensions (i.e. variables of the problem).

illustrates a fitness landscape of a very simple organism with only one gene (whose potential values are encoded by illustration's X axis).

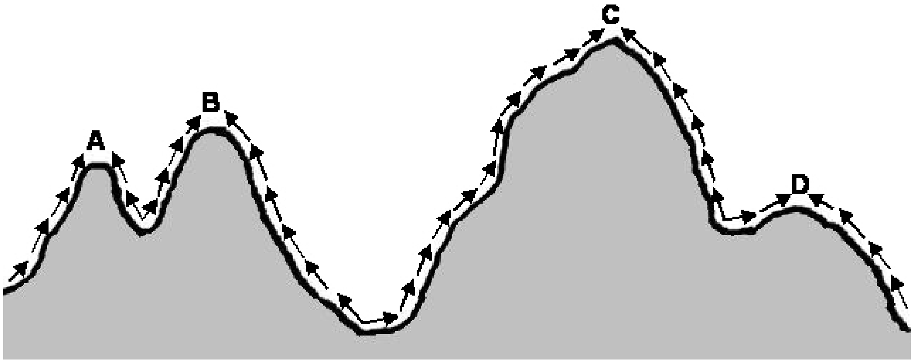


Figure 8: Possible fitness landscape for a problem with only one variable. Horizontal axis represents gene's value, vertical axis represents fitness.

Multidimensional hiking

Every arrow on the figure represents one possible individual. Its length represents the variation which can be brought in by the mutation operator. The fact that individuals always tend to move « upwards » indicates that selection pressures are involved. It has to be added that without the implementation of the crossover operator, the globally optimal state (encoded by point C) could not be attained for individuals who haven't originated at the slopes of C. Only some sort of crossover operator could ensure that individuals who attained the local optima (encoded by peaks A, B, D) could be mutually recombined (for example B with D) in a way that shall allow them to leave the locally stable states and approach the globally optimal C. The fact that genetic algorithms, thanks to « crossover » operators, can combine two individuals from diverse sectors of the fitness landscape, allow them to find solutions to problems where heuristics based on « gradient descent » would normally fail.

Ruggedness of fitness landscapes

An important property of fitness-landscape is its "ruggedness". Some fitness functions can yield landscapes as flat as Pannonian Plane: the algorithm will need a long time to find there a hill if ever a hill there is. Other may yield landscapes as rugged as mountains of northwest Vietnam: nothing is certain on such landscapes where even the slightest mutation can produce huge decrease or increase of fitness. Ideal landscapes are those which are rugged but not too much: as on the slopes of Himalaya, a steady progress towards some locally optimal -not neccessarily the highest but sufficiently high- vantage point can be assured.

C.f. NK Theory introduced in Kauffman (1995) for further discussion of landscape ruggedness and ways how it can be potentially tuned.

END FITNESS FUNCTIONS AND LANDSCAPES 8.7.1.0

Canonical Genetic Algorithms

Canonical genetic algorithm (CGA) is a genetic algorithm applied on populations (n-tuples) of binary strings (individuals) of length l. Each among l bits is considered to be a gene and each string of such genes is considered to be a potential solution to the problem which is to be solved. Given that the initial population is randomly generated, the CGA proceeds as follows: In CGAs, fitness proportionate selec-

Listing 1: Canonical Genetic Algorithm

```
initialize the population
determine the fitness of each individual
perform selection
repeat
	perform crossover
	perform mutation
	determine the fitness of each individual
	perform selection
until some stopping criterion applies
```

Operators in CGA

tion (8.7.1) is used as the selection operator. Mutation operates independently on each gene of each individual and consists of stochastic bit flipping of current gene's value to its opposite. A "one-point crossover" (4) is most commonly used in CGAs, which consist of randomly chosing a section locus of the chromosome, dissecting two selected parent individuals A and B along the section locus and creating two children individuals C and D as a concatenation of sections previously encoded in two distinct parent organisms, i.e. $C = A_1B_2$ and $D = B_1A_2$.

CGA convergence and elitism

CGAs being thus defined in (Holland, 1975; Goldberg, 1990), it has been demonstrated by (Rudolph, 1994) that such pure CGAs are unable to converge to global optimum of the problem they tend to maximize. This is so because even if CGA would be able to discover the optimum, the unceasing activity of mutation operators would force the system to depart from such an ideal state. On the other hand, if ever one implements the elitist trick of keeping the most fit individual, such convergence is assured. Thus, Rudolph's theoretical « analysis reveals that the convergence to the global optimum is not an inherent property of the CGA but rather is a consequence of the algorithmic trick of keeping track of the best solution found over time.» (Rudolph, 1994) It is principially because of CGA's

1. theoretical ability to converge to global optimum
2. simplicity and architectural elegance

that our method of Evolutionary Localization of Semantic Attractors (ELSA, 10.4.7) is, *in essentia*, nothing else than a CGA endowed with elitist strategy.

END CANONIC GENETIC ALGORITHMS 8.7.1.0

Parallel Genetic Algorithms

Parallel Genetic Algorithms (PGAs) add another level of complexity to traditional GAs. In PGAs is the global population of solutions divided into multi sub-populations which, most of the time, evolve independently from each other. One can understand such sub-populations as different societies or species evolving on isolated islands. Only during so-called "migratory periods" do the sub-populations communicate with each other: most often by means of "sending" the most fit individual to another receptor sub-population. Grid (A,B), hierarchical (C), ring (D) and multi-hierarchical (E,F) architectures of such interinsular migratory relations are depicted on Figure 9.

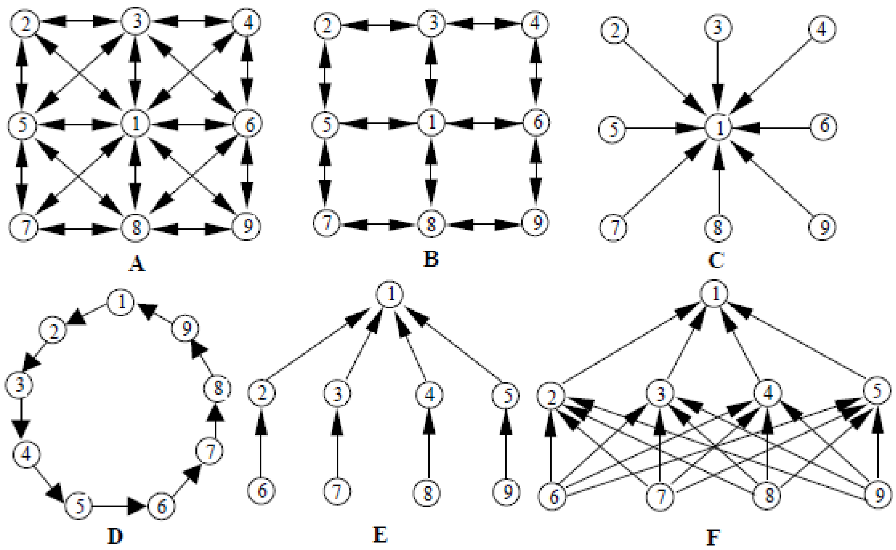


Figure 9: Different architectures of Parallel Genetic Algorithms. Reproduced from Sekaj (2004)

By introducing multiple independent populations, PGAs allow to put in equilibrium the selective pressure (i.e. preference of better individuals) and population diversity (i.e. gene dissimilarity). In traditional single-populated GAs, these two forces oppose each other: by increasing the selective pressure an engineer reduces the diversity and thus exposes himself to danger of converging "just" to a locally optimal state. On the other hand, by favorizing too much diversity, one can slow down significantly the convergence rate. To find the equilibrium between these two forces is indeed an art.

PGAs solve this tradeoff problem by allowing to increase selective pressures in one sub-population while augmenting the diversity of the other. The gain seems to be particularly significative in case of heterogenous PGAs whereby diverse sub-populations implement diverse search strategies. Another improvements in case of problems with "rugged" fitness landscapes can be attained by introducing "sub-population re-initialisation" into the process. That is, an exchange of population whose diversity is too low, for a completely new, randomly generated population. Such « re-initialisation is able to remove differences between homogenous and heterogenous PGA's or between different PGA architecture types respectively. However, all the presented PGA modifications can speed up the search process and prevent the search algorithm from a premature convergence» (Sekaj, 2004).

It seems that adding another level of complexity to GAs increases the probability of finding the globally optimal solution. It is true that even the traditional single-population GAs explore the search space in multiple directions, in PGAs, however, is such exploration qualitatively augmented. By their faculty to centralize the decentralized; by their ability to speeden the convergence to optimal solutions of diverse problems as well as by allowing for hiearchical[10] stacking of independent information-processing units, PGAs are reminiscent of so-called deep-learning methods principially based on hierarchical stacking of diverse connectionist networks. And what's more, by being partially localized and partially globally-integrative, PGAs can offer a possibly interesting means how to simulate certain functions of human brain (c.f. 2.8) which seems to dispose of analogic properties.

END PARALLEL GENETIC ALGORITHMS 8.7.1.0

8.7.2 EVOLUTIONARY PROGRAMMING & EVOLUTIONARY STRATEGIES

Evolutionary programming (E.Prog) and evolutionary strategies (E.Strat) are methods whose overall essence is very similar to GAs. There are, however, some subtle differences among the approaches.

In E.Prog, mutation is the principal and often the only variation operator. While recombination is rarely used, « operators are freely adapted to fit the problem at hand» (Kennedy et al., 2001). E.Prog algorithms often double the size of population by mixing children with parents and then halving the population by selection. Tournament selection operator is often used.

10 In study of Sekaj (2004) hierarchical architectures C, E and F seem to be the most successful in approaching the global solutions of two specfic mathematical functions.

Another difference is that while GAs were developped in order to optimize the numeric parameters of mathematical function under study – and variation thus directly modifies the genotype – in E.Prog, one mutates the genotype but evaluates the fitness according to phenotype. E.Prog is thus often used for construction and optimization of such structures like finite state automata (Fogel et al., 1966). A self-adaptation approach (Bentley, 1999) allowing for mutation of the parameters of the evolution itself – e.g. the mutation rate – is also frequently used.

Such an approach of « evolving the evolution » is also used in E.Strat which where discovered - in parallel but independetly with Holland's GAs – by Rechenberg (1971). The biggest difference between E.Prog and E.Strat is thus fact that E.Strat often recombines its individuals before mutating them. Popular and well-performing strategy thus seems to be :

1. Initialize the population
2. Perform recombination using P parents to form C children[11]
3. Perform mutation on all children
4. Evaluate children population and select P members from it.
5. If the termination criterion is not met, go to step 2 ; terminate otherwise.

Given that in certain simulations (c.f. **??**), we shall

1. encode solutions by means of non-numeric chromosomes
2. evaluate the fitness of individuals by means of additional « phenotypic algorithms »

we consider the works of Fogel & Rechenberg to be precursors of our approach.

END EVOLUTIONARY PROGRAMMING & STRATEGIES 8.7.2

8.7.3 GENETIC PROGRAMMING

Contrary to GAs, E.Prog and E.Strat which operate upon the chromosomes (vectors) of fixed length of numeric/boolean/character values, do individuals evolved by means of Genetic Programming (GP) encode programs of arbitrary length and complexity. In other terms, one may state that while above-mentioned EC methods look for the most optimal solution of a given problem, GP tends to produce a hierarchical tree structure encoding a sequence of instructions (i.e. a program)

11 Frequently used C/P ratio is 7

able to yield optimal solutions to a whole range of problems. Simply said : GP is simply a way how computer programs can automatically « discover » new and useful programs.

The most important thing to do in order to prepare a GP framework is to specify how shall be the resulting individuals (programs) encoded. Original choice of the founder of the discipline, John Koza, was to encode all individuals as trees of LISP S-expressions composed of sub-trees, which are, themselves, also LISP S-expressions. Within such arborescent S-expressions, the terminal (i.e. leave nodes where the branches end) nodes represent program's variables and constants while the non-terminal nodes (i.e. internal tree points) represent diverse functions contained in the function set (e.g. arithmetic functions like +, -, *, / ; mathematic functions like log, cos ; boolean functions like AND, OR, NOT ; conditional operators if/else etc.)

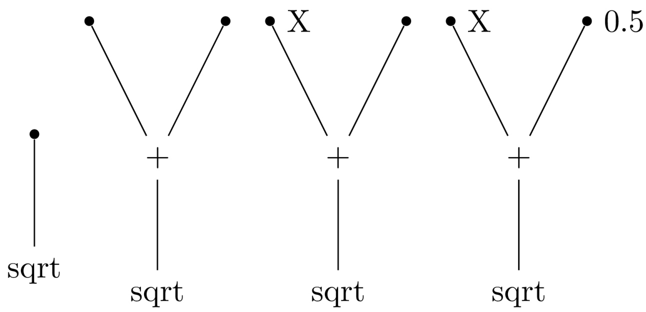


Figure 10: Sequence of steps constructing the program sqrt(x+5)

Figure 10 illustrates how, during the initial run of the algorithm, an individual – calculating, for example, the square root of X+5 – could be possibly randomly generated by implementing a following procedure :

1. « Root » of the program tree is randomly chosen from the function set, it is the function sqrt.

2. The function sqrt has only one argument (arity(sqrt)=1), therefore it will take only one input from the randomly determined functor + (addition)

3. Functor + takes two inputs (arity(+)=2), therefore the tree bifurcates into two lines in this node. It randomly choses, as the first argument, the constant 5 ; and the variable X as the second argument.

Note that in step 3, both arguments were chosen from the terminal set. If they would have been chosen from the function set, the tree could bifurcate further. In order to prevent such growth of trees ad infinitum, a limiting « maximal tree depth » parameter is more than often implemented in GP scenarios.

Once such a program has been generated, one can evaluate its fitness by confronting it with diverse input arguments and comparing its output with a golden standard. Such a random-program generation & evaluation is repeated for all N initial candidate programs, subsequently the most individuals are selected and varied. While GP's selection techniques can sometimes closely ressemble selection techniques as used in GAs, variation operators are often of essentially different nature. This is so, because GP's not individual genomes or their linear sequences can be mutated or crossed-over, but rather complex and hierarchical networks of expressions. In a case of cross-over, for example, one switches whole sub-tree encoded within one individual, for a sub-tree encoded within another one.

GP-based solutions cannot be expected to function correctly if they do not satisfy the theoretical properties of closure and sufficiency. In order to fulfill the closure condition, each function from the non-terminal set must be able to successfully operate both on output of any function in the non-terminal set and on any value obtainable by a member of the terminal set. Even behaviour of some simple operators thus has to be a priori adjusted (e.g. return 1 in case of division by zero) in order to assure correct functioning of the resulting program.

On the other hand, sufficiency property demands that the set of functors and terminals is sufficiently exhaustive. Otherwise the solution could not be found. One can not, for example, hope to discover equation for generating the Mandelbrot set if the initial set of terminals does not contain the notion of imaginary number, nor does the function set contain any other explicit or implicit reference to the notion of complex plane. Thus, while the closure constraint delimits the upper bound beyond which the discovery of the solution is not feasible, the sufficiency constraint delimits the lower bound of the minimal set of « initial components » which have to be defined a priori, so that discovery of the adequate program should be at least theoretically possible.

Other theoretical notions as well as diverse subtleties (special operators, methods how to distribute the initial population in the search space, fitness function proposals, domains of application, etc.) of practical implementation, are to be found in possibly the most important GP-concerning monography (Koza, 1992).

Grammatical evolution

Grammatical Evolution (Gr.Ev) is a variant of GP in a sense that it also use evolutionary computing in order to automatically generate computer programs. The most important difference between Gr.Ev and GP is that while GP operates directly upon phenotypic trees representing program's code itself (for example in form of LISP expressions), Gr.Ev uses the evolutionary machinery for the purpose of generating grammars, which would subsequently generate the program code.

In Formal Language Theory (c.f. also item 10.2), grammar is represented by the tuple {N, T, P, S} where N denotes the set of non-terminals, T the set of terminals, S is a symbol which is member of N and P denotes the set of production rules that substitute elements of N by elements of N, T or their combinations1. Consider a grammar G exhaustive enough to encode programs able to perform arbitrary number of operations of addition or subtraction of two variables: Such a grammar contains three non-terminals, non-terminal

Listing 2: An example of grammar G.

```
N = {expr, op, var}
T = { +, -, x, y}
S = expr
P = {   <op> ->  + | -
        <var> ->  x | y
        <expr> ->  <var> | <expr> <op> <expr> }
```

<op> which could be subtituted for either terminal + or terminal - ; non-terminal <var> which could be subtituted for either terminal x or terminal y ; and non-terminal <expr> which could be substituted for either a non-terminal <var>, or a sequence of non-terminals <expr> <op> <expr>.

x+x
x+y
y+x
y+y
x-x
x-y
y-y
y-x
x+x
x+x+x
x+x-x
x+x+y
x-x+y-y
y+y+x+x+y-x
etc.

The fact that in this last production, the non-terminal <expr> is present both on left and right side of the substitution rule gives this grammar a possibility to recursively generate infinite number of expressions. As may be seen in the listing to the left, even a very simple grammar -with only four terminal symbols and three non-terminal symbols to each of which are associated only two production rules- can theoretically -i.e. if given infinte amount of time for application of production rules- produce an infinite number of distinct individual programs able to perform basic arithmetic operations with two variables.

Generation of a given resulting expression is determined by the order of application of specific production rules, starting with non-terminal symbol S. Such a sequence of application of production rules is called derivation. For example, in order to derive the individual « x+x », one has to apply production rules in following order:

Listing 3: Production of expression x+x.

```
        S = <expr>
        <expr> ::= <expr> <op> <expr>
        <expr> ::= <var>        # <var> <op> <expr>
        <var> ::= x             # x <op> <expr>
        <op> :: = +             # x + <expr>
        <expr> :: = <var>       # x + <var>
        <var> :: = x            # x + x
```

while the individual « y-x » would be generated, if ever the starting symbol S should be expanded by a following sequence of production rules :

Listing 4: Production of expression y-x.

```
S = <expr>
        <expr> ::= <expr> <op> <expr>
        <expr> ::= <var>        # <var> <op> <expr>
        <var> ::= y             # y <op> <expr>
        <op> :: = -             # y - <expr>
        <expr> :: = <var>       # y - <var>
        <var> :: = x            # y - x
```

In Grammatical Evolution, it is this « order of application of production rules» which is encoded in the individual chromosome. In other terms, individual chromosomes encode when and where distinct production rules shall be applied. Figure 11 more closely illustrates, and puts into analogy with biological systems, the sequence of transformations which every binary chromosome undergoes during the process of unfolding into fully functional program.

As the Figure 11 indicates, the approach of Gr.Ev is quite intricate and involves multiple steps of information processing. Whole process starts with binary chromosome subsequently split into 8-bit codons which yield an integer specifying which production rule to use in a given moment of program's generation. On many different layers does the « generation » process, as implemented in Gr.Ev, introduce and implement very original ideas like:

- « Degenerate genetic code » - similary to « nature's choice » to encode one amino-acid by means of many different triplets, can one encode application of a unique production rule by more than one codon.
- «Wrapping » - under certain conditions can be whole genome « traversed » more than once during the process of phenotypic expression. Specific codon can be thus used more than once during the compilation of single individual.

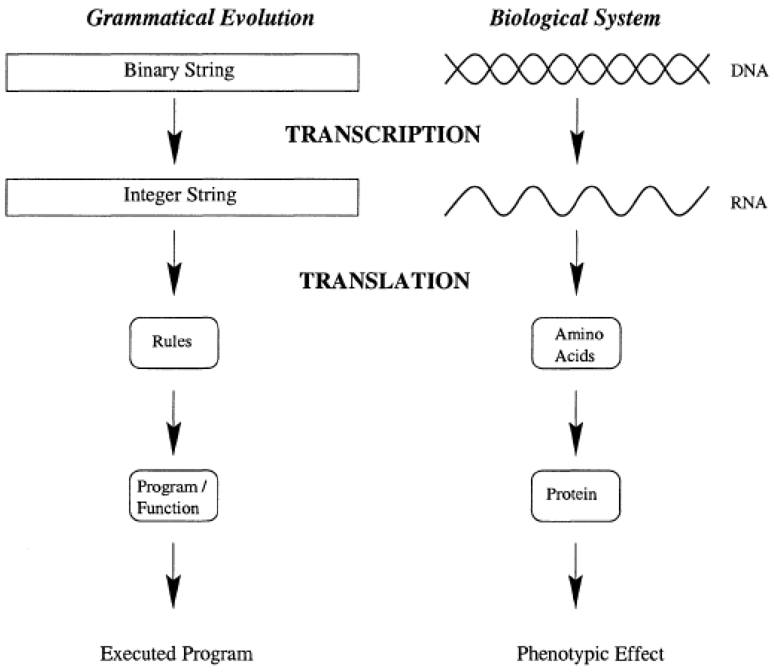


Figure 11: Sequence of transformations from genotype until phenotype in both Gr.Ev and Biological systems. Figure reproduced from O'Neil and Ryan (2003).

Rationale for usage of such « biologically inspired tricks » is more closely presented in the work of the founders of Grammatical Evolution field (O'Neil and Ryan, 2003) . They claim that the focus on genotype-phenotype distinction, especially in combination with implementation of « degenerate code » and « wrapping » notions, could result in compression of representation (& subsequent reduction of size of program search-space) and account for phenomenas like « neutral mutation », well-observed in biological systems, whereby a mutation occures in the genotype but does not have any effect upon the resulting phenotype. Another important advantage mentioned by O'Neill and Ryan is that Gr.Ev approach makes it very easy to generate programs in any arbitrary language. This is due to the versatility and generality of notion of « grammar ».

When compared with traditional GP technique, Gr.Ev was outperformed in a scenario when one had to find solutions to problem of symbolic regression. But in more case complex scenarios like « symbolic integration », « Santa Fe ant trial » or in scenario where one had to discover a most precise « caching algorithm », Gr.Ev significantly outperformed GP. Seminal work of O'Neil and Ryan (2003) presents also some other interesting examples of practical application of Gr.Ev, for example in the domain of financial market prediction.

It is worth underlining that while in many points (« grammar », « evolution ») does the work of O'Neilly and Ryan significantly overlap with ours, their aims significantly differ from our aim to interpret the process of language acquisition as an inherently evolutionary process. More concretely, while Gr.Ev tends to offer a very general toolbox to generate useful computer programs in arbitrary programming language and used for solving arbitrary problems, we confront the evolutionary computation machinery to shed some light upon diverse facets of one sole problem : that of «learning of first language».

Other important difference between the approach of Gr.Ev and the one we shall present in our thesis is that while in Gr.Ev, grammars are considered to be « generative devices », i.e. tools used for generation of programs, in our Thesis we shall use them as both « generative » and « parsing » devices. Another, even more fundamental difference is due to the fact that while « At the heart of GE lies the fact that genes are only used to determine which rule is applied when, not what the rules are» (O'Neil and Ryan, 2003) the evolutionary model of language-induction proposed in our Thesis shall aim to determine not only the order of application of the rules, but also the content of the rules themselves.

END GRAMMATICAL EVOLUTION 8.7.3.0

END GENETIC PROGRAMMING 8.7.3

8.7.4 TIERRA

Another example of how can one materialise evolutionary principles within an in silico framework is offered by Tierra, an artificial life simulation environment programmed between 1990-2001 by Thomas S. Ray and his colleagues. Since Ray is an ecologist, his objective was not to develop an EC-like model in order to find or optimalize solutions of a given problem, rather he aimed to create a system where artificially entities could spontaneously evolve, co-evolve and potentially create whole artificial ecosystems.

An artificial entity in Tierra's framework (Ray, 1992) is a program composed of sequence of instructions, chosen from instruction set containing 32 quite traditional assembler instructions somewhat tuned by the author so that their usage would facilitate « replication » of the code. Every artificial entity runs in its own « virtual CPU » but its code stays encoded in the « soup », i.e. piece of RAM which is potentially read-accessible to all other entities as well. Rare «cosmic ray » mutations flip the bits of « soup » from time to time, more variation is ensured by bit-flipping during the procedure whereby the entity replicates (i.e. copies) its code from the « mother cell » section of the soup to the « daughter cell » section.

Selection is in certain sense emulated by a so-called Reaper process which tends to stop the execution of programs which are either too old or contain too much flawed instructions. Other than that, there is nothing which ressemble the traditional notion of exogenously defined « fitness function ». For within Tierra, the survival (or death) of diverse species of programs is a direct consequence of species ability (or inability) to obtain access to limited ressources (CPU & memory).

Thus, after one seeds the initially empty soup with a manually constructed individual, containing 80-instructions allowing the individual to copy his code into the daughter cell of the memory, after the memory has been filled and the battle for ressources has started and once the mutation have generated sufficiently enough of variation, one can observe the emergence of dozens of new forms of replicable programs. Some of them being parasites, some of them being able to create algorithmic counter-mesures against parasites, one can literally observe an emergence of artificial yet living ecological system. It is therefore little surprising that Tierra could automatically evolve, among others, an individual containing just 22 instructions, capable of replication. That is, a replicator almost 4 times shorter than the replicator manually programmed by the conceptor of the system and injected into initial « soup ».

Currently the most famous descendant of Tierra is an AVIDA system (Ofria and Wilke, 2004). Contrary to Tierra, however, is every AVIDA's individual encapsulated within its own virtual CPU and memory space. Tierra's Darwinian metaphore1 of computer programs evolving by means of fighting for limited ressources is thus not so strictly followed.

END TIERRA 8.7.4

8.7.5 EVOLUTIONARY LANGUAGE GAME

Evolutionary Language Game (ELG) first proposed by Nowak et al. (1999) is a stunningly simple yet mathematically feasible stochastic model addressing the question « How could a coordinated system of meanings&sounds evolve in a group of mutually interacting agents ?».

In most simple terms, the model can be described as follows: Let's have a population of N agents. Each agent is described by an rxc associative matrix A. A's entry a_{ij} specifies how often an individual, in a role of a student, observed one or more other individuals (teachers) referring to object i by producing signal j.

Thus, from this associative matrix A, one can derive the active «speaker» matrix S by normalizing rows :

$$s_{ij} = \frac{a_{ij}}{\sum_{n=1}^{r} a_{in}}$$

while the «hearer» passive matrix H by normalization of A's columns:

$$h_{ij} = \frac{a_{ij}}{\sum_{n=1}^{c} a_{nj}}$$

The entries s_{ij} of the matrix S denote the probability that in P-representations of an agent-speaker, object i is associated with sound j. The entries $h_i j$ of the matrix H denote the probability with which, within C-representations[12] of the hearer, a sound j is associated with the object i.

Subsequently, we can imagine two individuals A and A', the first one having the language L (S,H), the other having the language L' (H', S'). The payoff related to communication of such two individuals is, within Nowak's model, calculated as follows:

$$F(A, A') = \sum_{i=1}^{r} \sum_{j=1}^{c} s_{ij} h'_{ji} = Tr(SH')$$

And the fitness of the individual A in regards to all other members of the population can be obtained as follows :

$$f(A) = \frac{1}{|P| - 1} \sum_{\substack{A' \in P \\ A \neq A'}} F(A, A')$$

After the fitness values are obtained for all population members, one can easily apply traditional evolutionary computing methods in order to direct the population toward more optimal states, i.e. states where individual matrices are mutually « aligned ». In Nowak's framework this alignment represents the situation when hearer and speaker mutually understand each other, i.e. speaker has encoded meaning M by sound S and hearer had subsequently decoded sound S as meaning M.

ELG beautifully illustrates how a mutually shared communication protocol can emerge from a population of randomly set sound-meaning associative matrices if there is some « mutual associative reinforcement » mechanism involved. This mechaninsm allows to transfer information from one individual to individual another. This is attained by creating a blank « student » matrix and then filling its elements, by means of stochastic « matrix sampling » procedure, in a way so that the resulting student matrix will partially correspond to | be aligned with matrices of pre-existing « teacher » (or teachers).

Further experiments with ELG are described in Kvasnicka and Pospichal (2007, 1999) and Hromada (2012b). All these studies point in the same direction and suggest that not only emergence of mutually shared communication protocol practically *ex nihilo* is possible whenever there exists a means of transfer of information among individuals, but also that without the presence of certain low amount of noise during the learning processs, the system as a whole would fail to

12 See following chapter to see closer introduction to what C and P-representations are.

converge to « communicatively optimal » state. In other words, ELG model indicates that presence of noise -a minimal yet not null amount of mal-transfered information- is necessary in order to assure that the population of mutually aligned sound-meaning matrices shall, sooner or later, converge into most communicatively optimal state.

The role of ELG model within the context of our Thesis is quite opened. For while it is the case that ELG sheds some light upon the question of emergence of language within a community of symbolicaly interacting agents, it does not, principially address the problem of language learning by a concrete individual. Thus, ELG is rather a model of macroscopic phylogeny than microscopic ontogeny - it addresses the problem of how small communities of homo habilis could, hundred years ago, gradually converge to system of signs within which, for example, « baubau » could mean a banana and « wauwau » mean a lion. Or, in less fatal and more vital affairs, it can be useful to synchronize activities problems related to dating, mating etc., as represented on Figure 12.

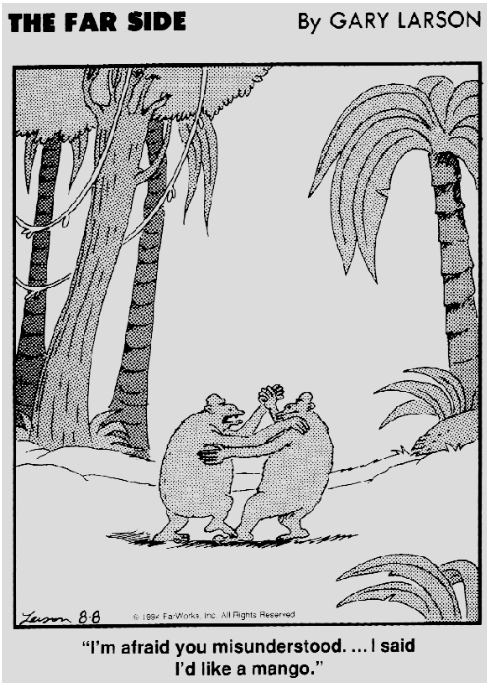


Figure 12: A case whereby mutual alignement of sound-meaning mappings can be useful. Reproduced from Kvasnicka and Pospichal (2007)'s reproduction in Pinker (2000).

Unfortunately, ELG wasn't explicitely constructed to address the problem of ontogenetic alignement, id est the problem of how toddlerese adapts to the motherese. But, we believe, it is not completely hors propos to imagine a slight variation of Nowak's model wherein one population of matrices would be much more stable (representing the linugistic competence of mother, parent or teacher agent) while the second population of matrices would represent the linguistic competence of a « child ». Given that the fitness function would somehow succeed to represent the degree of alignment between such « mother » and « child », we postulate that something like child's language competence could spontaneously emerge obe distilled and induced from ontogeny-oriented variant of Evolutionary Language Game.

END EVOLUTIONARY LANGUAGE GAME 8.7.5

In this section we have discussed more closely diverse applications of Evolutionary Computing (as defined in Section 3.1), namely

1. genetic algorithms (GA) and parallel genetic algorithms
2. evolutionary programming (EP) and evolutionary strategies (ES)
3. genetic programming (GP) and its variant grammatical evolution (GE)
4. an artificial ecology environment Tierra
5. model of *ex nihilo* induction of sound-meaning mappings called Evolutionary Language Game (ELG)

While some of these applications may strongly differ from each other they all materialize -sometimes in purely informatic or mathematic worlds; sometimes in worlds more material or even "social" - the basic premises of Universal Dariwnism. They all implement, in one way or another, reproduction, selection and variation of populations of information-encoding entities.

The content of

1. what these entities encode
2. ways how they encode it and how it varies
3. reasons why some structures are chosen into another generation and some not

all this varies substantially from application to application. But the trinity of principles: reproduction, selection, variation is implemented in all of them, otherwise they could not be, ex vi termini, labeled as EC implementations.

Dozens of analytical studies - related to topics as fitness-landscapes 8.7.1 or parallel genetic algorithms item 8.7.1 - could, sooner or later, find their accomplishement in a general, formal, and mathematical theory of evolution. Articulation of such theory could yield more rigorous a base for description of phaenomena which are nowadays explained in terms of somewhat vague, speculative and conjectural doctrine of Universal Darwinism. For the one who would decide to establish such a theory, EC could furnish tool as useful as was, for Kepler, the Galileo's telescope.

As was already indicated, the aim of this dissertation is not to furnish nor even discuss such general theory. The aim is to first use the conceptual prism of doctrine of Universal Darwinism in order to observe and interpret the phaenomena related to the topic of our interest - language acquisiton. And subsequently - in order to furnish a sort of *testimonium ex simulatione*- to use a most simple evolutionary model possible to demonstrate that it may be useful to conceive the problem of language acquisition in terms of gradual optimization and co-evolution of populations of linguistic functions and structures. We believe that for such a purpose, EC can furnish a very useful framework.

The reason behind this belief is simple - during few decades since its conception, EC-based systems have demonstrated their capability to find solutions to thousands of diverse problems and metaproblems. EC-based systems help designers and planners to invent optimal components, houses and cities; EC-based approaches are used to tune neural networks in robotic systems; EC-based systems help us not only to understand our world but also to change it.

Simply stated: Evolutionary Computing works.

END EVOLUTIONARY COMPUTING 8.7

Evolutionary Computing works because evolution itself works. And evolution - understood as gradual optimization of replicators - works, because it is a logical necessity.

Such is the doctrine of Universal Darwinism.

The goal of this chapter was to furnish a brief overview of diverse scientific theories and paradigms based on or inspired by UD's explicatory power. First was mentioned the biological evolution - it was the study of this form of evolution which gave birth to evolutionary theory. A discipline of Evolutionary Psychology was later discussed and partially criticized as being often too expansive in its aims. It was reiterated that the aims of this dissertation are not those of EP: while EP tries to explain diverse human skills as a result of biological evolution, the Hard Thesis postulates that human learning itself is an evolutionary process.

Evolutionary epistemology, Campbell&Simonton's explanation of individual creativity in terms of "blind variation and selective retention" and the notion of memetics were discussed as examples of evolution which is based on reproduction, variation and selection of non-DNA replicators. It was further precised that contrary to $EE_{2\text{-}1}$ and traditional memetics which study the evolution based on structures copied between the brains, we shall tend to put focus on evolution going on within the brain.

An existence of a sort of 3^{rd} replicator is thus posited. Asides nucleic acids - which furnish the material base for evolution of Nature; and asides memes - which represent the basic units of evolution of

Culture; a third replicator is posited in order to explain certain properties of a mind (1.1) which learns. To honor Piaget's work in genetic epistemology (8.4.4) we tend to call such replicator a "scheme".

By being internal to both mind&brain, such "schemes" are very elusive and it is of no surprise that they could potentially escape the attention of occidental "positivist" science. Even in case of other replicators, science took its time to recognize their nature and force. While breeding domesticated species for thousands of years, "science" was nonetheless ignorant of principles of evolution until a sort of crossover between Mendel's and Darwin's *ideae* occured. While being bombarded on a daily basis by propaganda memplexes and viral tweets, certain *scholae* have still somewhat difficult time to admit the sheer existence of memes.

And if the nature of such salient, objective, empiric phenomena escaped for such a long time the analytic regard of scientific enquiry, could there be anything done - in the limited scope of this dissertation - to demonstrate the existence of such "subjective" schemes ? After putting aside introspection as an invalid method of validating hypotheses in a positivist way, we see only three possible means of proving the existence of such 3rd replicator:

A. Study of reproduction of information within the brain by means of imaging techniques like fMRI, EEG etc.

B. Study of "schemes" when they are still observable, i.e. before they are interiorized.

C. Computational simulations

The path A, the path of neurosciences, is too costly and thus beyond our reach. Hopefully it shall be undertaken by others with more resources and more patience. But luckily, the price for undertaking paths B & C is negligeable and it is thus in this direction that we shall proceed. For in order to make progress on path B, one just needs to observe activity of minds who have not yet mastered the way how to interiorize into their subjective realm their perceptive and behavioral schemas. Such minds are, according to Piaget and even moreso by Vygotsky (1987): minds of children.

And when it comes to path C, nothing could serve us better than EC 3.1 branch of informatics. It has been suggested that EC is a sort of applied evolutionary theory: it can generate empiric proofs. Whenever a genetic algorithm discovers a useful solution which was not yet found, whenever a genetic programming scenario generates a piece of evolutionary art which the programmer haven't even dreamt of, a tangible -and often beautiful- proof is furnished.

A proof of belief that darwinists are definitely not further away than creationists from knowledge of a noumenic Principle governing our phenomenal world. END UNIVERSAL DARWINISM 8

9 DEVELOPMENTAL PSYCHOLINGUISTICS

Developmental Psycholinguistics (DP) is a scientific discipline studying changes occuring in human faculty of understanding and production of natural languages. As such, it is closely related to developmental psychology (a sub-field of psychology) and developmental linguistics (a sub-field of linguistics).

While developmental psychology thematises phaenomena development of human psyche, consciousness, mind, attention, reasoning, intellect, memory, perception, action, etc. on their own, DP does so always in relation to language. And contrary to linguistics, which often thematises language - or linguistic competence - as a product of some process P, DP ultimately strives to understand the process itself. In other terms, approaches common to DP « regard continuity of expression and function as critical clues to tracing the path children follow as they acquire language» (Clark, 2003).

Processing approach

We consider this distinction between "Product versus Process" to be of crucial importance for our tentative to align DP with UD. This is so, because the evolution itself is a process and hence it would be impossible to align the two paradigms if ever the linguistic faculty was understood solely as a static product. In other terms, alignement of DP and UD is possible only under the condition that the DP's main object of interest is not a static product, but a dynamic process.

As a name for such a process, we shall adopt the decision made by Harris (2013) and use the term «language development» preferably to widely-used term «language acquisition». A reason for this being the tentative to mark the fact that the child not only passively «acquires» the language from environmental input but rather gradually builds it, in interaction with its environment. Sometimes the term «language learning» is also used to denote the same process - a great take care has to be taken, however, not to forget that the "implicit and natural" way how a child learns toddlerese differs substantially from "explicit" drill used in learning of second, third, foreign, etc. languages.

More active than plain acquisition

This being said, we can know define the process which is, ex vi termini, the main object of interest of any developmental psycholinguist:

9.1 LANGUAGE DEVELOPMENT (DEF)

Language development (LD) - or ontogeny of natural language L in human individual H - is a constructivist process gradually transforming L into evermore optimized communication channel facilitating the exchange of information between H and her social surroundings.

END LANGUAGE DEVELOPMENT 9.1

Language Development is social and constructivist

The adjective constructivist indicates that LD should be, within the theory hereby introduced, considered as a process based on gradual internalization and modification of mental representations induced and re-induced by confrontations with external informations. Piaget's constructivist theory in relation to LD shall be more closely described in 9.4.3. Note also that by introduction of terms "verbal communication between H and social surroundings", the definition 9.1 places emphasis on the social aspects of human language. By doing so, it embraces so-called socio-pragmatic approach to LD (c.f. 9.4.4) more closely than so-called generativist and nativist ones (c.f. 10.2).

Optimized language allows to mean more but say less

But the key component of LD's definition is the notion of "optimization". This notion, which goes hand-in-hand with the notion of "facilitation of exchange of information" refers to the fact that, as language L develops - in infancy and beyond - it usually makes it possible to encode more precise an information with smaller quantity of signal. By integrating the notions of "optimization" and "facilitation of information", definition 9.1 thus ultimately states that language development is indeed a process which, if healthy and well-adopted to environment, makes it possible to successfully exchange ever subtler and subtler meanings (signifiées) encoded by shorter - or at least not longer - sequences of articulated symbols.

Conditions of success of a communicative act

An information can be successfully exchanged between human sender and receiver if and only if following conditions are fulfilled:

1. C_1 the sender is able to encode the information into the signal
2. C_2 the signal can be decoded by the receiver
3. C_3 the result of such decoding attracts receiver's mind limitely close to the state intended and anticipated by the sender

One can speak about success in intepersonal communication only if the communicative act fulfilles all of these conditions.

Of producing and parsing

As was already pointed out in 5.1, linguistic signals are usually strongly sequential and analysable into finite numbers of distinct discrete elements. When sender encodes his intention into such a sequence, (s)he is said to produce or generate the linguistic utterance. When receiver decodes it, he is said to parse the utterance. Ideally, when sufficiently strong a morphism exists between such meaning-encoding and signal-decoding interactors, the result of such parsing would have, as a consequence, a precious moment which humans call "understanding".

Of comprehension

Understanding, or comprehension, is closely related to the condition C_3. The fact that humans are able to understand each other, the fact that speaker and listener, writer and reader can share intentionality is, according to usage-based theorists of LD, something which seems to be a unique propensity of human species (c.f. 9.4.4 for further introduction of usage-based theories).

It is important to realize that in spite of disposing, at certain level of abstraction, of a sort of symmetry, production and comprehension are nonetheless distinct processes. It is as with movement of a hand which involves different muscles when the hand is going up and different ones when the hands move in the opposite direction; as with human endocrine system which uses one hormone to promote a certain activity and a completely different hormone to inhibite it; as with multitudes of other biological and cognitive phaenomena which seem to be the mirror images of each other but in fact are not: production and comprehension are distinct. Distinct mechanisms are implemented to generate a sentence and distinct to parse it. Distinct brain regions are involved. Hearing is not speaking with roles of speaker and heared simply inverted: it is something fundamentally different.

Of asymmetry between production and comprehension

The existence of such a mismatch between linguistic production and linguistic comprehension is so evident that many linguistic theories ignored it, or at least set it aside, as secondary. Practically all linguistic schools drawing inspiration from the Formal Language Theory (FLT) (10.2), e.g. the generativist tradition, do not care much about this distinction. This is so because at the level of abstraction where FLT is postulated, parsing is practically the same thing as generation and the only thing which differs is the direction in which rules are applied. . It is true that when parsing, system proceeds from surface structure towards deep structure by always substituting left-side of the production rule for the right-side; when generating, system proceeds from the deep structure towards the surface structure by substituting right-sides of the production rules for the left-sides. But all the rest - the content of production rules, the alphabet, the lexicon, the very computational machinery - is the same.

Of symmetry between production and parsing

Given more theoretical and much less empirical aims of FLT, one can understand the reasons why it practically ignores the mismatch between man's language comprehension and man's language production. One could even praise FLT's conceptors for the fact that by pointing to the level of abstraction where production and comprehension meet, they point to some potentially fundamental unity. For adopting an attitude where production is a sort of inverted parsing can, ineed, yield some interesting and potentially useful computer programs. But to build psycholinguistic theories and ignore the principle which every parent *feels* and every language teacher *knows*, such an attitude has to necessarily result in an inconsistent theory or a mal-functioning model.

Of human condition and insufficiency of FLT

In order to evit such an epistemologic disaster, a somewhat more mundane principle is being posited:

9.1.1 CENTRAL DOGMA OF DP (DEF)

C-representations precede P-representations.

END CENTRAL DOGMA 9.1

Of C- and P-representations

Eve Clark, who had coined these (C|P)-representation terms further clarifies: « Children set up a representation for each new word or phrase they notice in the speech they hear, attach meaning to it, and adjust the representation in the light of further analyses. They can use it to access that meaning when they next encouter that form. As they hear more language, they add to their store of such representations. These representations for comprehension (C-representations) consist first of an acoustic template, to which children then add information about meaning, syntax, and use...Children also represent the information needed for producing each expression. For this, they need specifications for articulating the sounds in the target word or phrase. Their representations for production (P-representations), then, necessarily differ from C-representations.» (Clark, 2003)

In simple terms, the central dogma states that humans understand language before they can speak it. A child comprehends what an *airplane* means long before it will be able to pronounce that word correctly. According to Clark, comprehension is always ahead of production, even in adult age when the mismatch is much less visible than in childhood.

C-representations are not passive

Important thing to realise is that utility of C-representations goes far beyond some passive involvement in recognition and comprehension of words and phrases. This is so, because C-representations can also influence and determine the direction of construction of P-representations. C-representations can provide targets with which P-representations are gradually aligned. This is how Clark describes the process:

« How would this work? Suppose a child is trying to produce *snow*. If children can access the C-representation for *snow*, they can compare their own production with the C-representation, detect any mismatch, and repair their own utterance. The C-representation is a model of what the word should sound like so others can recognize it. Under this view, C-representations provide model targets for what children produce. They also provide the target that the product of a P-representation must match. So as children adjust their P-representations to match what they hear from others, they align them with their C-representations. It is this gradual alignment that mirrors changes in children's own production of words and phrases.» (Clark, 2003)

In other words, what Clark tacitly indicates is that not only does human language development involve gradual adaptation of one's internal representations (C-ones) to structures observables in the external environment, but also that LD involves a sort of gradual adaptation of one set of internal representations (P) to another set (C). In light of such a theory can many common phenomena, like canonical babbling (9.2.2), for example be interpreted as highly useful and possibly inevitable ways how infant's linguistic faculty tunes and bootstraps itself in partially auto-programming and auto-poietic fashion.

Another reason why we consider the principle C-precedes-P can to be of certain interest for this dissertation is, that it indirectly addresses the debate we have already raised when discussing the hypothesis stating that "learning involves reproduction of information-encoding entities" (2.4). For if we accept that C-precedes-P, we have to accept, in the first place, that representation of the word mama is somewhow distinct from the P-representation of the same word. And more: if we accept that the C-representation of the word mama is distinct from the P-representation of the word mama, yet refers to the very same mother-referent in the external world, we have to accept that the information contained in two representations has to be, at least partially, the same.

Central Dogma implies information replication

We thus end up with two distinct representations, C and P but both originating in C and pointing to the referential content to which C sole referred when it got set up. Couldn't this mean that informational content of locus which encodes C (e.g. in Wernicke's area) got replicated into independent cortical locus which encodes P (e.g. in Broca's area) ? Couldn't the neural basis of such process be somewhat similar to processes postulated by neural darwinists (8.6), for example the one depicted on figure Figure 6?

We let the reader (him | her)self to answer these and similar questions. Note, however, that answering these answers with "yes" would suggest that the statement which have been labeled hereby as a central dogma of developmental psycholinguistics, does indirectly support the thesis that language development is a form of evolutionary process.

9.2 DEVELOPMENT OF TODDLERESE

The goal of following subsections is to present facts related to development of multiple facets of toddlerese. In 5.1, the toddlerese was defined as a protovariant of the natural language, and natural language was defined as a system composed of prosodic, phonologic, morphologic, syntactic, semantic, and pragmatic structures and principles (4.1). None of these layers is to be ignored by somebody aiming to have an adequate vision of development of toddlerese.

Facets of toddlerese

But taking into account all scientific discussions which were, since end of 19th century, pre-occupied with elucidation of mystery of LD's universality, speed and the fact that in case of healthy individuals, LD is practically always successful, taking into account all such scholastic schisms is not a path to knowledge neither. Thousands of experiments and observations were done, hundreds of books published, dozens of theories and even whole doctrines were unleashed, sometimes sentencing whole generations of linguists into the scholastic hell filled with infinities, recursive rules and utterly inconvenient formalist games. In order to evit such a destiny, following paragraphs shall restrict themselves to very "minimalist" presentation of few evident or experimentally well-verified LD-pertaining facts.

Thus, the brief expose hereby introduced will be only very rarely concerned with any linguistic phenomena beyond the state of the toddlerese, whose upper bound was operationalized, in 5,at 30 months (i.e. at age 2;6). And given that the "operational thesis" (6) restricted the scope of our interest to textual modality of human intepersonal communication, we shall present in closer detail the psycholinguistic studies pertaining to development of morphosyntactic and semantic faculties. In contrast to these, prosodic, phonic and pragmatic layers shall be described much more superficially than they rightfully merit. For when it comes to language as was known to all our human predecessors, it was indeed the pragmatic and phonetic aspect which were at the core and inception of it all.

9.2.1 ONTOGENY OF PROSODY, PHONETICS AND PHONOLOGY

Prosody, Phonetics, Phonology (PPP)

Prosody is all that relates to tempo, rhythm, stress and intonation of speech. Phonetics is concerned with articulation, acoustics and audition of physical properties of speech signs. Phonology, on the other hand, is less "material" and more "cognitive" in a sense that it is not concerned with such physical characteristics of phonemes like amplitude, frequency or timbre, but rather with systems of abstract categories and rules whose existence is directly or indirectly observable in case of any human cognitive system which was exposed to phonemes and is able to perceive them.

Prenatal C_{PPP}-representations

Human beings are sensitive to language even in the prenatal period. The study of DeCasper and Spence (1986) has shown that new-born infants prefer to listen to a story which they have already "heard" *in utero*. Given the fact that in uterus, frequencies about 1kHz are weakened by transmission through maternal tissue, this preference could be explained principially in terms of prosodic and not phonemic information. Another study had indicated that even 4-day old newborns are able to distinguish between mother language (e.g. French, in case of French newborns) and a foreign language (Russian, English etc.) filtered by a 400Hz low-pass filter Mehler et al. (1988). Clark summarizes the results of both studies in a statement: « what infants are attending to are the prosodic properties of the speech they have been exposed to prenatally» (Clark, 2003).

During approximately first eight months which follow the birth, infants are capable to discriminate practically any phonetically plausible contrast between two sounds. But before attaining one year of age, children loose this capacity to distinguish practically any sound from any other and their perceptual filters become more and more adapted to phonology of language spoken in their social environment. In other words, « infants can discriminate nonnative speech contrasts without relevant experience...there is a decline in this ability during ontogeny...data...shows that this decline occurs within the first year of life, and that it is a function of specific language experience.» (Werker and Tees, 1984)

Decline in universal discriminative capacity

It shall be indicated multiple times in this dissertation that sometimes a loss or limitation can serve a creative purpose. Such is also the case, we believe, in case of the above-mentioned loss of capacity to distinguish practically any phoneme from any other. For by losing this capacity, an infant also gains something: she gains the capacity to distinguish language from non-language, *the mother language* from a language spoken by an alien passing by. When this problem is resolved, child's cognitive system can *focus* more efficiently upon the upcoming problem: that of discovery and extraction of recurring patterns in and from the speech stream.

Less is more 1

Set of experiments, performed by Jusczyk and his colleagues, focused principially on infants' ability to "hear" such regularities. One type of regularities are prosodic ones, for example syllabic stress patterns (stronger stress on first syllables in English). Another regularities are, of course, due to repetitive occurences of the same words. In order to remark that the word X was repeated, an infant has to be able to somehow identify the word as something which was already heard. The study of Jusczyk and Aslin which focused on perception of monosyllabic words has shown that « some ability to detect words in fluent speech contexts is present by 7 and half months of age» (Jusczyk and Aslin, 1995). The same study has also indicated that 6-month old infants still lack such ability to perceive (monosyllabic) words as perceptual units.

9-month-olds match lexical patterns

At 9 months, infants are able to identify sequences of two and more syllables: « 9-month-olds appear to be capable of integrating sequential and suprasegmental information in forming wordlike (multisyllabic) phonological percepts, 6-month-olds are not» (Morgan and Saffran, 1995). Another study indicates that 9-month-olds also prefer to listen to words of their ambient (mother) language and not words from another language Jusczyk et al. (1993). Before attaining first year of age, children are able not only to discriminate but also to identify familiar phonemic chunks of various sizes, extract them from the speech stream and potentially associate with contextual information and other sensory modalities (visual, tactile etc.). It is therefore reasonable to assume that 9-month old healthy infant already disposes of dozens of C-representations which can be labeled as protolexical. Their transformation into full-fledged lexical structures shall be discussed in next section.

Ontogeny of infant's faculty to produce intelligible verbal signal is no less fascinating. It starts, of course, with the cry of a new born who is able to obtain any wished change of environment (food, warmth, diaper change etc.) with one loudly and adamantly expressed bit of information. But after cca 2 months, an infant starts to produce more gentle cooing sounds which seem to express, contrary to crying, infant's satisfaction or agreement with the current state of environment. In three and four months which follow get these two modes of verbal production - crying and cooing - still and more refined and are evermore accompanied by facial and gestural expressions. And sometimes - when the cooing vowel-like "ooo" and "aaah" are co-articulated with some occlusive consonants, thus forming sounds like "uuum", "baaa" or "maaa" - one can constate the occurence of marginal babbling.

Of cry and cooing

And then, somewhere between six and ten months, comes canonical babbling.

9.2.2 CANONICAL BABBLING

« Canonical babbling consists of short or long sequences containing just one consonant-vowel (CV) combination that is reduplicated or repeated.» (Clark, 2003)

END CANONICAL BABBLING 9.2

Consonants occurent in canonical babbling are more than often voiced and labial[1] (b), labionasal (m), velar (g) and little bit later - when child already has some teeth to block the airflux with - also dental (d). [2] Few months which follow shall be subsequently dedicated to *variation* of both the enveloping intonation contour of the babbling as well as the syllables contained in the babbling sequence: canonical "mamama" sequences shall thus evolve into sequences like "mamapapadadada?". At cca 1 year of age « many babbled sequences sound compatible with the surrounding language using similar sound sequences, rhythm and intonation contours.» (Clark, 2003)

Development of babbling

It is during the period of babbling when first "words" appear. And according to growing amount of evidence, the development of first words is a natural and continuous prolongation of babbling phase. Elbers and Ton, for example, summarize their analysis of monologues of a Dutch boy Thomas in the six weeks following acquisition of his

Emergence of first words

1 In our bachelor's thesis we had emitted the hypothesis that prominence of labial closures in early babbling is to be associated with succion.

2 Development of canonical babbling

first word (1;3-1;5) with the conclusion: « new words may influence the character and the course of babbling, whereas babbling in turn may give rise to phonological preferences for selecting other new words» (Elbers and Ton, 1985). For example the frequency of t-like vowels occurent in the babbling sequences has increased significantly (from 15% to 40%) in the period when Thomas started to use his t-containing word ("aut(o)").

Preference and avoidance

Toddlers thus seem to be selective in word forms which they pronounce, « working first on what they can already do and only after that moving on to harder problems» (Clark, 2003). When they do not have enough practice with a certain sound or a word form, they tend to avoid it. This hypothesis was to be demonstrated by an ingenious experiment designed as follows: « during 10 bi-weekly experimental sessions, 12 children (1;0.21 - 1;3.15) were presented with 16 contrived lexical concepts, each consisting of a nonsense word and four unfamiliar referents. For each child, eight words involved phonological characteristics which had been evidenced in production (IN) and eight had characteristics which had not been evidenced in production or selection (OUT)» (Schwartz and Leonard, 1982). The results of the experiment, presented on 2 made it evident that while children's ability to understand is independent from the form of the word-to-be-understood, toddlers and pre-toddlers prefer to "mention" mainly those things, whose names contain only familiar phonetic forms (i.e. IN words).

	IN WORDS	OUT WORDS
Produced spontaneously	33	12
Understood correctly	54	50

Table 2: Children avoid production of words with unknown characteristics. Reproduced from table Clark (2003) based on data in Schwartz and Leonard (1982).

To get from babbling to rich spectrum of intelligible words is not an easy task. Every child uses its own unique strategy to solve it; every child traverses a different "path" in order to align its linguistic structures to those of her social environment. As the author of a thorough study comparing acquisition of phonology by three children put it: « each of the three children is exhibiting a unique path of development with individual strategies and preferences and and idiosyncratic lexicon» (Ferguson and Farwell, 1975). We consider it important to underline that rarely are these path a linear descent from random babbling to optimal (i.e. correct) pronounciation. As can be seen not only in the data collected by Ferguson & Farwell it is often rather the contrary which is the case: « although the children tend to be quite accurate in their first production, their accuracy often decline over time, so

From babbling to words

later versions of the same words appear to be further from the adult targets» (Clark, 2003).

Of variation of PPP structures

What seems to be common, however, to all those paths is that they flourish with variation. As William and Terese Labovs have observed during the longtitudional observation (1;3-1;8) of their daughter Jessie, she revealed « continuous exploration, experimentation, practice and intense involvement with linguistic structure» (Labov and Labov, 1978). For 3 months of Jessie's life, this experimentation was concerned solely with words "cat" and "mama"; overall she had pronounced each of these terms at least 5000 times during the 5 months of observation. « In summary, what might be regarded as a rather flat plateau in Jessie's development, upon closer inspection, revealed a constantly changing series of small experiments where she progressively scrutinized and tried out different phonological options.» (Clark, 2003)

This small experiments can be often characterized in terms of application (or non-application) of specific simplification routines. These routines, which we shall call "variation operators" in iii can be coarsly divided into three big groups of

1. Substitutions
2. Assimilations
3. Transpositions

Of substitutions

Substitutions are due to simple replacement of one sound (or group of sounds) with another sound or group of sounds. Common is voicing of initial voiceless consonants ("pie" pronounced as [bay]), devoicing of final ones ([nop] <- "knob"), gliding ("ball"->[baj]) etc. Also, children often do not pronounce some parts of word at all. These *omissions* - which can be understood as special cases of substitution whereby one sound is substituted for a "blank" or "non-terminal" sound which is not articulated - are also very common especially in case of consonants at initial ("tram"->[am]) or final ("pes"->[pe]) positions.

Of assimilations

Assimiliations « refer to the effect of sounds on those preceding or following them within a word or across word-boundaries» (Clark, 2003). Assimilated can be only one or few features, e.g. in ("orol" -> [olol] does the lateral feature of final "l" override the trill feature of "r") for backward lateralization or ("balon"->[balol]) for forward lateralization. But whole cluster of features or even sounds can be assimilated as well: in particular is this the case in syllable reduplication whereby one syllable completely overrides the other ("wasser"->[vava]).

Of transpositions

Another group of simplification procedures & variation operators are transpositions. Known as "metathesis" in historical and evolutionary linguistics (8.5) and analogic, mutatis mutandi, to so-called chiasms (Hromada (2011); Dubremetz (2013)) in rhetorics these switches in order (AxB->BxA) are already at play in production of toddlers ("KOstOL"->[okol]).

All these examples shall be in closer detail discussed in iii. And in the second volume of this thesis, these cases shall be formalized and subsequently embedded as "variation operators" into evolutionary computation scripts. But for the purpose of this expose let's just limit ourselves to the constatation that the sequence of application of similiar routines shall, in course of ontogeny, allow the child to converge from a quasi-random babbling to correct articulatory program able to produce word X.

END ONTOGENY OF PPP 9.2

9.2.3 ONTOGENY OF LEXICON AND SEMANTICS

Of semantics

Raison d'etre of language is to communicate meanings. Semantics is a scientific discipline devoted to study of meanings. Meaning - also called *signifie* in tradition established by (de Saussure, 1916) - is a fairly abstract entity which does only rarely, if ever, exists on its own. In language, meanings are always coupled with "signifiants", i.e. with material phonetic or graphemic forms which denote some specific meaning.

Of sign

Signifiant, signifié and information related to morphosyntactic properties (c.f. 9.2.4) form a triad which, taken all together, composes a word. In modern linguistics, words are sometimes considered to be members of "lexicon". A lexicon is simply a set of all words internalized by and represented within the individual cognitive system. In DP, the process of acquisition of lexicon is also known as vocabulary development.

We consider the process of vocabulary development to be, in huge extent, reducible to the problem of construction of semantic categories. Under such view, the problem of understanding of a new word W could be understood as the problem of:

Of semantic categories

1. detection of recurrence of W in speech
2. establishment of mapping|association between W and corresponding semantic category C[3]
3. reducing or increasing the extension of C so that it is neither too specific nor too general

None of these problems is computationally trivial but children nonetheless solve both of them with stunning swiftness and ease. We think

3 We consider it important to precize that within the theory hereby proposed, semantic categories - understood as points, regions or subspaces of some sort of "absolute semantic space" - can be *shared*, i.e. accessed by multiple mutually independent cognitive agents.

that this is so, because human brain (2.8) is principially a pattern-detecting computational device whose principal objective, especially during initial stages of ontogeny, is to subsume huge amount of contextual multi-modal information under and into as-neatly-as-possible packaged categories. Under such view a word W, a signifiant, is not only a "label" for its respective conceptual category; it is also a stimulus triggering a completely unvoluntarily categorization process. As Kyra Karmiloff and her mother Annette put it: « there is a dynamic feedback between developing cognitive skills and growing vocabulary, and words can act as an invitation to form a category» (Karmiloff and Karmiloff-Smith, 2009).

Since we shall return later (10.4) to more theoretic discussion of what semantic categories "are" in computational sense, and how mapping between them and their labels can be constructed in a general computational system, let's just focus on the question "What are particular aspects of acquisition of semantic categories constructed in human children?".

As infants gradually overcome perceptual limitations of the newborn state they tend to see the world evermore clearly. This subsequently makes it possible that « very young infants can and do perceive even the most subtle differences between and across category members. One study showed that three-month-olds could not only differentiate between cats and dogs (a between-category distinction), but also distinguish among different kinds of cat (a within-category distinction))» (Karmiloff and Karmiloff-Smith, 2009).

During first year of age, initial C-representations are being formed by associating such representations of perceptual categories with co-ocurrent representations of most frequent and salient forms which the infant succeeds to detect and identify in her linguistic environment. Interaction with such environment - consisting mainly of mother, father, siblings or other "tutors" - is dynamic, repetitive and goal-oriented. Roger Brown describes it as a "word game" of which the child is a principal player:

Of word game

« The tutor names things in accordance with semantic customs of the community. The player forms hypothesis about the categorical nature of the things named. He tests his hypothesis by trying to name new things correctly. The tutor compares the player's utterances with his own anticipations of such utterances and, in this way, checks the accuracy of fit between his own categories and those of the player. He improves the fit by correction.» (Brown, 1958)

To understand what object is meant by what name is not an easy task. For how does a child know, that word "milk" means the life-strenghtening liquid and not white color, liquid in general, something to drink or vessel in which it is stored? A possible answer can be: by application of diverse lexical constraints (LCs). Among multiple LCs mentioned in the litterature, we consider these:

Of utility of lexical constraints

1. whole-object assumption
2. basic-level assumption
3. taxonomic assumption
4. mutual exclusivity and fast mapping

constraints to be of biggest importance during the toddler stage of LD.

The whole-object assumption « presupposes that children already have categories of objects, such that objects can be represented as whole entities distinct from their locations or from their relations to other objects or places.» (Clark, 2003). It is evident that endowing humans with could be quite useful for our survival as species: to be able to immediately percieve and label a lion as a lion is more "fit" a strategy than to invest computational resource's in seeing details of lion's fur or whiskers. The same applies for *the basic-level assumption*: ability to parition the world into the basic-level categories Rosch (1999) which are not too general (above basic level), nor too specific (below basic level) is crucial to survival. In comparison to one's ability to categorize a shark as a shark, is the ability to categorize these predators into below-basic-level categories as blue, white or tiger sharks or as members of above-basic-level category of chordates, somewhat secondary.[4].

Of whole-object assumption

Another LC which is quite closely related to Rosch's theory of basic-level categories and prototypes (10.4.1) is the *taxonomic assumption* which presupposes that labels should be *a priori* extended to object of the same kind and not the object which is thematically related. Its validity was demonstrated by experiment in which « children saw a series of target objects (e.g., dog), each followed by a thematic associate (e.g., bone) and a taxonomic associate (e.g., cat). When children were told to choose another object that was similar to the target ("See this? Find another one."), they as usual often selected the thematic associate. In contrast, when the instructions included an unknown word for the target ("See this fep? Find another fep."), children now preferred the taxonomic associate.» (Markman and Hutchinson, 1984)

Of taxonomic assumption

While above-mentioned LCs are useful heuristics for determining either the nature or scope of categories-to-be-constructed, LCs of fast mapping and mutual exclusivity are heuristics facilitating the discovery of relation between the label (signifiant) and semantic category (signifie). Thus, « the mutual exclusivity constraint stipulates that in a given language an object cannot have more than one name, so if the child already knows the word "car," he will not think a new word refers to cars. In other words, in the early stages of word learning, the child does not expect synonyms. The second constraint, fast mapping, stipulates that novel words map onto objects for which the child

Of fast mapping and mutual exclusivity

4 This does not apply for professional biologists and philosophers, of course.

People: mommy (1;0), daddy (1;0), baby (1;3)
Food: banana (1;4), juice (1;4), cookie (1;4), apple (1;5), cheese (1;5)
Body parts: eye (1;4), nose (1;4), ear (1;5)
Clothing: shoe (1;4), sock (1;6), hat (1;6)
Animals: dog (1;2), kitty (1;4), bird (1;4), duck (1;4)
Vehicles: car (1;4), truck (1;6)
Toys: ball (1;3), book (1;4), balloon (1;4)
Household objects: bottle (1;4), keys (1;5)
Routines: bye (1;1), hi (1;2), no (1;3)
Activities: uh oh (1;2), woof (1;4), moo (1;4), ouch (1;4)

Table 3: Words produced by at least half of children in the monthly sample. Reproduced from table in Clark (2003) based on data from Fenson et al. (1994).

does not already have a name.» (Karmiloff and Karmiloff-Smith, 2009) Both lexical constraints of mutual exclusivity and fast mapping can be understood as direct implications of principle of contrast.

The Principle of Contrast (DEF)

« Every two forms contrast in meaning.» (Clark, 1987)

END THE PRINCIPLE OF CONTRAST 9.2.3

Of usefulness of principle of contrast

The importance of this fairly trivial principle in regards to LD is not to be underestimated. Acquisition of any kind of form-meaning mappings can be significantly catalysed by the sole fact that PoC applies. Take, for example, an information-processing agent which knows only what "mama" means, but often hears an expression "mama a tato" when it simultaneously sees her mother and father. Discovery that the form "tato" denotes "father" would be trivial for an agent with PoC embedded among her information-processing procedures. And quasi impossible or very (computationally) costly for an agent who does not.

Of insufficient and excessive generalization

Thus, with aid of very restricted number of heuristic-like principles and constraints, and in combination with contexts which repeat themselves day after day and week after week, small infants shall start to associate first linguistic forms to first conceptual categories. But the sole establishment of this association between the word and the category is not sufficient. The scope, the extent, the region of semantic space covered by and attributed to the specific category have to be delimited as well. Before it shall be the case the child shall commit many errors of either insufficient or excessive generalisation. For example, in case of insufficient generalisation, it shall sometimes apply a generic label ("dog") to denote just one specific canine ("lessie"). And in case of excessive generalisation, it shall denote with a label ("cat") even referents ("lynx") upon which such label is not commonly applied by child's linguistic community. Clark (2003) offers a nice example, extracted from Kuczaj's transcripts [5] contained within CHILDES corpus, in which a child (2;4) learns a new word which shall help her to narrow down a general verbal meaning:

```
I (wanted to have his orange peeled) : Fix it.
T : You want me to peel it?
I : Uh-huh. Peel it.
```

Section 13 shall present some more detailed results related to such microconversations resulting in a correction of child's semantic category. For the time being let's just suggest that such parental or sibling corrections could be quite easily integrated into a darwinian model of language ontogeny either as a sort of selection or variation operator. Such kind of an exogenous, environment-originated perturbations gradually divide infant's conceptual space into structure of partitions functionally isomorph to structure of partitions embodied in child's tutor. Table 4 illustrates in a very brief but nonetheless *parlant* way an example of how relations between few labels and subjacent categories changed in ontogeny of one particular child.

Of usefulness of exogenous feedback

WORD	INITIAL AND SUBSEQUENT REFERENTS	MORE APPROPRIATE WORD
papa	father/grandfather/mother (1;0)	mama (1;3)
	any man (1;2)	Mann (1;5)
Mann	pictures of adults (1;5)	
	any adult (1;6)	Frau (1;7)
ball	ball (1;0)	
	balloon (1;4)	balloon (1;10)

Table 4: Case of development of word | meaning mapings. Based on data in Barrett (1978).

Thus, an important « part of learning a word meaning is also learning what the extension of each term is, by learning what counts as a possible referent. Children also try out some words in ways that are hard to link to any identifiable use. The target word itself may not be identifiable, and the general absence of adult comprehension typically leads to the word's being abandonded» (Clark, 2003). We propose to interpret this tendency to "abandon not identifiable words" as

5 In transcripts of conversations with children we shall label child-directed sentences with I (meaning "infant") and adult-generated sentences with T (meaning "tutor").

Of selection of vocables

a sort of selection. Cumulation of multitudes of such selective events combined with playful variation inherent to every healthy child shall, so we argue, gradually attract child's mind into a state where she shall dispose of language of her surroundings.

Of graduality of word-learning

And learning of concepts is indeed gradual. Analyses of maternal journals and estimations suggest that at 12 months of age, children understand on average at least ten words (Menyuk et al., 1991). In following months increases the size of the lexicon only slowly; topics for the words which the child understands and is subsequently able to produce are also quite restrained:« not surprisingly, young children talk about what is going on around them: the people they see every day; toys and household objects they can manipulate; food they themselves can control; clothing they can get off by themselves; animals and vehicles both of which move and so attract attention; daily routines and activities; and some sound effects» (Clark, 2003). Table 3 contains a list of words produced at given age by at least 50% among 1803 children whose parental reports were studied by Fenson et al. (1994).

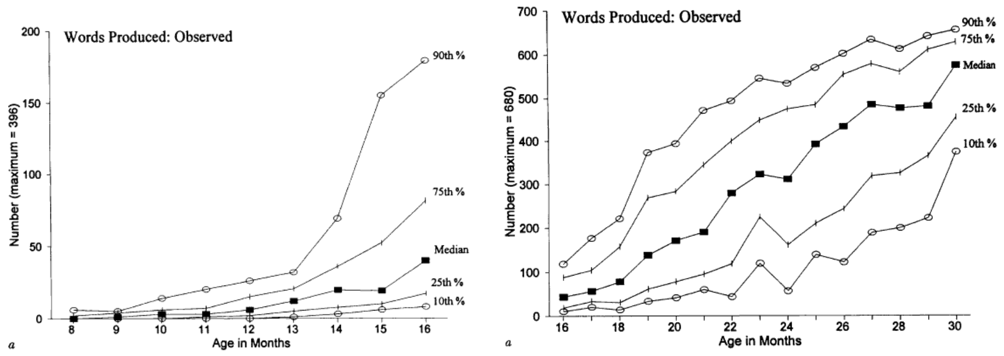
Of proto-imperatives and proto-declaratives

From the perspective of end-state language are many among these first words a specific-object-denoting nouns. But children often use them with function of verbs or, more specifically, as imperatives. Thus, when saying "milk" a small child expresses her wish, want and need meaning of "wanting milk" or "make me get that bottle". Only months later shall be such proto-imperatives accompanied with proto-declarative statemens meaning "look, mother, there is milk!". This distinction between proto-imperatives and proto-declaratives is not to be underestimated since it seems to stem from child's *growing will to share information*. Since we shall return later (9.4.4) to this properly human tendency to share information, intentionality and attention with others, let's just express our agreement with the statement that « Using language simply to share a common experience with the listener is particular to human communication. Animals tend only to use communication in a proto-imperative way» (Karmiloff and Karmiloff-Smith, 2009).

Of vocabulary explosion

It is approximately in the period of gradual passage from proto-imperative to proto-declarative use of language, i.e. between 16-20 months, when the rate of acquisition of vocabulary shall start to accelerate. This phenomenon, known as vocabulary explosion or vocabulary spurt, starts to express itself when child's productive vocabulary attains approximately 150 words and can be described as follows: « Prior to the vocabulary spurt, children learn on average about three words per week. But when they enter the vocabulary spurt stage, their learning of new words increases dramatically to about eight to ten words per day» (Karmiloff and Karmiloff-Smith, 2009). We shall discuss the phenomenon of vocabulary spurt in somewhat more quantitative terms in 10.1.2, during the discussion of Logistic law. Figure 13

Figure 13: Development of productive vocabulary in early (a) and late (b) toddlerese. Figures reproced from Fenson et al. (1994).



(a) (b)

shows the development of productive vocabulary in 1803 children as measured by McArthur infant and toddler communication development inventories.

Authors like Marchman and Bates (1994) interpret occurence of this and other similar LD-related phenomena in terms of attainment of *critical mass*.[6]. It seems indeed reasonable to postulate that some sort of qualitative change of toddler's linguistic faculties occurs during the period when she enters the vocabulary spurt: for approximately in the same period the toddler shall start to juxtapose words side by side and construct first phrases. And that marks the advent of morphology and syntax which shall be discussed in following section.

Of critical mass in word-learning

Before we end this very brief overview of word learning among toddlers, let's just reiterate the founding that « one of the interesting characteristics of words is that their meanings do not remain static; they can change» (Karmiloff and Karmiloff-Smith, 2009). And this "change" is a part of process which is usually called "learning". And this "learning" can be, we suggest, plausibly interpreted as a particular case of evolutionary process which, during ontogeny, divides one's semantic space (10.4) into categories which shall tend to overlap with categories categories "out there".

9.2.4 ONTOGENY OF MORPHOSYNTAX

In traditional linguistics, most fundamental meaning-carrying units of linguistic analysis are not individual words, but so-called morphemes. That is, prefixes, suffixes, word roots or other materially

6 Phenomena which occur when only a certain critical mass is attained are best studied by the theory of complexity (c.f. Kauffman (1995) for gentle introduction). Such phenomena are often related to a so-called "phase transition" (e.g. water → ice; fuel → reactor) which can be accompanied with not only quantitative but also qualitative transformation of the observed system.

encoded signifiants encoding a particular signifie. The particular system of mutual interactions of diverse categories of morphemes yields a particular morphology. When interaction of morphemes surpasses an individual word and multiple words are concatenated in a full-fledged utterance, *an information can be contained also in the way in which diverse components (words) are ordered*: in utterance's syntax (σύν -> "together",τάξις -> "ordering"). Since in many languages, the distinction between morphology and syntax seems to be very fuzzy[7], some linguists prefer to speak simply about morphosyntax.

Similiary to many other forms of human activity (e.g. object-manipulation, food-preparation, rituals etc.) are human languages compositional and combinatorial. Compositionality can be defined as follows:

Compositionality (DEF)

« The meaning of a signal is a function of the meaning of its parts, and how they are put together.» (Brighton et al., 2003)

END COMPOSITIONALITY 9.2

while combinatoriality means that theoretically infinite - yet practically vaste but finite - of complex constructions can be obtained by means of combining of finite amount of elements (morphemes).

Evolutionary and computational advantages of compositionality and combinatoriality of natural languages being adressed elsewhere (Brighton et al. (2003); Kvasnicka and Pospichal (1999); Pinker (2000)) let's now focus on other characteristics related to development of syntax in practically any healthy human child.

Of variability in LD

One such "universalium" is that there exists both inter- (i.e. children of different communities acquire different languages) and intra- (i.e. children of the same community acquire their language differently) linguistic variability. The variability of developmental trajectories is in fact so huge, that one could plausibly argue that there are no two children in the world - not even twins[8]- which would acquire language in an absolutely identic way.

Of non-determinism in LD

This is so, becaue language-learning is strongly dependent on individual perspective as well as context within which the learning occurs. Contexts from which child acquires language structures involves not only auditive, but also visual, emotional, social, etc. dimensions, and internalization of structures thus involves many factors. Since some of these factors are stochastic, language-acquisition itself can NOT be a fully deterministic process. This is the second "universalium".

7 Take, for example, the German word of the year 1999 "Rindfleischetikettierungsüberwachungsaufgabenübertragungsgesetz" which, in fact, is quite minimalist in comparison to words uttered by classical sanskrt poets. Are the rules which govern composition of such words the rules of morphology, or the rules of syntax?

8 In this context, we consider it worth mentioning that twins often develop a sort of their own language, or "idioglossia", whose potential conflict with ambient language can slow down twins' language development.

Asides compositionality, combinatoriality, variability, context-boundedness and non-determinism, we consider it worth to mention these other characteristics which could be considered as universal and axiomatic: *Other LD universalia*

- graduality: children tend to master shorter structures before they master longer structures
- cumulativity: children tend to "build upon" what they already know
- specificity: children tend to learn individual patterns in individual contexts of social interaction
- repetitivity: scenes in which children acquire individual structure X contain certain recurrent features
- inductivity: specific structures can be transcontextually crossed-over to yield structures corresponding to more general meanings than the ones with which the child was already confronted

Of course the list does not end here and other properties like (Tomasello, 2009), recursivity (Chomsky, 1957), syllabicity (Jackendoff, 2002) or importance of substitution were rightfully highlighted. Since some of these shall be discussed in 9.4 and 10.2, let's now lay aside the generalities and focus upon facts. First, production:

Being still in their babbling phase, children first produce one-word "holophrases" which they succeed to fit into an individual intonational contour. As intentions they want to communicate get more and more complex, children couple these with movements like approaching or running away; with gestures like pointing, nodding or shoulder-shrugging; or even with more complex manipulation like object bringing, throwing or showing. In sum, « gestures appear to help young children communicate before they can pronounce the longer phonological sequences required for combining words» (Clark, 2003). *Of first phrases*

As the temporal span of intonational contours increases[9] and as child improves its pronounciation of individual words - thus reducing the cognitive cost related to phonetic aspects of the utterance - she succeeds, normally around cca 18 months of age , to fit multiple words under the *vault* of a single intonational contour, thus creating a first two-word construction. *Of 2-word constructions*

According to (Tomasello, 2009, pp. 104), this primordial "word combinations" stage have two distinctive features:

- they partition the scene into multiple symbolizable units
- they are composed only of concrete pieces of language, not categories

9 Possibly because of slowing-down of the "internal oscillator" observable in experiments with so-called spontaneous tempo (c.f. 9.2.6)

A concrete example MAMA NENE (meaning "mother-breast") shall be further discussed in 12.7.1.

Child's ability to concatenate two words and integrate them into a single intonational contour is swiftly followed by emergence of so-called pivot schemas.

Pivot schema (DEF)

A two-word schema in which « one word (the "pivot") recurres frequently in the same position in combinations, and the other word *varies*[10]» (Braine and Bowerman, 1976)

END PIVOT SCHEMA 9.2

A canonic example of what is meant by pivot words and pivot schemas is presented on table reproduced on Figure 14. This table lists all comprehensible two-word combinations, noted by the mother, which the toddler named Andrew spontaneously produced during first five months after leaving the single-word stage.

Figure 14: Corpus of two-word utterances produced by a toddler Andrew. Reproduced from Braine and Bowerman (1976).

ANDREW'S WORD COMBINATIONS

more car[a]	no bed	other bib	boot off	see baby
more cereal	no down[c]	other bread	light off	see pretty
more cookie	no fix	other milk	pants off	see train
more fish	no home	other pants	shirt off	
more high[b]	no mama[d]	other part	shoe off	hi Calico
more hot	no more	other piece	water off	hi mama
more juice	no pee	other pocket	off bib	hi papa
more read	no plug	other shirt		
more sing	no water	other shoe	airplane all gone	airplane by[h]
more toast	no wet[e]	other side	Calico all gone[f]	siren by
more walk			Calico all done[f]	mail come
outside more	down there		all done milk	mama come
	clock on there		all done now	what's that
all broke	up on there		all gone juice	what's this
all buttoned	hot in there		all gone outside[g]	mail man
all clean	milk in there		all gone pacifier	mail car
all done	light up there		salt all shut	our car
all dressed	fall down there			our door
all dry	kitty down there		byebye back	papa away
all fix	more down there		byebye Calico	look at this
all gone	sit down there		byebye car	pants change
all messy	cover down there		byebye papa	dry pants
all shut	other cover down there		Calico byebye	
all through	up on there some more		papa byebye	
all wet				

Of affinity of pivot words

In Andrew's case, the pivot words are "more", "no", "all", "other", "there", "off", "all gone", "all done", "byebye", "hi" and "see". It can be immediately seen that pivot words tend to be juxtaposed with words belonging to specific linguistic categories ("more" with "nouns", "all" with adjectives or participles). Where are categories, there is generalisation and where is generalisation, there is productivity and, indeed,

10 Word "varies" put in italics by the author of this Thesis.

had been such productivity of pivot schemas experimentally demonstrated in (Tomasello et al., 1997). In retrospect, its author concludes it as follows: « 22-month-old children who were taught a novel object for an object knew immediately how to combine this novel name with other pivot-type words already in their dictionary» (Tomasello, 2009).

Another interesting result of the same study was that « children combined the novel nouns productively with already known words much more often than they did the novel verbs – by many orders of magnitude» (Tomasello et al., 1997). But because categories like "nouns" and "verbs" are results of adult categorization of certain lexical phenomena and not necessarily categories pertinent to child's own linguistic experience, let's just limit ourselves to trivial constatation that specific pivot words have affinity to words with specific features. And vice versa.

Of item-based constructions

These mutual affinities between "constant" pivot words and their variable "complements" result in emergence of populations of microsystems of productive order, which (Tomasello, 2009, pp. 117-127) calls "item-based constructions". When observing his daughter, Tomasello had realized that: « almost all ... multi-word utterances during her second year of life revolved around the specific verbs or predicative terms involved. This was referred to as the Verb Island hypothesis since each verb seemed like its own island of organization in an otherwise unorganized language system...Within any given verb's development there was great continuity such that new uses of a given verb almost always replicated previous uses and then made one small addition or modification.» (Tomasello, 2009)

Of islands of order

Other experiments have indicated the validity of the claim that the stage of "pivot schemas" naturally develops into a stage of such "constructional islands" of productive order. For example, the study of Pine and Lieven (1997) had shown that children between 1 and 3 years of age tended to use the determiner "the" juxtaposed with one set of verbs and determinet "a" juxtaposed with another, with rare overlap between the sets. In a parallel study conducted with the same group of 12 toddlers, the same authors have observed that 91.6% of first 400 distinct utterances could be "traced back" to only 25 initial patterns Lieven et al. (1997). Since many results of these studies are English-specific (e.g. the importance of prototypical constructions like "want+X", "verb+it" etc.), we consider it important to emphasize those conclusions of these authors which seem to point towards more "universal" a direction: « Our metaphor would be of language developing initially as a number of different islands of organization which gradually link up...These islands are initially segments (either words or phrases) which the child has identified to the extent that she can start analysing other systematic relations between what comes before, after or within them...We think, rather, that the data can support a view of structure as emergent.» (Lieven et al., 1997)

Based on research of Lieven and her colleagues as well as on his own (Tomasello, 2009, p.308) lists three basic operations by means of which a child can produce an utterance:

1. retrieval of a rote-learned concrete expression and the repetition of the same form as was already heard
2. retrieval of an utterance-level construction and its modification in order to fit the current situation
3. « combining constituent schemas» (Tomasello, 2009)

Note that the first operation can be aligned with notion of "imitation" and thus "replication of information", the third can be interpreted as a "cross-over" and the second is - so states our Thesis - equivalent to what universal darwinists call "variation operators". Tomasello lists three principal means of structural modification:

Of operators of morphosyntactic variation

1. extension: concatenation of the constituent to the end or beginning of an expression (e.g. ich auch + Yoga -> ich auch Yoga)
2. injection: « inserting a new constituent into the middle of an utterance-level construction or expression (the way a German child might insert auch[11] [too] into a schema where nothing had ever before appeared» (Tomasello, 2009)
3. slot-filling: inserting new content into a slot in the item-based construction (e.g. Brot + essen X -> Brot essen)

It is evident that such "slots" are, in fact, categories and they are denoted by what is in formal linguistics 10.2 called non-terminal symbols. As category-representing symbols, they are undoubtably a consequence of a category-construction (CC) process[12]. In the long run, the output of CC process should be a set of categories which are functionally equivalent to categories shared by, and inherent to, child's social surroundings. But what was already said about lexical and semantic categories hold also, mutatis mutandi, for the grammatical ones: before the gap between the ambient and the individual is bridged, before the structure of partitions inherent to the latter converges to the structure isomorph to the former, discrepancies between the two systems are to be observed.

Most salient and best studied among such discrepancies is group of phenomena labeled as "over-regularization" . Traditionally, over-regularization is supposed to account for cases whenever the child applies the production rule beyond the scope of its validity. The most

Of over-regularization

11 C.f. 12.7.2 for closer discussion of productivity of "auch" in case of one specific child.

12 We prefer to speak about CC and not simply about "categorization" to mark the distinction between the process by means of which a category is built, and the process during which an already built category is used in order to "categorize" diverse blobs of stimuli observable in the world

famous example of overregularization in English is that in certain stage of their development, practically all children tend to apply the rule Vpast $\rightarrow$ VPresent+ed on all verbs. Thus, especially during period when their mean length of utterance (MLU) is cca 4-5 words, do child generate past participles like « throwed » or «braked» which they had never (or very rarely) heard. Another interesting aspect of over-regularization is that often, children used the correct forms BEFORE they start producing incorrect over-regularizations: « Initially, children's uses of -ed past tense are all accurate. They may say melted or dropped, but not, as they later do, runned and breaked.» (Maratsos, 1988)

Sooner or later -but often sooner than later- are practically all grammatical over-regularizing discrepancies corrected and child's linguistic behaviour aligned with that of her surroundings. It is difficult to describe this fact without taking for granted "the principle of precedence of the specific":

The principle of precedence of the specific (DEF)

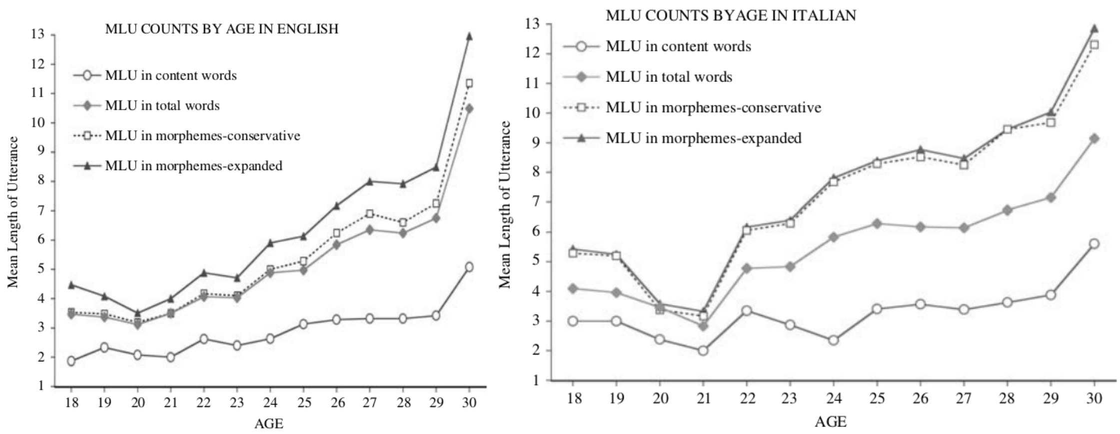
« Whenever a newly acquired specific rule (i.e. a rule that mentions a specific lexical item) is *in conflict* with previously learned general rule (i.e. a rule that would apply to that lexical item but also to many others of the same class), the specific rule eventually takes precedence.» (Braine, 1971)

END PRINCIPLE OF PRECEDENCE 9.2

This principle is defined her in terms of "rules". But as shall be seen in 9.4, the notion of "rule" is crucial only for certain - and not all- theories of language and LD. Thus, too much focus on notion of "rule" can turn out to be misleading, moreso in this section of our *expose* where our objective has been to focus more on empirical and less on theoretical considerations of process of development of individual morphosyntactic representations.

The body of empiric research which have explored this or that facet of language acquisition, is indeed vaste. For example, even a simplistic synthesis concerning the developmental, cross-linguistic or clinical aspects of MLU would easily account for a monography thick as a brick. But in this Thesis we cannot dedicate to this topic more than space than that which is dedicated to 15. Idem for other fascinating LD-related topics like cue competition (MacWhinney, 1987) in both comprehension and production, or acquisition of verbal skills related to negations, questions or word order: all these problems, and many others, are simply too specific to be addressed appropriately there, were only the most general principles of LD are sought to be addressed.

Figure 15: Mean length of utterances produced by English and Italian children of different age (in months) . Figures reproduced from Devescovi et al. (2005).



(a) (b)

Of importance of input

However, what should be addressed and re-addressed, emphasized and re-emphasized is the importance of linguistic input. This is so, because both content and distribution of linguistic input significantly influence the content and distribution of resulting representations and structures. This constatation may seem trivial, but is less so when one realizes "how special" the content of child-directed input and its distribution is. Since the content shall be more closely discussed in secion 9.3, let's now end this brief overview of development of child's LD with the question:

Y a-t-il some particular distributional, statistical, computational property of linguistic input which facilitates the internalization of morphosyntactic representations?

And the answer seem to be: yes there is, and the seems to be somehow related to the fact that acquisition of linguistic representations is governed, similary to many other cognitive functions, by the principle of distributed practice.

Principle of distributed practice (DEF)

« Given an equal number of exposures, distributed (or spaced) practice at a skill is almost always superior to massed practice.» (Tomasello, 2009) END PRINCIPLE OF DISTRIBUTED PRACTICE 9.2

In other words, humans in general and children in particular internalize better when they are confronted with the structure-to-be-internalized within contexts of N different sessions (ideally on different days), and worse when they are confronted with it N times during the same session (on the same day). In relation to LD, this phenomen was first noticed in study by Schwartz and Terrell (1983) who noticed that both group of 1-3 year old children who heard the new word once par session and group of children who have heard it twice per session, have, in fact, both needed approximately 6-8 sessions to learn it. Thus, « when the absolute number of presentations was held constant, distributed (infrequent) presentations led to greater acquisition than massed (frequent) presentations.» (Schwartz and Terrell, 1983)

Of distributed practice

Similar results were subsequently obtained in studies of acquisition of grammatical constructions. For example, Ambridge et al. (2006) conclude their study: « for grammatical constructions, children are more able to analogize across exemplars and extract a relational schema when those exemplars are more widely distributed in time than when they are temporally contiguous» (Ambridge et al., 2006). And since the possibility that something like "principle of distributed practice" exerces its force not only in case of acquisition of human verbal behaviour but also in development and optimization of other cognitive functions and skills, the same authors conclude: « a single set of general learning and cognitive processes is responsible for the acquisition of both individual lexical items (the lexicon) and regular and irregular grammatical constructions (the grammar)» (Ambridge et al., 2006).

Agreeing with such conclusions of which Piaget would be undoubtably quite fond of, we terminate this section with expression of our belief that no matter whether taking place in the lexical or morphosyntactic domain, we consider such processes to be principially based on iterative, gradual and non-deterministic optimization of populations of internal representations which replicate, vary and are subjects of selection. A belief, which we shall try to defend in what shall follow.

END ONTOGENY OF MORPHOSYNTAX 9.2

9.2.5 ONTOGENY OF PRAGMATICS

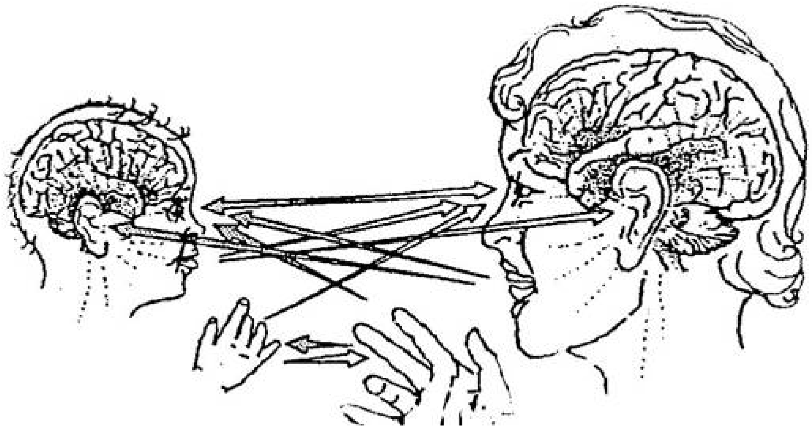
Of pragmatics

Pragmatics goes hand in hand with practice. In linguistics, pragmatics is all that is somehow involved in production or comprehension of utterance but is not contained within the utterance itself. Thus, pragmatics is all that encompasses and envelops the communicative act; pragmatic layer contains all the context within which the utterance is exchanged.

Of social context

It was already stated that language is a social entreprise and the context within which natural language utterances are exchanged is thus principially a social context. In such context, multiple human agents are in mutual interaction and exchange of sequences of linguistic symbols is only one among many other ways how these interactors modify each other's mental states. Other important channels of communication between two prototypical human subjects -mother and a child- are illustrated on 16

Figure 16: Some modalities of information exchange between mother and her child. Reproduced from Trevarthen (1993).



But the extralinguistic context is not limited to facial expression and gests. Nor introduction of olphactoric (pheromonal) or haptic communication would make the notion of "context" complete. For the context *par excellence* is given by the very spacetime region within which the linguistic exchange takes place, the region which contains specific physical objects or embodies certain processes.

Of spatiotemporal context

Somewhat contrary to what is displayed on 16, linguistic communication is rarely dyadic. Much more often it refers to object or state-to-be-attained which is external to both members of the interacting couple. It is rarely by chance that two humans encounter each other: more often they go towards each other because they want to be with each other. Even if the object of such wanting can be a simple "being with the Other".

What's more, both interactors have mental states and they both have intentions. And to make things even more complex, they use language in order to mutually modify their mental states. They use language in order to augment the probability that their intention shall be materialized. With the help, and through the act, of the Other.

Of intersubjective context

Thus, an infant whose cry/not cry signal emittor is not appreciated anymore has to change her strategies. She has to learn what formulas work best in what contexts, what should be said and when, where, in what order and how it should be said so that the mental states of the Other should be modified appropriately. A Herculean task extending well beyond childhood and puberty towards adolescence and beyond: pragmatic knowledge seems to be acquired from the very first until the very last breath of one's ontogeny. Having many forms -from cry of a newborn to wisdom of and old man; from benevolent lies to ma-

Of most difficult a task

nipulative propaganda- acquisition of pragmatic knowledge seems to be too difficult nut to crack.

This awareness of diverse intricacies and complexities of the pragmatic layer as well as the respect in front of maximas (Grice, 1975) and values which are its foundation; and in agreement with the principle « Whereof one cannot speak, thereof one must be silent.» (Wittgenstein, 1922), we decide not even trying to discuss computational aspects of pragmatics-related phenomena.

END ONTOGENY OF PRAGMATICS 9.2

9.2.6 PHYSIOLOGICAL AND COGNITIVE DEVELOPMENT

One cannot speak about early development and ignore the vast amount of physiological and cognitive changes which children undergo. Between the birth and the end of toddler stage, height of children's body almost doubles and both the weight and lung volume more than triple. Muscles strenghten, bones ossify. Fontanelles close, thus envelopping the brain within the fully enclosed resonator called skull. Primordial reflexes appear and disappear.

Of ontogeny of body

In first 12 months only does the average brain volume increases from 369 cubic centimeters to 961 cc. This increase, however, is not to be explained in terms of increase in quantity of neurons (gray matter) but in terms of increase of glial cells. In context of what was already said about Neural Darwinism (8.6), we consider it important to underline that during development, the number of neurons in fact, decrease due to the process known as "synaptic pruning".

Of ontogeny of brain

It is also during the early year of development when the linguistic faculty gets entrenched in the specific hemisphere of the brain. While there is still an ongoing debate concerning diverse aspect of such processus of "lateralization" (see Clark, 2003, pp. 387-391 for overview), it is nonetheless commonly accepted that the hemisphere of installation of developing linguistic faculty is determined in first 20 months of age.

Of hemisphere lateralization

Among hundreds of other neurological and physiological changes which shall occur with apodictic necessity in any healthy human child, there are three which we consider to be particulary important for the linguistic development yet thtacitly undiscussedd by psychodevelopmental linguists.

First is related to a relatively trivial fact that in comparison to other primates, human teeth erupt very late (Holly Smith et al., 1994). This, on one hand, allows for much longer breast-feeding and hence temporally reinforced emotional and social bonding between the mother and the child while, on the other hand, makes it impossible for a child to articulate sounds with dentals or alveodental acoustic fea-

Of teeth eruption

tures. Child's ability to correctly generate language of her surrounding is not only cognitively but also physiologically limited.

The second physiological change to which we would like to point attention in context of ontogeny of human linguistic faculty is related to rythmical behaviour. As suggested by practically whole tradition of research dedicated to psychology of rythm at least from Fraisse (1974) to Provasi et al. (2014), every human can be characterized in every stage of his development by a so-called Spontaneous Motor Tempo (SMT). By being both the tempo in which people tap when asked "Please tap on the table with Your hand in the most natural speed" and, also, the tempo which people choose as most natural when asked to choose between multiple tapping sequences, SMT seems to be a fundamental cognitive phenomenon integrating both passive (perceptive or even C-structure) and active (productive or even P-structure) components.

Of tuning of the internal oscilator

In regards to ontogeny, it's worth to underline that SMT tends to slow down with age, which can help children to maturate « from facile acquisition of relatively brief events, such as phonetic categories, to enhanced proficiency with longer events» (McAuley et al., 2006). It is also only with age that children acquire faculty to process and generate rythmic patterns wider range of tempos, id est tempos which significantly differ from the SMT of their endogenous tact-giving "oscillator".

Given the importance of tempo for control of repetitive or oscillatory activity including not only language but also walking - which, coincidentally or not, appears approximately in the same period when children leave the phase of canonical babbling and enter the phase of word productions - we believe that the study of SMT and other rhytm-related phenomena could by useful for any further study of human cognition. Within the scope of the current Thesis, however, shall these phenomena serve only a peripheral role.

Of sleeping

At last but not least, the third non-negligeable development occurring in early childhood is related to changes in length, distribution and composition of sleep *cycles*. Thus, « the newborn infant spends two-thirds of each 24-h period asleep; by 6 months he spends half of his time asleep and half of his time awake. Sleep consolidation is another important aspect of infant's sleep development. By the age of 6 months sleep condensed into fewer periods of longer duration so that sleep periods are lengthened from 4 to 6 h.» (Gertner et al., 2002)

But not only do new-borns sleep significantly more than toddlers which sleep significantly more than older children and adults; not only they sleep not in one nocturnal block as adults do but in multiple blocks, both diurnal and nocturnal, but - and this is an important "but" - they spend significantly more time in dreaming "rapid eye movement" (REM) phase than they shall ever do in the future: « REM sleep assumes a high proportion of total sleep in the first days of life and its amount and ratio diminish as maturation proceeds» (Roffwarg et al., 1966).

Beautiful line of research studying pre-sleep "crib talk" monologues (Nelson, 2006) aside, the relation of LD and sleep has not yet been studied in an extent it merits. Note, however, the words with which author of few among very few experiments studying the impact of sleep upon processing of linguistic stimul concludes her results: « memory consolidation associated with sleep introduces flexibility into learning, such that infants recognize a pattern at test regardless of whether it is instantiated exactly as it was before. Sleep then sustains the learning of previously encountered information in a form that enables children to generalize to similar but not identical cases, and it also introduces flexibility into learning» (Gómez, 2011).

Of memory consolidation

Consistently with such conclusions, we end this short *expose* with a statement that from the point of view of theory of intramental evolution of linguistic representations which we hereby aim to introduce, such "memory consolidation" ocurrent during sleep could be interpreted in terms of activity of mutation and cross-over operators acting upon and mixing already encoded structures.

END PHYSIOLOGICAL AND COGNITIVE DEVELOPMENT 9.2

9.3 MOTHERESE

Child's closest social environment are her parents, most notably her mother. Hundreds of studies were conducted to study the nature of «motherese», a special simplified child-directed which mothers use when speaking with their children. While some studies point in divergent directions, they more or less agree that « maternal speech has certain characteristics that distinguish it from speech to other adults. These characteristics are in essence simplicity, brevity and redundancy» (Harris, 2013).

Of basic features of motherese

Other characteristics generally associated with child-directed speech are:

1. uses higher pitch (mean fundamental frequency of speech to 2-year olds is cca 267Hz and 198 Hz in case of speaking to adults)

2. exaggerated intonation (wider pitch range)

3. slower speech due to both more and longer pauses

4. contains repetitions and variation sets (9.3)

Clark (2003) summarizes the properties of child-directed speech as follows: « adults consistently produce shorter utterances to younger addressees, pause at the ends of their utterances around 90% of the time (50% in speech to adults), speak much more fluently, and frequently repeat whole phrases and utterances when they talk to younger children. They also use higher than normal pitch to infants and young children, and they exaggerate the intonation contours so that the rises and falls are steeper over a larger range (up to one-and-a-half octaves in English).» (Clark, 2003)

Multiple studies indicate an existence of causal link between the quantity and simplicity of motherese utterances and speed of child's linguistic development. More concretely, it had been observed that « mothers' choice of simple constructions facilitated language growth» (Furrow et al., 1979), while more complex style can slow their development down. Other studies precise that « children who showed the earliest and most rapid language development received significantly more acknowledgments, corrections, prohibitions and instructions from their parents» (Ellis and Wells, 1980). Other means how LD can be stimulated are variation sets.

Variation sets

Variation set is composed of two or multiple subsequent utterances which are all derived from one common item-based construction. « Variation sets are identified by three types of phenomena: (1) lexical substitution and rephrasing, (2) addition and deletion of specific referential terms, and (3) reordering of constituents.» (Küntay and Slobin, 2002)

A most simple form of VS is a sequence of utterances $U_1 ... U_x$ sharing the same word W serving as a link between subsequent utterances.

```
MOT we lost that piggy bit .
MOT so that bit goes there .
MOT and (.) we've lost that horsie bit which is a bit of a pain .
MOT that bit goes there .
MOT and that bit there .
MOT shall we try and find the lost bits ?
MOT found two bits .
```

Just a little bit more complex are VS where "linking" between $U_1 ... U_x$ is done first with one word (or construction) W_1 which, in certain moment co-occurs with another word (or construction) W_2 which subsequently "links" to following sentences. In an illustratory example extracted from data/Eng-UK/Lara/1-11-27.30.cha transcript of CHILDES corpus (MacWhinney, 2014) do words "bit" and "that" fulfill such a fixing role.

More complex variation sets involve slight variation of longer expressions. For example, the transcript data/Eng-UK/Lara/1-11-27.30.cha of the same mother-daughter couple taken 7 months later, contains a following VS:

```
MOT you have to sing to her , Lara , if you want her to go to
    sleep
MOT sing her a lullaby
MOT you have to sing her a song
```

In this VS, it is a whole expression "sing something to someone" which is varied, first by removal of a non-obligatory dativ marker "to" and subsequently by variation of the object of singing, from lullaby to a song.

Many researchers argue that exposure to variation sets can facilitate acquisition of both semantic and syntactic categories and/or rules. For example, Küntay and Slobin (1996) have observed that

1. use of variation sets is positively correlated with child's acquisition of certain verbs
2. VS make up cca 20% of child-directed speech.

Similar observation, i.e. that 1/5th of child-directed speech are variation sets, was obtained by Brodsky et al. (2007). In a study involving the analysis of CHILDES corpus, the authors explain the advantages of VS in information-theoretic terms: « variation sets seem to be ideal environments for learning lexical items and constituent structures...a pair of utterances that have nothing in common is not informative, and neither is a pair of identical utterances. An optimally informative pair would therefore balance between overlap and change.» (Brodsky et al., 2007)

Note that the notion of «variation set» can be intepreted in UD-consistent terms, given that:

1. repetition is equivalent to «replication in time» and every single instance of the utterance can be therefore considered as an independent,individual structure
2. alteration of form between subsequent utterances can be interpreted as a consequence of variation operator influencing production of new sentences

To illustrate the extent of some variation sets, 1st appendix (**??**) contains the longest variation game discovered in, and extracted from, the CHILDES corpus.

END VARIATION SETS 9.3

Studies like those of Harris (2013) suggest that there the releation between the complexity of motherese and complexity of child's own production is in fact reciprocal. Thus, mothers adjust their language according to the stage of child's linguistic development.

In context of current Thesis, we propose to interpret the mutual *getting closer* of motherese and toddlerese in terms of adaptation and co-evolution of two populations of linguistic structures. Mother adapts to toddlerese of her child, toddler adapts to motherese: both co-evolve.

In the long run it is the adult who leads the dance. This is so because internalization of the language of the Other is in child's and not parent's vital interest. Thus, child's P-structures which shall correctly modify the adult's behaviour in an intended direction could be considered more fit and thus more prone to intramental replication than P-structures which shall not yield an intended effect.

END CHILD-DIRECTED SPEECH 9.3

9.4 LANGUAGE ACQUISITION PARADIGMS

In practically no modern scientific discipline is the age-lasting trialectics between realists, nominalists and idealists so ardent as in psycholinguistics. Different perspectives and terminology being adopted, it is nonetheless the eternal "problem of universals" which is being targeted. Travested into rationalists, mentalists or nativists, one group does its best to convince the public that intangible "general" is prior to the "specific"; in the other camp, the empiricists battle for their belief that the observable and specific is prior to general.

In course of centuries do hours of hand-waving, ink-spilling and dozens of metaphysical chimeres accompany the process whose unfolding is supposed to bring scientific community ever closer to "most fit" narrative about origins and development of language in onto-, phylo- or even cosmo- (De Chardin et al., 1965) genesis. Let's now glance on few among its most distinctive figures:

9.4.1 CLASSICAL

First observations

One among first tentatives to describe the process of language acquisition was done by Saint Augustine in his Confessions: « Passing hence from infancy, I came to boyhood, or rather it came to me, displacing infancy. Nor did that depart,- (for whither went it?)- and yet it was no more. For I was no longer a speechless infant, but a speaking boy. This I remember; and have since observed how I learned to speak. It was not that my elders taught me words (as, soon after, other learning) in any set method; but I, longing by cries and broken accents and various motions of my limbs to express my thoughts, that so I might have my will, and yet unable to express all I willed, or to whom I willed, did myself, by the understanding which Thou, my God, gavest me, practise the sounds in my memory. When they named any thing, and as they spoke turned towards it, I saw and remembered that they called what they would point out by the name they uttered. And that they meant this thing and no other was plain from the motion of their body, the natural language, as it were, of all nations, expressed by the countenance, glances of the eye, gestures of the limbs, and tones of the voice, indicating the affections of the mind, as it pursues, possesses, rejects, or shuns. And thus by constantly hearing words, as they occurred in various sentences, I collected gradually for what they stood; and having broken in my mouth to these signs, I thereby gave utterance to my will.» (Augustine, 1838)

Not being a theory *per se*, Augustine's naive but honest reflexion concerning the origins of his own mental faculties had nonetheless a non-negligeable impact upon all theories of LD which had followed. From our current perspective it, this Augustine's *Confessio* could be most probably labeled as a precursor of associationist school. This is so, because Augustine principially explains the ontogeny of his semiotic faculty in terms of associations between "words" and "things".

Associationism was among one of those very rare prototheories of cognition which have succeeded to survive one and half millenium which had followed. David Hume, John Locke or J.S. Mill had adhered, in one way or another, into the camp of those who were convinced that great deal of mental phaenomena -or possibly ALL phaenomena- could be principially explained in terms of mind's tendency to "link" its internally represented "signs" with things-in-world or other "signs". Thus, thanks to its usefulness in learning and evidence in introspection, associationist had succeeded to cross the centuries to see the day when the neurologist Donald Hebb postulated the material (i.e. neural) basis of what was before considered to be solely mental phaenomena:

« When one cell repeatedly assists in firing another, the axon of the first cell develops synaptic knobs (or enlarges them if they already exist) in contact with the soma of the second cell.» (Hebb, 1964)

Since the discovery of the phenomenon which is often referred to by saying "cells that fire together, wire together; neurons that fire out of sync fail to link" the Hebb's rule helped to yield explication of such neuroscientific challenges as "emergence of mirror neurons"[13] Keysers and Perrett (2004). The core idea behind many functional artificial neural network architectures - e.g. Hopfield networks Hopfield (1982)- is also essentially Hebbian. In 10.4.2 shall be the Hebb's postulate mentioned as a potential explanation of validity of a so-called "distributional hypothesis" which is *the* idea behind practically all efficient computational models of semantic vector spaces.

Behaviorism is another school of thought which was derived from associationist school and in which the Hebb's can play a role of the

13 It was already indicated, during our discussion of memetic theory, that mirror neurons are often mentioned in relation with imitation. But in theory hereby introduced, they can also be understood as neural substrate for bridging C-representations with P-representations.

most fundamental principle. For in its very essence, behaviorism simply substituted the notion of association between two signs with the notion of conditioning. By adopting the terminology of stimuli and reflexes, of rewards and punishement; and by renouncing to any methods which were not strictly positivist and empiric, the behaviorist school had renounced to any tentatives to understand internals of the mind. Since behaviorist precepts worked -and worked not only when applied on Pavlov's dogs or Skinner's pigeons but on humans as well- and because science lacked both computers and subtle experimental neuroimaging apparati, the tentatives to explain man's mind in terms of reinforcement of relations between stimuli and reflexes was the principal pre-occupation of western psychology of 1st half of 20th century.

And it would have possibly dominated until now if ever the central figure of the field, B.S. Skinner, hadn't decided to apply the behaviorist doctrine into the domain of linguistics. Skinner's book Verbal Behavior Skinner (1957) claimed that language is learned by operant conditioning, id est, that child learns language because the expressions of her verbal behaviour - the fact that she utters X and not Y - are reinforced by parental rewards. For example:

« In all verbal behavior under stimulus control there are three important events to be taken into account: a stimilus, a response and a reinforcement. These are contingent upon each other... The three term contingency ... is exemplified when, in the presence of a doll, a child frequently achieves some sort of generalised reinforcement by saying doll.» (Skinner, 1957)

In Skinner's theory, vectors of reinforcement are not necessarily just dolls, milk and breasts but can be quite abstract: the very parental attention can be rewarding, lack of it may punish. By focusing on child's and parent's basal needs, wants and behaviours with which they attain them, Skinner articulated a theory which has some overlaps with current interactionist and sociopragmatic theories of LD 9.4.4. The idea of founding the fitness functions of our grammar induction systems not only on measures internal to the system, but also on environment's responses to the system (14.4.1) is also traceable back to similar behaviorist point of view.

9.4.2 GENERATIVISTS AND NATIVISTS

Then came Chomsky. In his revolutionary Syntactic Structures (Chomsky, 1957) he proposed to adopt a rule-based, algebraic, transformationalist approach to explain the mystery of grammars able to generate infinite number of utterances out of finite sets of elements. Two years later, young Noam had gained in prominence by overt and uncompromising review (Chomsky, 1959) of aging Skinner's Verbal Behaviour. More carpet bombing than review-like, this critique - in some

Of Chomsky and Skinner

circles considered as the most influential rhetoric exercice of 20th century - irreversibly hallmarked the rupture: the turn from "behaviorist" to "cognitive" approach to study of language acquisition.

Whether one rightfully praises Chomsky like (Gardner, 1985b), or rightfully criticize him as Tomasello (2009), one has to admit that he was one among the first who attempted to interpret linguistic representations and processes as fundamentally computational phenomena. Thus, the surface structure of an utterance was to be understood as an output of series of substitutional rules acting upon a certain deep structures offered as an input. While the notion of substitution rules was already known to sanskrit scholar Panini more than two millenia before Chomsky and was in 19th century practically deified by *neogrammarians* who spent a non-negligeable effort to use the notion of universally applicable rule to explain the mystery of historic language change (8.5), it was nonetheless Chomsky who, strongly influenced by his predecessors Jakobson and Harris, developed a theory whereby substitution rules are supposed to be acting also in individual human cognitive systems.

Of generativism

Panini's Grammar (APH)

Panini's (cca 400 BC) grammar is the oldest attested work in descriptive linguistics. Composed at the end of Vedic period and at the beginning of Classical period, it contains 3996 rules of Sanskrit morphosyntax and, in lesser extent, also of semantics. It was transfered orally -from masters to their students in myriads of schools spreadout through whole India- in form of sutras, i.e. verses to be memorized.

Of first rule-based grammar

Grammar begins with Shiva Sutras which enumeration of 14 fundamental phonological classes from which one can generate 281 pratyāhāras, i.e. classes of second order which are to be subsequently processed by application of one or more among 4000 "Ashtadhyayi" substituion rules and meta-rules. In terms of modern linguistics these 14 sutras list all sanskrit terminals phonemes (16 vowels and 33 consonants) and associate them with anubandha labels (non-terminals). PERL-consistent transcription of Shiva sutras follow: every line presents one sutra in form of a substitution rule. Parantheses contain individual phonemes; symbol enclosed between second and third / symbol denotes the anubandha.

Shiva Sutras (SRC)

```
s/ ( a | i | u ) / ṇ /
s/ ( Ṛ | ḷ ) / K /
s/ ( e | o ) / ṅ /
s/ ( ai | au ) / c /
s/ ( ha | ya | va | ra ) / ṭ /
s/ la / ṇ /
```

```
s/ ( ña | ma | ṅa | ṇa | na ) / M /
s/ ( jha | bha ) / ñ /
s/ ( gha | ḍha | dha ) / ṣ /
s/ ( ja | ba | ga | ḍa | da ) / ś /
s/ ( kha | pha | cha | ṭha | tha | ca ṭ|a | ta ) / V /
s/ (ka | pa ) / Y /
s/ (śa | ṣa | sa ) / R /
s/ ha / L /
```

END SHIVA SUTRAS (SRC) 9.4.2.0

Description of the means by which 281 (14*3 + 13*2 + 12*2 + 11*2 + 10*4 + 9*1 + 8*5 + 7*2 + 6*3 * 5*5 + 4*8 + 3*2 + 2*3 +1*1 - 14 - 10) pratyaharas are to generated from the list of classes listed above, and description of almost 4000 rules (e.g. *vṛddhir ādaiC*) which subsequently allow the production of so-is-believed, of whole corpus of one amongst the most complex languages ever known to man, all that surpasses the objectives of this dissertation.

What does not surpass it, however, is the question: "How could Panini (or any lineage which preceeded Panini) ever discover such a grammar?". Computationally, the task is enormous. Revelatory explanations - so popular in India - aside, we see only one possible answer: by means of intramental evolution.

END PANINI'S GRAMMAR 9.4

Not only Panini, Jakobson, Harris have influenced Chomsky, but also Turing whose idea of a symbol-substituting machine crossed over with discovery of transistor, thus yielding new and powerful generation of computers around the time as Chomsky entered MIT. And it was in and by and through contact with computers that Chomsky understood the generativity of rule which is applied many times and which can consider its own past outputs as its present or future inputs. Recursivity once understood, followed an insight that recursivity is present in natural languages, e.g. in expressions like:

Of recursivity

She knows that he knows that he knows that she knows...

Et caetera et caetera, theoretically ad infinitum. For an old-school generativist, the very theoretical possibility to realize such an *infinite regress* constitutes a sort of a proof of belief that grammars, understood as systems of rules which can be recursively applied upon sequences of symbols chosen from a finite alphabet, can ultimately generate infinite amounts of such sequences. [14].

14 C.f. the "Halting problem" in (Hromada, 2008) for closer theoretical discussion of this "infinitist" fallacy which is, we believe, the source of many problems which haunt the generative linguistics from the very moment of its conception.

That recursion, combined with substitution, can simulate and/or generate *more than anything*, was already known to Cantor and Godel, let along Turing. But contrary to Godelian proving-god-through-arithmetics and Turing's Enigma-breaking, Chomsky had decided to program the "universal machine" with a goal in mind which his audience could understand: language. To make things formal and contrary to centuries of knowledge which says otherwise, language was subsequently reduced to a set of sequences of symbols chosen from the alphabet (10.2). Other definitions, axioms and theorems had followed, often with huge importance for subsequent development of evermore complex assemblers, parsers and compilers of artificial languages. For example, it is difficult to imagine how informatics could move from assembler to C++ or PERL without having at its disposal the theoretical framework of Chomsky-Schützenberger containment hierarchy of formal grammars.

Of Chomsky's contribution to computer science

Thus, a formal system was developed which turned out to be useful for certain subdiscplines of informatic science. But as is often the destiny of formal systems which fail to delimit their domain of applicability, its proponents started to confound the map with the territory. As a result, practically whole generation of linguists following the "discovery of generativity of recursivity" got lost in the labyrinth of futile tentatives trying to fit the expressive diversity of the natural into the monolithic framework able to account only for simple among the artificial. Thus, in a somewhat paradoxical turn of events, were thousands of linguists transformed into pigeons pecking X-bar conditioned by the reinforcement principle of "publish or perish". New models and theories with names as binding as "Government and Binding" or as non-minimalist as "Minimalist program" were proposed and turned into full-fledged doctrines shared among the castes of initiates. To give at least some meaning to the evermore ezoteric symbol-substituting *passe-temps*, a noble quest was launched: the quest for a so-called Universal Grammar.

Of operant-conditioning of linguists

In its very essence, the notion of Universal Grammar (UG) is Chomsky's answer to the problem raised by Gold and related to the fact that from one specific language sample, one can induce multiple grammars which are able to generate such sample. But if multiple grammars can be obtained, how could a child now which one is the correct one? Chomsky's answer was: because her choice is constrained by "something" innate: her majesty the UG[15]

Of Universal Grammar

One can explain "innate" either in creationist or emergentist terms. Hoping that nativists do not belong to the first group, there is only one way how existence of UG could be explained: by evolutionary

15 Note that there is a certain symmetry between couples (Turing Machine, Universal Turing Machine) and (Grammar, Universal Grammar). In both couples does the unicity of the latter furnish a frame for the diversity of the former. But there is a difference as well: while UTM *allows* to emulate any Turing Machine, UG is supposed to *constrain* the set of relevant grammars.

Of reconciliation

process. This is, we believe, a point of reconciliation, a point of convergence between non-creationist yet nativist doctrines which postulate UG, and theory of intramental evolution as hereby introduced. On the other hand, there is also a significant point of divergence: while we suggest that it is more computationally feasible to produce language-generating and language-constraining representations during ontogeny, Chomsky's "nativist" disciples like Pinker (1994) spent significant part of their carreer arguing for the fact that UG is somehow produced by evolution which is phylogenetic. Simply stated: nativists believe that UG is encoded in DNA.

As was already indicated, the *raison d'etre* of UG is to constraint and direct learning of grammar. The problem that English children learn grammar of english and Chinese learn grammar of chinese was "solved" by a so-called Principles & Parameters theory as follows: during acquisition of language L_X, child extracts from the utterances she hears a set P_X of parameters specific to L_X, inserts these parameters P_X into the UG, thus obtaining the specific grammar G_X able to generate L_X. Id est:

$$G_X = UG(P_X)$$

Of power of stimulus

According to nativists, a child would not be able to learn G_X without UG's intervention in the acquisition process. The necessity is supposed to be both empiric, and theoretic. The empiric necessity is related to the problem of "poverty of stimulus" that the utterances children here are qualitatively incorrect and quantitatively insufficient to account for the fact that child, indeed, learns language of its environment. Unfortunately for nativisits, vaste body of rigorous empiric research (Clark, 2003; Karmiloff and Karmiloff-Smith, 2009; Tomasello, 2009) in DP indicates that the notion of "poverty of stimulus" was nothing else than a chimere and that reality of a healthy child surrounded by a healthy social environment is rather the contrary: one should not speak about poverty, but rather about "power of stimulus". A scientist who agrees with the statement « in summary, child-directed speech and other sources of language – overheard speech, stories read aloud, speakers heard on radio or TV, for instance – provide such rich input that children should eventually learn enough of their language for all their needs» (Clark, 2003) can thus discount the poverty of stimulus as simply irrelevant.

On theoretical grounds, the necessity to have something like UG embedded in human Language Acquisition Device (LAD) is often claimed as a necessary consequence of the Gold's Theorem, an important result obtained in a "learnability theory" sub-branch of formal language theory.

Refutation of Gold's Theorem (APH)

In short, the theorem postulated by Gold (1967) states that « Any class of languages with the Gold Property is unlearnable.» (Johnson, 2004) In Gold's formal system, class C of languages has the Gold Property if and only if:

1. C contains a countable infinity of languages L_i such that $L_i \subset L_{i+1}$ for all i>0
2. a further language L_∞ such that for any i>0, x is a sentence of L_i only if x is a sentence of L_∞ , and x is a sentence of L_∞ only if x is a sentence of L_j , for some j>0

Learnability, on the other hand, is defined in terms of

1. environment E, which is supposed to be an infinite sequence of sentences of the language to be learned
2. an ideal learner, which « learns L given E iff there is some time t_n such that at t_n and all times afterward, the learner correctly guesses that L is the target language present in the environment» (Johnson, 2004)

Gold's Theorem therefore simply states that infinite [16] set of mutually embedded languages are «unlearnable» by a system which never forgets and which is exposed to infinite environment. It is difficult to see, however, how these purely theoretical relation between purely theoretical infinite sets can have any further implication for concrete individual languages, understood as finite sets of utterances exchanged in specific extralinguistic contexts. Given that lifetime of an individual human learner is always finite, the linguistic environment of such a learner also has to be finite. The first condition of what "learnability" means in Gold's formal system is thus irelevant to human learners.

Of finiteness of environment of human learner

The second condition is irrelevant as well: human mind is not a storage system which faitfully internalize and store for "all time afterwards" every piece of linguistic data to which it was exposed, unmodified. Humans forget, children moreso. It is therefore somewhat unclear how the very notion of learnability and unlearnability, as defined by Gold, could apply to human beings in general and children in particular.

Of forgetting

This being said, we find it appropriate to state that more than an important proof telling us something about LD in human beings, whole fuzz about Gold's Theorem is more an evidence of how a whole a multidisciplinary scientific endeavour can get stuck for decades in a blind alley just because of an inter-disciplinary quaternio termino-

Of fallacy of four terms

16 Taking Gold's Theorem seriously in regards to language learning is equivalent, mutatis mutandi, to belief that children shall never learn basic arithmetics because in order to understand addition, they would first have to be confronted with all integers between one and infinity.

rum (Sokol, 1998) fallacy. In other words the term «unlearnable» in Gold Theorem is just a term which Gold uses within his tautological statement to denote certain properties of certain infinite hierarchically embedded sets of sequences of symbols, and as such does relate only abstractly to concrete condition of human learners.

END REFUTATION OF GOLD'S THEOREM 9.4.2.0

A theoretical pillar of necessity to postulate UG being somewhat undermined by the above aphorism, there are not many reasons for not using the lex parsimoniae of William of Occam to raze the notion of DNA-encoded UG away from the terminological toolbox of 21st century linguistics. This does not mean that there are no faculties and features which would be universally present among all human languages[17]. Take for example the fact that in all human cultures, people group consonants and vowels into specific clusters. Given the universality of the phenomenon, one is more than tempted to agree with authors like Jackendoff (2002) (a nativist of second generation) and state that syllabization is a component of UG. But when one realizes that syllabization is potentially just a consequence of concrete application of a deeper cognitive processus, known as "chunking", upon articulatory programs constrained by the trivial fact that consonants cannot be pronounced without vowels, one can immediately ask whether syllabization - or any other components of so-called UG - are in fact not a consequence of particular interactions of more general cognitive processes and neurophysiological characteristics of human beings (9.2.6) and their particular social environments.

Of syllables and chunks

Somewhat contradictory to what term "universal" normally means, a *fidele* nativist would consider the following equation:

LD = General cognitive processes (Linguistic Input + Extralinguistic Input)

as an unexcusable heresy. This is so, because he considers UG to be the core component of a language-specific cognitive module and not a general cognitive process. For some nativists, there is only one domain where mind uses rule-based grammars: language. Others may be ready to accept that some other domains of human activity - from walking, dancing, body excercising and mating through music-generating, food preparation and object creation to ritual-performing, healing or simple arithmetics - can also be rule-based and encapsulated in specific cognitive modules Fodor (1983) or even have sorts of grammars of their own. Thus, all nativists believe that foundations of language faculty are to be found in genes, only few, however, would accept that triggering of those very same genes could result in activity of drumming or salsa-dancing as well.

Of non-linguistic grammars

17 Some universals like compositionality, graduality, specificity etc. were mentioned in section 9.2.4

To summarize: more than half a century ago, the generativist approach to language threw an energetic spark into muddying waters of structural linguistics, thus igniting a passionate interdisciplinary debate between linguistics and computer science. Strictly formalist, transformationalist approach had failed to furnish a complete, consistent and elegant framework for the study of natural languages but significantly facilitated construction and further development of artificial and programming languages. What failed in greater extent, however, was the nativist entreprise aiming to discover DNA-encoded "innate" predispositions specific to language. In spite of two generations of effort of linguists, psychologues, geneticians or clinicians, no "language gene" was discovered and the answers to questions asking

Of success and failure of chomskian entreprise

- Which composents of Language Acquisition Device are innate?
- What is the nature of its core, the Universal Grammar?
- Which processes are purely language-specific and which are more general?

seem to be obfuscated as they ever were, potentially because its terms mean slightly different things for computer scientists, different things for logicians, different things for psychologues and different things for linguists. But the lack of answers to such question notwithstanding, an orthodox nativist position adopted by Chomsky had nonetheless resulted in a state profitable for everybody: on beginning of 21st century, the vaste majority of all cognitive scientists agrees that man's language faculty is a result of interaction between at least two major components:

Of consensus

1. cognitive and physiological characteristics tuned by phylogenesis of human species
2. input to which the language learner is exposed during its prenatal and postnatal ontogeny

and the whole *nature & nurture* debate is not led anymore in terms of mutually exclusive *either/or*, but focuses more on degree and forms of mutual triggering and epigenetic interactions between innate and acquired programs. Let's now leave the discussion of those who emphasize the importance of the first component and focus on those who emphasize the role of the second component: empiricists and constructivists.

END GENERATIVISTS AND NATIVISTS 9.4.2.0

9.4.3 EMPIRICISTS AND CONSTRUCTIVISTS

Empiricists argue that human knowledge arises principially from experience (ἐμπειρία). They thus explain the acquisition of a certain word or expression in terms of *percieved* contexts within which the child hears a given word or expression. Empiricist paradigm is thus quite similar to associanist, and in lesser extent also behaviorist paradigms mentioned above (9.4.1).

But what about acquisition of structures and principles which are not salient, evident or even percievable at all? What about acquisition of all those directly unpercievablable entities - be it rules, schemas, patterns, templates or something else - which determine the result of linguistic comprehension and production, yet are not present *per se* in any utterance? What about all word order, long-distance dependency or chiasmatic principles which definitely have to be somehow encoded in the mind - because they act - but are detectable only through consequences of their actions? Because they express themselves only through their instances, and because they instances vary, pure empiricism has to encounter serious epistemological problems when explaining acquisition linguistic representations operating with and on more general levels of abstraction solely through sensory experience.

Through hundreds of years of both theoretical reflexion as well as methodic experimentation, empiricists gradually evolved into constructivists. Being firmly rooted in phenomenology of everyday human experience, constructivists do not deny existence of more general representations. They simply state that all those concrete-surpassing quantifiers, rules, principles, categories, templates or schemas are as natural a consequence of exposure of mind to repetitive, contextualized stimuli, as photosyntesis is a natural consequence of exposure of plant's leaves to light. For constructivists and their connectivist descendants, mind's hardware - the brain - is a generalization device *par excellence* and therefore there is truly nothing mysterious about the fact the mind is able to transcend the concrete and the arbitrary.

Thus, contrary to their nativist counterparts who postulate that child's mind tries to "deduce" concrete grammar of her ambient language by entering the specific parameters into the formal system of universal grammar, constructivists postulate that child's mind in fact "induces" her grammar from and out of myriads specific utterances she hears. In the famous Chomsky vs. Piaget debate, the position of father of all constructivists could be characterized as follows: « Piaget maintains that one's linguistic structures are not defined by the genome, but instead, are 'constructed' by 'assimilating' (organizing) things in the environment in terms of pre-linguistic structrues, and 'accomodating' (modifying) these as they prove insufficient. This mode of functioning, called 'reflective abstraction' is innate, as is some elementary reflex behaviour (e.g., sucking, grasping), but the cognitive structures, even the pre-linguistic ones, are not.» (Piatelli-Palmarini, 1980)

Asides two modes of schema application and modification which Piaget called 'assimilation' and 'accomodation', and which we have already discussed in presentation of Piaget's Genetic Epistemology theory (8.4.4); and asides notion of 'reflective abstraction' which Piaget uses to explain cognitive development occuring well after the age of toddlerese, Piaget introduced other terminology which we consider to be particulary useful when aiming to explain certain facets of LD: circular reaction, schema coordination and interiorization.

Circular reactions are related to the propensity of cognitive system to repeat, reproduce and reactivate its schemas. Primary circular reactions occur between 1-4 months and are triggered by child's discovery that acts, originally performed by accident, can bring about a pleasing consequence, and subsequently repeating the action. Secondary circular reactions occur between 4-8 months of age, are still repetitive and habitus-forming but also involve external objects (e.g. switch switching). Schemas thus formed are subsequently mutually combined and recombined, coordinated and recoordinated thus generating still bigger a variety of behaviours and habits which the child finds out to be useful, pain-reducing or simply pleasurable.

According to Piaget's theory, schemas of first 18 months of age are principially sensorimotor. But later, after the basic perception-action couplings were mastered and optimized in sufficient extent, child leaves the "sensorimotor stage" and enters "the preoperational stage" wherein she starts to "internalize" the schemata. Internalization does not mean that child simply creates neural representations of her sensorimotor couplings: such "neural substrate encoding" is axiomatic for any organism with nervous system and takes place already in prenatal development. Internalization in Piaget's theory means that a child creates neural representations of mental substitutes - symbols - which themselves refer to certain sensory,motor and later also symbolic "realities". Internalization shall subsequently allow the child to execute certain operations purely mentally, without need to materially realize them in physical reality: it is in great extent thanks to internalization that the child can find the shortest way out of an unknown space without the need to physically toddle through all possible paths.

Parallely to Piaget, at the other border of Europe, but in a less Kantian and somewhat more "dialectical materialist" a space, the mind of Lev Vygotskij was slowly converging to practically identical conclusions: the process of internalization of schemas was to explain the ontogeny of thinking. Believing that « the internalization of socially rooted and historically developed activities is the distinguishing feature of human psychology is the distinguishing feature of human psychology, the basis of the qualitative leap from animal to human psychology» (Vygotsky, 1978) and knowing that language is potentially the most important exemplar among such "socially rooted and historically developped activities" Vygotskij went even further than Piaget and postulated that thinking is a form of internalized language: thoughts are inner-speech utterances.

There are, of course, subtle differences between theories of Piaget and Vygotskij. For example, while Vygotskij's theory focuses more on social and cultural forces behind the split between the outside "social" speech and the inner speech, Piaget's theory emphasizes child's individual,egocentric, knowledge-constructing acts. But it would be false to state that Piaget wasn't aware of importance of social aspects for development of cognitive functions, as is evident, for example, from the statement « The individual would not come to organize his operations in a coherent whole if he did not engage in thought exchanges and cooperation with others.» (Piaget, 1947) One can rencocile such point of with Vygotskij: « An operation that initially represents an external activity [e.g., egocentric speech] is reconstructed and begins to occur internally [e.g., private and internal speech]... An interpersonal process [e.g., social language] is transformed into an intrapersonal one [i.e., inner speech]... The transformation of an interpersonal process into an intrapersonal one is the result of a long series of developmental events.» (Vygotsky, 1978).

Such long series of developmental events is, according to the theory of intramental evolution, equivalent to evolutionary process wherein diverse schemata are replicated through processes of internalization and articulation and are selected by their ability to induce intended changes in the (social) environment. In other terms: pragmatic, environment-related concerns are to be present in function evaluating survival and reproduction fitness of such structures.

END EMPIRICISTS AND CONSTRUCTIVISTS (SRC) 9.4.3.0

9.4.4 SOCIO-PRAGMATIC AND USAGE-BASED PARADIGMS

Piaget's and Vygotskij's theories are not theories of LD. They are much more general: they are theories of development of knowlege and thought; they are theories of learning. Under light of such generality, concrete particularities are secondary: thus neither Piaget nor Vygotskij offer specific quantitative values which are to be defined by any engineer aiming to reproduce LD-like processes in silico. They rather offer a general gradual, environment-oriented and ludic framework within which once can do so: that alone suffices.

But what about concrete aspects of child's social learning (Bandura and McClelland, 1977) of language? One can hardly speak about science if specific processes and representations are not experimentally explored and verified; there is no science where specific relations and correlations between variables and phenomena are not evaluated.

Only entreprise which allows to find concrete answers to concrete questions by analyzing concrete phenomena, is truly scientific.

Jerome Bruner was among the first scientists who have performed a detailed analysis of child's LD and interpreted their findings in "social environment" terms. By performing a longitudinal study focused on two boys - Richard and Jonathan, by visiting their homes every fortnight since they were 5 (resp. 3) months old until they were 24 (resp. 18) and by taking half-hour audiovideo recordings of their playing with their mothers; Bruner initiated a paradigm both rigorous yet "natural" and completely non-violent (because with mother and in home environment).

During first months, Bruner focused on games through which the child learns how to manage interactions with the closest social environments. Through games with a cloth clown which the mother makes gradually appear, disappear and reappear; or through the game of peek-a-boo, the infant gradually learns the basic conditions of social and participatory activities . Bruner concludes these observations with words: « If the "teacher" in such a system were to have a motto,it would surely be: "where before there was a spectator, let there now be a participant".» (Bruner and Watson, 1983), indicating that such games help to establish social conventions upon which latter language use shall be based.

Later, Bruner had explored how referential meaning of first words is born through mother-originated object-highlighting and child-originated pointing. Or he analyzed the motherese articulated during the picture-book reading to discover that « The variety of mother's utterance types in book reading is strikingly limited. She makes repeated use of four key utterance types, with a surprisingly small number of variant tokens of each. These types were (1) the Attentional Vocative, e.g. Look; (2) the Query, e.g. What's that?; (3) the Label, e.g. It's an X; and (4) the Feedback Utterance, e.g., Yes.» (Bruner and Watson, 1983). Complete list of such constructions which occured more than once during session at 1;1.1, are presented on Table 5.

It is also worth noting that such utterances were observed to occurs almost always in the sequence:

1. Attentional Vocative
2. Query
3. Label
4. Feedback

Some members of the sequence can be left out - e.g. attentional vocative is left out when mother simply responds to what the child does - but Bruner noticed that the order of utterances is practically never switched. Along with the extra-linguistic context (e.g. book-reading), can be undertood as *the first format*.

TYPE / TOKENS	FREQUENCY
I. Attentional Vocatives	**65**
Look !	61
Look at that!	4
II. Query	**85**
What's that?	57
What are those?	8
What are they doing?	6
What is it?	5
III. Label	**216**
X (=a stressed label)	91
It's an X	34
That's an X	28
There is an X	12
An X	12
That's X	6
There is X	6
Lots of X	5
They are X-ing	5
More X	3
They are X	3
These are the X	3
The X	2
IV. Feedback	**80**
Yes	50
Yes, I know	8
It's not an X	5
That's it!	3
Isn't it?	2
Not X	2
No, it's not X	2

Table 5: Utterances classified as tokens of the four major types of the motherese. Reproduced from table 4.2 in (Bruner and Watson, 1983, pp.79-80).

Format (DEF)

« A format is a standardized, initially microcosmic interaction pattern between an adult and an infant that contains demarcated roles that eventually become reversible.» (Bruner and Watson, 1983)

END FORMAT 9.4

Contrary to pivot schemas () or variation sets (), which are "individual" in the sense that they constructed, stored and articulated by a signle individual, are formats interactive, mutual and shared. It is also important to emphasize the extra-linguistic and pragmatic facets of such a microcosmic scene: « format is a routinized and repeated interaction in which an adult and child do things to and with each other» (Bruner and Watson, 1983).

Successful unfolding of a "format" from its beginning until its very end is possible only if both participants succeed to focus their attention upon the same object of interest. Such "joint attention" is the cornerstone of not not only Bruner's theory, but also of Tomasello's. As a primatologue *par formation*, Tomasello stresses out the fact that humans are only apes which use symbols

- to acknowledge sharing of attention with others
- redirect attention of others to external objects, states or processes
- modify mental states of others

. In other terms, humans are capable of "intention-reading", they are able of "joint attention" surpassing the dyadic I-You relation by integrating an external object (or mental state) into triadic I-You-it relation (Buber (1937)).

For Tomasello, intention-reading « is the foundational social-cognitive skill underlying children's comprehension of the symbolic dimensions of linguistic communication» (Tomasello, 2009). It is supposed to be a domain-general skill allowing not only linguistic communication, but many other practices as well (rituals, tool and house manufacture, co-ordinated warfare, healing, nonreproductive mating etc.) and is strongly intertwined with other phenomena studied by *theory of mind* (imitation, perspective-taking etc).

The principal reason why intention-reading is supposed to be foundational is its ability to attribute *function* to diverse linguistic expressions or their components (e.g. words): « identifying the functional roles of the components of utterances is possible only if the child has some (perhaps imperfect) understanding of the adult's overall communicative intention—because understanding the functional role of X means understanding how X contributes to some larger communicative structure.» (Tomasello, 2009). Thus, not only the observable "microcosmos" within the articulation of the utterance AXB took place, but also "intention of the speaker" has to be understood if the child is to understand the function of X'use in her language. Which, according to Wittgenstein, is equivalent to meaning of X.

Asides the *foundational* processes of intention-reading and joint attention, ambition of Tomasello's theory is to understand the nature of cognitive processes related to: "

1. schematization and analogy, which account for how children create abstract syntactic constructions out of the concrete pieces of language they have heard

2. entrenchment and competition, which account for how children constrain their abstractions to those that are conventional in their linguistic community

3. functionally based distributional analysis, which accounts for how children form paradigmatic categories of various kinds of linguistic constituents (Tomasello, 2009)

also involved in LD process.

Crucial to explanation of these processes is the nature of their input and output representations. Contrary to Bruner, whom his generativist Zeitgeist forced to interpret his data through the terminological prism of "deep" and "surface" structures, has Tomasello conceived his theory in the period when generativism was already somewhat "out of mode". Thus, instead of mysterious UG, isolated lexicon, a monolithic set of transformation rules and omnipresent arborescent structures is his *usage-based theory of language acquistion* based on "item-based constructions", "expressions", "schemas" and "templates" containing "slots" and variable through diverse operators.

A big theoretical advantage of these forms of representation is their ability to can encode multiple levels of generality at once. As such, they have no problem to account for acquisition of such fixed or semi-fixed idiomatic expressions like *ça va?*, *gonna*, *dunno* or *kick the bucket* which seem like rule-generated but are in fact learned by rote. Generative models based on rules -supposed to be generally applicable- and lexicon -whose members are supposed to be as concrete and atomic as possible- have huge difficulties with such fixed or semi-fixed entities[18]

Usage-based models, on the other hand, do not have any problem whatsoever in accounting for existence of such hybrid structures. As

18 Take as an example *ça va?*, the French equivalent of *How do You do?* The form is not purely rule-generated because otherwise a completely normal demonstrativ pronoun ça cannot be substituted for other demonstrativs (il, elle) without the whole losing completely its meaning. On the other hand, it cannot be a member of lexicon neither because it is decomposable and the second component *va* (i.e. "goes") can be, in some argot contexts, substituted for specific verbs (i.e ça tourne; ça roule).

Tomasello puts it: « The impossibility of making a clear distinction between the core and the periphery of linguistic structure is a genuine scientific discovery, and it has far-reaching theoretical implications...it suggests that language structure emerges from language use, and that a community of speakers may conventionalize from their language use all kinds of linguistic structures—from the more concrete to the more abstract, from the more regular to the more idiomatic, and with all kinds of mixed constructions as well.» (Tomasello, 2009). In usage-based linguistics, mind is free to mix terminals with non-terminals as is wanted, necessary and appropriate for realization of communicative intention framed by a specific situation.

This being said we consider it futile to try to summarize all details of Tomasello's broad and detailed synthesis. Instead, reader is hereby invited to read the monography: its lecture should significantly facilitate the interpretation of what shall follow in Part ii and Part iii and could be used as a sort of prolegomena to these. For the purpose of the present exposé, let's just conclude with the following citation: « As children attempt to read the intentions of other persons as expressed in utterances, they extract words and functionally coherent phrases from these utterances, but they also create item-based constructions with open slots on the level of whole utterances. Few theorists of language acquisition deal with these humble creations, and those who have dealt with them (e.g., Braine, 1976) have not provided an account by means of which they **evolve** into more abstract and adult-like constructions.» (Tomasello, 2009)

As a follow-up to this citation, both phenomenological and computational exploration of extent in which evolution of populations of such "humble creations" could be characterized as a process involving intramental replication, variation and selection, is now defined as a principal objective of this dissertation.

END SOCIO-PRAGMATIC AND USAGE-BASED 9.4.4

Having thus glanced at the history of just a few among *legions of savants* who spent time of their lives seeking, in one way or another, to propose an answer to the question:

"How is it possible that humans understand each other yet often not agree with each other?"

we conclude this brief overview with a simple truism which could, with little bit of good-will, reconcile all-above mentioned positions:

"Because we can do so and want to do so."

and note that the debate of what "can" and "want" means in context of "understanding" and "agreement" surpasses by far the scope of our current proposal.

end language acquisition paradigms 9.4

Intention behind less than 50 pages of this chapter was to acquaint the reader with certain facts, concepts and theories related to language development. LD was principially defined as "constructivist process" (9.1) and it was indicated that ontogeny of language competence can be understood as a process of gradual optimization of one's linguistic structures and processes. Difference between "comprehension" and "production" of language use was emphasized and a "dogma" was postulated, stating that in developing mind, language comprehension is to precede the language production.

Certain facts specific to all facets of linguistic competence were brought to reader's attention. These were presented in order to indicate that certain domain-general processes - related to contrast detection; category construction aided by input-driven distributional analysis; schematization; pattern-matching etc. - operate on all levels, from prosodic to pragmatic and beyond. Gradual increase in diversity and complexity of representations was observed in many different cases: in learning of phonotemplates, in vocabulary development, in construction of item-based constructions from pivot schemas etc.

Input and social interactions were also often mentioned. Brown's "word game" and Bruner's "formats" were discussed in case of vocabulary learning and "variation sets" were told to facilitate the acquisition of morphosyntax. Special properties of "motherese" were praised for their abilities to significantly reduce the computational complexity decyphering process. This was the reason for adoption of somewhat critical a stance towards formal "learnability" and "nativist" theories which see perfect learner in imperfect environment there, where we tend to see an imperfect learner in the perfect environment.

For this reason we prefer to end this chapter with the citation from a book which, asides (Clark, 2003; Tomasello, 2009) was our first *guide* in the evercomplex labyrinth of LD-related data and theories:« If grammar were innately specified in the infant brain and simply triggered by hearing the correct forms, why would it take so long to manifest itself? In such a case, one might expect grammar to be an inherent part of the child's output from the start. After all, by the time infants reach their first birthdays they will have had considerable exposure to linguistic input and, as we saw, they already have a significant receptive vocabulary. The average three-month-old has already had approximately 900 waking hours or 54,000 minutes of auditory input. And these calculations do not even take into account the last three months of intrauterine life...» (Karmiloff and Karmiloff-Smith, 2009). Hundreds of thousands of minutes when the child is still a toddler, millions of sequences of linguistic tokens pre-processed and pre-formated by those-who-love-the-one-to-whom-they-speak: plenty of high-quality data to proces by ever-evolving populations of cognitive schemata. Plenty of information to induce useful patterns from.

END DEVELOPMENTAL PSYCHOLINGUISTICS 9.4

10 COMPUTATIONAL LINGUISTICS

Computational linguistics (CL) is a discpline positioned at the intersection between linguistics and informatics. The extent of this intersection is huge because both informatics and linguistics have one important property in common: on the most formal level, they both deal with sequences of symbols. In this chapter, such abstract and theoretical perspective shall be more closely discussed in section dedicated to Formal Language Theory 10.2 and its modular counterpart theory of Grammar Systems 10.2.3. Subsequently, more "real-life" problems related to Natural Language Processing (10.3) shall be mentioned, with special focus being put on the problems of:

Plan of the chapter

- geometrization of semantics attained by projection of natural language corpora into N-dimensional vector space
- part-of-speech tagging and part-of-speech induction which make it possible to automatically attribute grammatical category membership to different tokens occurent in the corpus
- grammar induction which makes it possible to infer grammar G_L of language L from the corpus C_L

But before doing so, let's now briefly discuss that sub-discipline of CL, which is older than CL itself.

10.1 QUANTITATIVE AND CORPUS LINGUISTICS

Centuries before first computers were invented, the preceptors had already been counting words in different corpora. Panini and his disciples contemplated the Vedic corpus (9.4.2) in order to invent the most cognitively efficient means of transmission of the Corpus through human cerebral wetware without ever writing it down, Dominicans were creating concordancy tables of biblical verses, Arabs analyzed the Quran and kabbalists the Torah: and it cannot be excluded that practically all members of these otherwise divergent currents had found particular pleasure in doing so.

The advent of computers have changed such an opaque *hermeneutic passe-temps* of some most devoted philologues into a full-fledged and highly empiric science. The symbol-reading and symbol-manipulating faculties of Turing machines embedded in first thousands, then millions, then billions of transistoric flip-flops have allowed to process all words of one's library in few seconds. Frequencies of occurence of a word W - i.e. the answer to the question "How many times does

Frequency of occurence

the word W occur in corpus C ? - were evaluated for still bigger and bigger corpora; probability distributions of relative frequencies - i.e. F_W normalized by number of all words in C - were assessed.

And new evidence was given that natural language corpora contain such salient regularities, that one can or even must explain them in terms of mathematical "laws".

10.1.1 ZIPF'S LAW

The basic form of Zipf's law can be expressed by an equation:

$$f_W * r_W \approx C$$

where f_W is a frequency of occurence of word W in the corpus and r_W is word's rank in the table where all words of the corpus are sorting according to their frequency in descending order (i.e. the most frequent word has rank 1, the second has rank 2 etc.) and C is a constant. In other terms, Zipf's law states that the frequency of a word is inversely proportional to its rank in the frequency table, which is equivalent to the statement that « the frequency of a word in a text and its rank is approximately linear when plotted on a double logarithmic scale» (Ferrer-i Cancho and Elvevåg, 2010). In terms of probability distributions, this law states that frequencies of occurence of words in the corpus are independent and identically distributed random variables with distribution

Formulation of ZL

$$p(f) = \alpha f^{-1-1/s}$$

id est, the "power law"/Pareto distribution with exponent *s*.

G.K.Zipf was profoundly convinced that this regularity is an expression of a domain-general cognitive eco(nom|log)y *principle of least effort*. More concretely, he conjectured that the observed regularity is a consequence of tendency of linguistic system to attain the state of *vocabulary balance*, i.e. the state in which two oposing forces, force of unification and force of diversification, characterized as follows: « on the one hand, the Force of Unification will act in the direction of decreasing the number of different words to 1, while increasing the frequency of that 1 word to 100%. Conversely, the Force of Diversification will act in the opposite direction of increasing the number of different words, while decreasing their average frequency of occurrence towards 1. Therefore number and frequency will be the parameters of vocabulary balance.» (Zipf, 1949)

Vocabulary balance

Next generations of linguists -Chomsky included- and mathematicians - e.g. Benoit Mandelbrot - were less enthusiastic when it came to the importance which Zipf attributed to his "law". The point of conflict was not whether the frequencies of words in natural language texts follow the power law distribution: each new analysis had demonstrated that it is, verily, the case. The argument arose when

Critics of ZL

some authors started to considered Zipfian distributions as a tautological necessity, as a phenomenon emerging anytime, even in randomly generated artificial corpora. One famous study concluded: « Zipf's law is not a deep law in natural language as one might first have thought. It is very much related the particular representation one chooses, i.e., rank as the independent variable» (Li, 1992) and reiterated Mandelbrot's remark that ZL is "linguistically very shallow".

ZL defended

On the other hand, more recent article convincingly demonstrates that « good fit of random texts to real Zipf's law-like rank distributions has not yet been established. Therefore, we suggest that Zipf's law might in fact be a fundamental law in natural languages» (Ferrer-i Cancho and Elvevåg, 2010). A lateral support for such claim comes also from study which focused on "evolution" of ZL - notably the exponent s - in language ontogeny. Given statistically significant observations that « in children the exponent of the law tends to decrease over time while this tendency is weaker in adults...Our analysis also shows a tendency of the mean length of utterances (MLU), a simple estimate of syntactic complexity, to increase as the exponent decreases. The parallel evolution of the exponent and a simple indicator of syntactic complexity (MLU) supports the hypothesis that the exponent of Zipf's law and linguistic complexity are inter-related» (Baixeries et al., 2013). We add that it would be somewhat difficult to observe such ontogeny-related modifications of ZL's exponent if ever ZL was just a pure artefact owing its existent to one's choice of mathematical formalism.

ZL in language ontogeny

At last but not least, we consider it worth reiterating that similar Zipf-Mandelbrot distribution were observed in sciences other than linguistics. In ecology, for example, the distribution between number of species observed as a function of their abundance is understood as a zipfian phenomenon (Mouillot and Lepretre, 2000). Given that ecology is principially a science about equilibrium-seeking systems consisting populations of entities which interact and replicate, we consider the fact that similar scaling phenomena operate both

ZL and ecology

- in realm of words - and, Zipf would add, also in realm of "meanings" because « words are tools that are used to convey meanings in order to achieve objectives...the reader may infer from the orderliness of the distribution of words that there may well be a corresponding orderliness in the distribution of meanings because, in general, speakers utter words in order to convey meanings» (Zipf, 1949)

- in ecology

to support the Thesis that certain neurolinguistic structures intramentally interact and replicate. END ZIPF'S LAW 10.1.1

10.1.2 LOGISTIC LAW

Another among multiple "quantitative laws" which seems to be of particular interest for anyone aiming to understand and create evolutionary models of language ontogeny is the "logistic law" often known as Piotrowski's law. This law postulates that language development follows the logistic curve formalizable into mathematical notation as

$$p(t) = \frac{1}{c + a \cdot e^{-b \cdot t}}$$

whereby t denotes time, $p(t)$ denotes quantitified value of an observable property of lingustic system in time t, e is euler's constant and a, b, c are parameters of the model. We consider it important to mention that what research of Best (2006) and other participants in the "Gottingen project" indicates is that the law applies not only on ethnogenic, cultural and historic (i.e. *Sprachwandel*, c.f. 8.5) but also on ontogenic development of linguistic systems (i.e. *Spracherwerb*). *LL's ubiquity*

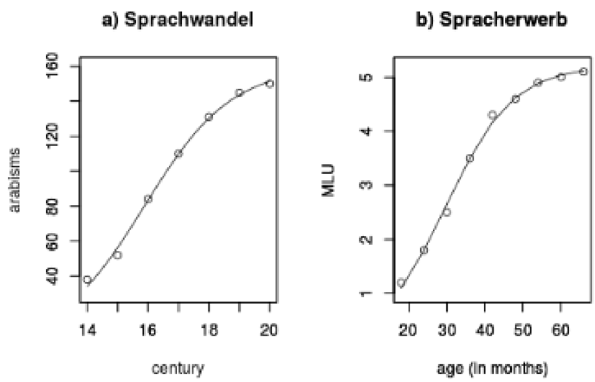


Figure 17: Logistic law in relation to historic and ontogenetic linguistic processes. Data taken from Best (2006).

Figure 17 illustrates examples of these two cases:

- points on image (a) represents increase of amount of words of arab origin into german between 14th and 20th century while the line represents the ideal logistic curve with parameters (a=7.41, b=0.696, c=160)
- right image (b) represents gradual increase of Mean Length of Utterance (9.2.4)

While members of certain schools may argue adamantly that many phenomena in both ethnogeny and ontogeny can be explained or even modelised in terms of logistic curves, other data[1] shall rightfully oblige others to express certain scepsis to capacity of logistic

1 C.f., for example, Figure 15 to see some data which, while aiming to represent practically the same phenomena as Figure 17 seems unsubsumable under the logistic curve.

curve to *cover* practically ALL quantitative aspects of language acquisition process, and to do so with sufficient statistic significance. On the other hand, some phenomena like the "vocabulary spurt" (c.f. 13) seem sometimes to follow the logistic first-slow-then-fast-than-slow-again so faithfuly, that it would be unwise to to apriori ignore such a salient, formal and high-order analogy between ethno- and onto-geny of intra- and inter-personal linguistic eco-systems.

Be it as it may, instead of trying to adequately address the Haeckel-like conjecture *some processes of linguistic ontogeny are formally isomorph to certain processes in linguistic ethnogeny*, it seems more appropriate to focus the attention of the reader upon the fact that closing the previous paragraph with the plural form of the term *eco-system* was intentional. This is so because it was indeed ecology where logistic curve models where deployed for the first time: introduced in 1838 by Pierre-François Verhulst as a model whereby *the reproduction of population is proportional to both the existing population and the amount of available resources*, and canonized later by (Lotka, 1925) as *the law of population growth*, it is closely related to predator-prey (or Lotka-Volterra) differential equations which are, even more than hundred years since its conception, still consider as a model of reference of population dynamics of biological and ecological systems within which two or more species interact.

Law of population growth

END PIOTROWSKI'S LAW 10.1.2

In this brief overview of Corpus and Quantitative linguistics we have mentioned two hallmark "laws" postulated (or discovered?) by proponents of this discipline: Zipf's law and Piotrowski's logistic law. We have indicated that both of these laws have certain ontogeny-pertinent aspects which make them worthy of interest not only to researchers interested in historical linguistics, but also to those known as "psycholinguists".

What's more, it was also indicated that there exist a certain analogy, a certain *partage* of features, between developmental psycholinguistics and ecology:

Towards ecology of intramental representations

- not only frequencies of words in corpus, but also abundances in ecology are Zipf-distributed
- logistic curves are used to model not only rate of (pro | intro)duction of new words into the copus, but also population dynamics of diverse mutually-interacting species within a specific ecosystem

Given that there exists a certain formal similarity between models of dynamics occurent within ecological or linguistic systems, transposition of certain principles from ecology into psycholinguistics may seem to be appropriate.

END QUANTITATIVE AND CORPUS LINGUISTICS 10.1.2

10.2 FORMAL LANGUAGE THEORY

Formal Language Theory (FLT) is a computational theory of of formal languages and formal grammars. Being rooted in apodictic definitions of computer science, mathematics and logics, its aim is to offer solid, coherent and scientifically valid framework useful for

1. design of new artificial (e.g. programming) languages

2. elucidation of structure and function of natural languages

Of FLT's usefulness

No-one denies that when it comes to the first objective mentioned above, the practical utility of FLT-originated concepts and principles is demonstrated anytime a computer translates the source code into machine code. It is true that without any solid theory *thematising the rules of production and parsing of symboling sequences*, it would be highly problematic to procceed all the way from romantic intuitions of lady Ada Lovelace, notebooks of Gottlob Frege, Zuse's Plankalkül through assembler, C, C++ all the way to parsers, linkers, and compilers of modern high-level programming languages like Python, PERL or R.

Of FLT's uselessness

But the capper evidence that FLT can also yield a framework useful for the attainement of the second goal, is yet to be furnished. In spite of effort initiated by Chomsky's focus on generativism (), that is, in spite more than half of century of intellectual work of thousands of most brilliant minds of their generation, no pure FLT-based model[2] was proposed, which could account for diversity of forms of even such a morphologically poor language as English. Sadly for science, sectarian disputes within FLT community are of envergure which makes it impossible to answer even the most trivial problems, like that of positioning of natural languages within Chomsky-Schutzenberger hierarchy.

This being said, let's just introduce the conceptual pillars upon which the FLT stands.

10.2.1 BASIC TENETS (DEF)

) FLT is based on notions of symbols, sequences and sets. Thus,

1. alphabet A is defined as a finite set of symbols including the empty symbol ϵ

2. string S is defined as an ordered sequence of concatenated symbols contained in A

3. language L is defined as a set of strings over A

2 By pure FLT model, we mean a model that does not contain any statistic components.

4. * (Kleene star) is a free moinoid unary operator generating all possible strings over a certain alphabet,

$$A^* = \bigcup_{i \in N} A_i = \{\varepsilon\} \cup A \cup A_2 \cup A_3 \cup A_4 \cup \ldots$$

A^* therefore denotes the infinite set of all possible strings over A and language L is either a subset of, or equivalent to A^*, i.e. $L \subseteq A^*$

Given this, grammar G_L of language L is a means how to characterize which among the members of A^* are to be contained in L, and which are not. In traditional FLT, Grammar is defined as follows:

Grammar and Rule(DEF)

A grammar G is a tuplet $\{V_T, V_N, X, R\}$, where V_T is the set of terminal elements, V_N is the set of non-terminals, X -an "axiom symbol" is a member of V_N ($X \in V_N$), and R is a finite set of N rules $R = \{r_1, r_2, ..., r_N\}$.

A (rewriting | (produc | substitu)tion) rule r has a form $foo \rightarrow bar$ and fundamentally denotes 2-ary substitution operation wherein the first operand foo is substituted by second bar, or vice versa.

END GRAMMAR AND RULE(DEF) 10.2.1

The expression *vice versa* is quite important here, for it denotes that grammar can be useful in both

1. (produc | genera)tion of terminal string-expression E of language L started by input of "entry axiom" X which takes place when rules are applied in right-wise order (i.e. foos are substituted by bars) and is to be terminated only when the string does not contain any non-terminal symbols.

2. parsing / comprehension of string (sentence) started by input of E and terminating when the some substitution transform E or its derivates into X. In other terms, this scenario occurs when rules are applied in right-wise order (i.e. bars are substituted by foos within the string) and terminates when the working string does not contain any terminal symbols.

Practically all currently widely used notations take this symmetry between substituens (that-which-substitutes) bar and substituendum (that-which-is-substituted) foo, as granted. Table 6 illustrates "plain", "compressed" and "uncompressed" grammars written down in three common notations[3]

3 In all notations we follow the common convention of denoting non-terminal symbols with uppercase characters (e.g. "B", "M", "S", "X") and terminal symbols with lowercase (e.g. "a", "b", "m" ...) characters.

Note that the "uncompressed" and "compressed" grammars of G_L are equivalent only when it comes to language they cover, but not in the way how the G_L is represented. They are functionally but not structurally isomorph. Thus, where "uncompressed" grammars represent disjunction in terms of multiple trivial rules (for any disjunction, one has as many rules as there are disjunct elements), compressed grammars represent a disjunction by one rule only. The price, however, is the need to introduce the disjunctive symbol | and to use it every time when disjunction needs to be marked.

	S-notation	**Backus-Naur notation**	**PERL-notation**
Plain	$X \rightarrow baba$	$X :=< mama >$	$s/X/mama/$
	$X \rightarrow mama$	$X :=< baba >$	$s/X/baba/$
Compressed	$X \rightarrow SS$	$X := SS$	$s/X/SS/$
	$S \rightarrow ba \mid ma$	$S :=< ba > \mid < ma >$	$s/S/ba \mid ma/$
Uncompressed	$X \rightarrow MM$	$X := MM$	$s/X/MM/$
	$X \rightarrow BB$	$X := BB$	$s/X/BB/$
	$M \rightarrow ma$	$M :=< ma >$	$s/M/ma/$
	$B \rightarrow ba$	$B :=< ba >$	$s/B/ba/$

Table 6: Diverse notations of three grammars covering the language $L = \{"mama", "baba"\}$.

For a logics-oriented reader, it may be useful to conclude that FLT considers languages and their respective grammars to be equivalent to a sort of formal system. Thus, the set of all strings A^* being understood as a set of all (i.e. both true and false) proposition, the string belonging to language L can be understood as true theorems and the act of *deriving* these by G is equivalent to theorem proving.

END BASIC TENETS 10.2.1

10.2.2 CHOMSKY-SCHÜTZENBERGER HIERARCHY (TXT)

Language L can be classed according to type and form of rules which its respective grammar G_L contains. Undoubtably the most common typology is the Chomsky-Schützenberger hierarchy of languages which classes all possible languages into one among four classes. These are defined as follows:

1. **unrestricted grammars** contain rules which can contain any combination of terminals and non-terminals in both substituens and substituendum

2. **context-sensitive grammars** have rules of the form $\alpha A \beta \rightarrow \alpha\gamma\beta$ with A a nonterminal and α, β and γ strings of terminals and | or nonterminals. Strings α and β may be empty, γ , however, must be nonempty.

3. **context-free grammars** which have rules of a form $A \rightarrow \gamma$ with A a being a nonterminal and γ being a string of terminals and/or nonterminals.

4. **regular grammars** have rules with one single non-terminal on a left-side and one terminal with max one juxtaposed non-terminal on the right side

Containment hierarchy

These classes of languages are mutually embedded. Thus, the class of regular languages is the specific subset of context-free languages and it follows that while any regular language is a context-free one, not any context-free language is a regular one. Idem for embeddings of higher-order: context-free languages are specific cases among context-sensitive grammars and context-sensitive languages are just a certain specific subset in the vast "unrestricted" ocean of Type-0 languages.

Types of automata

The main categorization being so canonized, great deal of FLT is occupied with study of algebraic and computational properties of these classes. It is thus known that languages produced by regular grammars can be recognized by finite state automata (FSAs), context-free languages can be recognized by non-deterministic push-down automaton, context-sensitive languages are recognizable by means of linear-bounded non-deterministic Turing machines while an arbitrary Type-0 is not to be recognized by nothing less complex than a Turing machine.

Usefulness of C-S hierarchy

It is indeed in such overlap regions between computer science and algebra, whereby the conceptualization of things in terms of C-S hierarchy finds its utmost utility. Utmost and practical: for as it was stated - but merits to be re-stated- purely theoretical explorations of mutual relations between diverse types of grammars and diverse types of symbol-manipulating automata can and indeed do have serious material consequences: faster encoding and faster decoding means faster machines.

Uselessness of C-S hierarchy

But in relation to diversity of expressions of natural languages, FLT taxonomies can be quite misleading. For example, as is nicely illustrated in overview (Jiménez López et al., 2000, pp. 87-97), even after decades of debates, "linguists" cannot even find an agreement whether English alone fits into class of context-free languages or whether it is more appropriate to consider it as *a priori* case-sensitive language. For experts coming from FLT-ignorant domains of linguistics - be it linguistic typology, comparative grammar etc. - such debates express nothing else than sad vaste of intellectual resources. Confronted on a daily basis with the astounding diversity of linguistic structures grounded in the substrate of their usage, adherents of such schools would at maximum dare to utter:

"In world of natural languages, nothing is certain and nothing is fixed. Asides the fact that natural languages belong into class of Type-0 languages. Maybe."

END C-S HIERARCHY 10.2.2

10.2.3 GRAMMAR SYSTEM THEORY (TXT)

A spin-off branch of Formal Language Theory which is of particular interest in regards to overall objectives of this Thesis is devoted to study of Grammar Systems (GS). A grammar system is a « set of grammars working together, according to a specified protocol, to generate a language» (Jiménez López et al., 2000). Thus, contrary to definitions of canonic FLT in which one grammar generate ones language, in GS several grammars work together in order to generate one language. Grammar Systems can be therefore considered as a sort of multi-agent variants of traditional «monolithic» FLT. Such multi-agent nature of GSs implies cooperation, communication, distribution, modularity, parallelism, or even emergence of complexity.

Polylithic model

Let's take as an example the most simple among the GS, so-called "language colonies", defined in (Kelemenová and Csuhaj-Varjú, 1994) as follows:

Language Colony (DEF)

A language colony colony C is an (n+2)-tuple $C = (T, R_1, ..., R_n, S)$, where

1. $R_i = (V_i, T_i, P_i, S_i)$, for every i, $1 \leqslant i \leqslant n$, is a regular grammar generating a finite language; R_i is called a component of C;

2. $S = S_i$ for some i, $1 \leqslant i \leqslant n$; S is called the startsymbol of C;

3. $T \subseteq \bigcup_i = 1^n T_i$ is called the set of terminals of C

   And the total alphabet of C is denoted by V, i.e. $V = \bigcup_i = 1^n (T_i \cup N_i)$

END LANGUAGE COLONY 10.2.3

Figure 18 illustrates a very simple bi-component (n=2) "language colony" variant of a GS. What is striking in case of even such a simplistic colony is that *the very fact of sharing and exchange* of strings between two otherwise finite regular grammars results in generation of an infinite language.

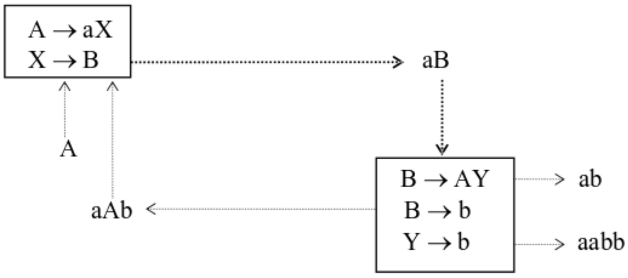


Figure 18: Emergence of "miraculous" infinite generative capacity by means of interlock of two finite grammars. Figure reproduced from Kelemen (2004).

Let it be reiterated: by allowing two or more finite components to communicate through a common symbolic environment, one can generate a set of strings - a language - with potentially infinite cardinality ! Kelemen (2004) denotes such behaviour - which is very common in the world of GS - with the term «miracle».

Blackboard model

The cornerstone idea of not only language colonies but also of any other GS is that diverse "component" grammars share a common "environment". This environment is nothing else than a shared string whereupon and wherein diverse components grammars apply their rules of production. In analogy to class (population) of individual students which together solve the problem on the blackboard they see, the term "blackboard model" is often used to denote the idea. For psychologues this model can be somewhat reminiscent of "working memory" accessible and accessed by diverse independent and encapsulated cognitive modules. Computer scientists, on the other hand, may see some similarity with multiple computational threads accessing the same address space in the shared memory.

Other Grammar Systems

Aside "language colonies" and GST introduces and precisely defines many other theoretical and formal constructs like "Cooperating Distributed Grammar Systems", "Parallel Communicating Grammar Systems" and "eco grammar systems". Notably due to life-long work of Erzsébet Csuhaj-Varju and substantial contributions by George Paun and Jozef and Alica Kelemens are these constructs developed in such a detail that it is practically impossible for us to introduce here, in extent and rigour they merit, the exact formalisms of GS theory in closer detail. Instead, we forward a potentially interested reader to the doctoral dissertation of Jiménez López et al. (2000) which contains many persuasive arguments for application of GS upon the study of natural human languages.

GST still mostly theoretical

On the other hand, the forereferred dissertation is limited by the fact that it mostly proposes to use the Grammar System Theory as a framework explaining the final, i.e. "adult" linguistic component, and not as a framework which could elucidate the very process of language development and language acquistion. In fact, we are not aware of any study which would use the GST as a theoretical explanatory framework for the process of LD, nor of any tentative aiming to implement GST in concrete programs, offering solutions to concrete practical "natural language processing" (NLP) problems.

END GRAMMAR SYSTEMS 10.2.3

FLT unites set theory, algebra and theory of formal systems in a highly abstract and subtle conceptual framework aiming to help us (and machines) to conceive more optimal sequences of operations within the realms of sequences (strings) of symbols. It introduces many useful notions like that of

1. terminal symbols, i.e. those symbols which materially occur in the articulated utterance (i.e. are parts of "signifiant")
2. non-terminal symbols, i.e. those symbols which denote generic properties inherent in and specific to the utterance
3. substitution rules and grammars (10.2.1)

which are, in one form or another, to be found in all linguistic theories at least since Panini (9.4.2). One simply cannot have a linguistic theory, no matter whether general, descriptive, generative, psycholinguistic or developmental without postulating both material observables (terminals), non-material non-observables (non-terminals) and something like a list of principles which relate the two.

Historic context of FLT's conception

Unfortunately, FLT was canonized in an era when computer scientists and computational linguists had to think about allocation of every byte of the memory[4] In such context, the CPU-register-manipulating recursive *while-loops* were considered as magical means of generation of big amount of output from minimal input. Thus, a sort of obsession with the notion of recursivity was born which led generativists to

1. tentatives to explain huge part (or all) human linguistic creativity in terms of recursivity
2. ignorance of the role which memory plays not only in concrete situations of linguistic performance, but also for overall stability of system underlying one's linguistic competence

No syntax without semantics

What's more, FLT is strictly about syntax. It is, *ex vi termini*, a self-encapsulated formal system and any tentative to make any reference to the world to the world of semantics beyond syntax is predesti-

4 However, the contemporary generation of computer scientists is not subjected to such constraints. Memory is cheap in the world where *640kb ought NOT to be enough for anybody*.

nated [5] to put FLT into state of irreversible havoc. For the world of meanings is the world of passionate contextual transpositions, useful metaphores, implicit ambiguities and fuzzy approximations; FLT, on the other hand, brings about the realm of evermore-abstract arborescent hierarchies of pure reason. Fitting one into another, subsuming syntax to semantics or semantics to syntax, thus seems to be at least as absurd a problem as the good old egg-chicken dilemma.

END FORMAL LANGUAGE THEORY 10.2.3

10.3 NATURAL LANGUAGE PROCESSING

Natural Language Processing (NLP) is a field of artificial intelligence and linguistics which explores machine's faculty to understand, produce and interact in natural languages. In contrast to both quantitative and corpus linguistics which mainly concentrates on discovery of general quantiative principles and sometimes on data-mining, as well as in contrast to FLT whose ultimate challenge is purely theoretical, is NLP concerned with concrete, practical and real-life problems of verbal interaction between humans and machines.

As was already noted in Chapter 4, the so-called Turing's Test (TT) is -at least in the canonic[6] form in which Alan M. Turing had proposed it- in its very essence nothing else than a NLP challenge. This is so because in the canonic TT, the interaction between the human tester and the artificial testee is mediated solely through written verbal modality.

The task of creating an artificial system which would truly pass the TT is not as easy as Turing and early computer scientists had believed. Natural languages are multi-layered structures whose components mutually interact both with each other as well as their external environments, the very personal identity of their host not excepted. Natural languages serve many goals - giving commands, transfer of information (or deceipt), telling stories - and often exploit highly irregular means with which these goals are attained.

Machines, on the other hand, are regular and ordered. If not programmed otherwise, they blindly follow the path towards the stationary state; if not programmed otherwise, they are unable to deal with any irregularity whatsoever. Thus, in order to bring the ordered world of machines together with the unpredictable world of living language, NLP engineers usually proceed step after step: one minute linguistic problem is understood, formalized and subsequently tackled with in one's source code. Then another.

---

5 Take, as an example, the introduction of Θ roles into Chomsky's Government & Binding Theory.

6 C.f. Hromada (2012a) for a description of taxonomy of TT-consistent scenarios allowing the evaluation of not only linguistic, but also emotional, spatial, visual, corporal, moral etc. intelligences of an artificial agent.

Indeed many are such problems:

- author attribution
- plagiate detection
- named entity disambiguation
- word and/or morphological segmentation
- sentiment analysis
- relationship extraction
- rhetoric figure detection (Hromada, 2011)
- automatic summarization
- discourse analysis
- anaphora resolution
- parsing
- automatic translation
- natural language understanding
- natural language generation
- question answering

all these are just few among dozens of other tasks which NLP experts aim to tackle. These are, in practice, almost always solved by means of adoption of NLP's ultimate methodology: the machine learning.

10.3.1 MACHINE LEARNING

subsec:ml

Machines can learn. That is, machines are able to discover underlying general patterns and principles governing the concrete input data and can subsequently exploit such general knowledge in contact with data which they have never seen before. They « can use experience to improve performance or make accurate predictions» (Mohri et al., 2012). And in everbigger number of domains, they do so still better and better than their human teachers.

Since the moment when machine learning (ML) was first defined, in relation to game of checkers, as « field of study which gives computers ability to learn without being explicitely programmed» (Samuel, 1959) has the discipline of ML evolved in an extent which is hardly compressible into a single book (Mohri et al., 2012) and certainly incompressible into limited scope of this subsection. This is so because not only does the number of domains of ML's application grow from year to year, but firstly because the quantity of distinct ML methods is already counter in dozens, if not in hundreds.

The general framework - sometimes also called "learning theory" (LT) - however, stays the same. No matter whether in psychology or in computer science, LT principially studies how an information-processing system (e.g. brain or computer) processes, represents and stores data sensed from external environment, how it internally transforms them and how outputs of such transformations influence subsequent activity of such a system (including sensing and processing of future data). There is thus a system that learns (the learner system, LS) , the learnt information (LI) and the process of learning (PL). Interactions among these three components, whether one should postulate less (e.g. in case when $LS \neq PL$) or more such components (e.g. in case when sensed data differs from learnt information), and many other - some stemming from neurosciences, other from pure mathematics - all such topics are to be explored by full-fledged LT.

A distinction which is most pertinant for the purposes of this Thesis - and one may argue that for the ML in general as well - is the distinction between **supervised** and **unsupervised** learning.

Supervised learning, called also learning-with-Teacher[7] is based upon the idea that a full cycle of a learning process consists of two stages:

1. **training | learning stage** - LS is first exposed to set of problems and their respective solutions, then aims to create the model associating the two

2. **testing | evaluation stage** - LS exploits the previously constructed model in order to furnish solutions to problems to which she wasn't exposed during the training stage. Its performance is then evaluated according to certain evaluation metrics.

In unsupervised learning, on the contrary, it is expected that the one-who-launches-the-program shall not furnish any explicit solution | answer-related information to the LS. The training phase is thus practically equivalent to the testing phase: both contain questions; neither contain answers. LS is simply furnished a huge dataset - in unsupervised NLP practice, the dataset is almost always equivalent to textual Corpus - and is asked to do something reasonable with it. Cluster the corpus contents into classes, for example.

While distinction between supervised and unsupervised seems to be crystal-clear for anyone practicing the NLP *fach*, the "cognitive plausibility" of fully unsupervised learning is more than discutable. Primo, the distinction turns out to be problematic for any models of phenomena in which the very order of exposure - i.e. the fact that

7 Or learning-with-Oracle, if the Teacher system is able to correctly solve the problem (e.g. furnish the answer) immediately after it received input sufficiently describing the problem (e.g. a meaningful question).

the corpus to which LS was exposed contains first the token *baba* and only later the token *mama* - can significantly influence the learning process. Thus, for models for which holds the statement « the engineer's decision to confront the algorithm with corpus X and not Y, and to do so in the moment T_1 and not T_2, is already an act of supervision» (Hromada, 2014b) the method cannot be considered as stricly un-supervised even in absence of any explicit answers.

Secundo, in case of modeling of LD processes, one cannot say that toddler undergoes "unsupervised" learning just because the input to which she is exposed does not contain any explicit corrections, cues or answers. The very corpus is the answer and - from toddler's point-of-view - the very authority of the adult who furnishes the corpus mints the corpus with justification of its truthfulness and validity. The very notion of "valid solution" or "correct input-output mapping" loses non-negligeable part of its importance when one realises that LSs which we aim to discuss here, can be conditioned to perceive agrammatical and false utterances as grammatical and true. No matter whether it is the case of a child in the middle of ego-centric stage or a victim of a propaganda machinery, it is often NOT the adequacy with external reality niether consistency with as big a set of propositions, which counts. Instead, it is the repetition, the frequency of co-occurrence, the self-referential and self-reinforcing set of references to the minimal "seeding" set of symbols, which counts and which directs the learning process.

Tertio, it is evident that both accuracy as well as speed of learning of solving of a particular class of problems is, at least in case of human learners, significantly catalyzed by the presence of a teacher skilled enough to adopt the input-to-be-taught to the momentanous state of LS. Vygotsky's "zone of proximal development" (Vygotsky, 1978) is too salient and too omnipresent a fact to be ignored: humans learn more efficiently with a skilled teacher. And this, as constructivists would also argue, is a domain-general fact which is to govern not only singing, drawing, cooking or bicycle riding but...all facets of natural language learning as well.

Principially for these "cognitive plausibility"-related reasons shall we attribute, in volume 2, a certain conceptual priority to supervised ML of evolutionary models of ontogeny of toddlerese. But before doing so, let's focus on that, which both supervised and unsupervised branches of ML have in common: evaluation.

Evaluation

It is in degree of sharing and conventionality of formal, quantitative and objective means of evaluation that science can be distinguished from art, and in lesser extent, also engineering from science. An NLP, as principially more a skill than a science, is not an exception in this regard. To paraphrase the same thing somewhat differently: it is not from existence of diverse means of evaluation, but from *partage commune* of the need to evaluate and knowledge of usefulness of such evaluations, from which the very productive unity of NLP stems.

Productive unity in the field there is, and diversity -luckily for field's survival- is there as well. Thus, what holds for already-mentioned diversity of NLP's learning methods, holds also for diversity of diverse evaluation metrics. This is so because there exists no wide agreement about the meta-criterion which could help to decide what criteria exactly should a good evaluation metrics fulfill. Hence, asides the fact that a good evaluation metrics should make the result of an arbitrary experiment as much comprehensible as possible even to an un-initiated greenhorn, and asides the observation that there, verily, exist evaluation metrics which *describe* certain classes of phenomena better than others, it should not be a priori accepted that there exists "the" evaluation metrics which is the best of all.

Things being as they are, the NLP and "information retrieval" communities often tend to use the traditional evaluation formulas for Precision and Recall:

Precision and Recall (DEF)

$$\mathrm{Recall} = \frac{\text{Number of retrieved relevant entities}}{\text{Total number of all relevant entities}}$$

$$\mathrm{Precision} = \frac{\text{Number of retrieved relevant entities}}{\text{Total number of all retrieved entities}}$$

PRECISION AND RECALL END 10.3.1.0

whereby the relevancy of the "relevant" document is defined to the external, ideally manually annotated *étalon* (i.e. golden standard), corrected by a human judge and subsequently furnished to LSD by the teacher or evaluator. Precision thus, in certain sense, carries information about how much is the set X, retrieved by the algorithm, stained with "false positives" which do not belong to X according to the golden standard. Recall, on the other hand, carries information about how many among the entities which are labelled as "true" in the golden standard, were selected (i.e. labeled as "positives") by the algorithm.

Values of the always are constrained in the interval [0,1] and can be further combined into their "harmonic mean", commonly known as F-Score:

$$F = 2 * \frac{\mathrm{precision} * \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}$$

which also yields a score from interval [0,1] whereby 0 is obtained by the worst possible and 1 by the ideally performing algorithm.

This being said, it should be evident that precision and recall are concepts useful especially in case of binary classification tasks, i.e. tasks in which one aims to categorize certain set of entitities into two groups (i.e. a is X or not-X). Given that the notion of binary distinction is indeed a powerful one it is not uncommon that some studies succeed to get crowned with laurel, thanks to some additional averaging, even when they use use precision & recall based metrics also for evaluation of pure multiclass classification problems, i.e. problems where one aims to categorize certain set of entitites into N>2 groups, or clusters.

Different measures were developed which target specifically the problem of multiclass clustering. The most traditional among these being purity, defined as:

$$\text{Purity}_{\Omega,C} = \frac{1}{N}\sum_{k} max_j|\omega_k \cap c_j|$$

where $\Omega = \omega_1, ..., \omega_K$ is the set of K clusters hypothetized by the LS, members $C = C_1, ..., C_N$ denote N classes present in the golden standard. During estimation of purity each among K hypothetized clusters is assigned to the class which is most frequent in the cluster. The accuracy of assignment is subsequently assessed by counting the number of correctly assigned documents and dividing by number of gold-standard classes (N). Similiarly to all notions closely introduced in this section, ideal results have value of 1 while bad results shall be close to 0.

Purity asides, litteraly dozens other measures for clustering accuracy performance have been already developped, see Rosenberg and Hirschberg (2007) for overview of most important among them. The same article also a measure called V-measure defined as:

V-measure (DEF)

$$h = 1 - \frac{H(\Omega|C)}{H(\Omega)}$$

$$c = 1 - \frac{H(C|\Omega)}{H(C)}$$

$$V = \frac{(1+\beta)hc}{(\beta h) + c}$$

where H(C) denotes entropy of collection of classes; H(Ω) denotes entropy of collection of hypothetized clusters; H(C | Ω) denotes conditional entropy of C given Ω and H(Ω | C) denotes conditional entropy of Ω given C; and β specifies the weight between the h and c[8].

V-MEASURE END 10.3.1.0

Asides the fact that its values are also from the values of interval [0, 1], V-measure disposes of multiple properties which makes it worthy of interest for anyone willing to use an elegant measure

8 β is often set to 1 in order not to bias the value of V neither towards homogeneity, nor towards completeness

of cluster evaluation. Not only is V-measure a harmonic mean of h (also called "homogeneity") and c ("completeness") and is thus strongly reminiscent of F-score, but it has also property of being stable in regards to variation of number of clusters. For these, as well as other reasons more closely elucidated in Rosenberg and Hirschberg (2007); Christodoulopoulos et al. (2010); Hromada (2014a), shall be V-measure used in "part-of-speech induction" chapter of the 2nd volume of this Thesis .

In order to work, Recall, Precision, F-score, Purity and V-measure require the golden standard which has, in NLP, often the form of a manually annotated corpora. These measures, based on "external criteria" must not be, *ex vi termini*, used to modulate the execution of an unsupervised learning process. In learning scenarios in which the only source of knowledge is pure non-annotated dataset, one is obliged to evaluate the clustering only according to criteria inherent in the dataset itself. Many such "internal criteria" have been already discussed in the litterature (e.g. silhouette coefficient, Dunn index, Davies-Bouldin index), one more - the "prototypicity coefficient" shall be introduced in volume 2.

Let's now now move forwards with just one little warning: in no way does the sketchy overview hereby presented pretends to be a complete overview of NLP evaluation techniques, let alone the learning methods themselves. Given the amount of research being done in the domain, this is simply impossible. Thus, in order to restrict this expose to reasonable length, topic of evaluation of continuous, i.e. "regression" ML models was completely set aside and all attention was concentrated upon the evaluation of ML algorithms which tend to "learn" models composed of two or more discrete categories. This design choice was mainly motivated by a belief that it is more reasonable to aim to explain functioning of language cognition in terms of categorization, and not in terms of regression[9].

EVALUATION END 10.3.1.0

At last but not least, it is important to mention that machine learning is able to yield programs and applications which work, and work very well. And it is indeed especially NLP which is, asides "computer vision" [10] a field in and for which the ML is developped. It is thus not too surprising that recent days have seen, for example in article of Karpathy and Fei-Fei (2014), results of some quite successful efforts to unite the two.

9 None that in reasoning that shall follow, operations acting upon continuous domains are not to be completely excluded. Take as an example the notions of 1) temporal half-life (i.e. decay interval) of a cognitive schema 2) selection of locally-nearest-neighbor according to similarity defined in cosine metrics.

10 C.f. (Hromada et al., 2010) for an older application of ML methodology in training of smile-detection classifiers.

MACHINE LEARNING END 10.3.1

ML-inspired methodologies for:

1. problem of ontogeny of semantic categories (equivalent to supervised learning of word meanings)
2. problem of ontogeny of morphosyntactic categories (also known as part-of-speech induction)
3. problem of ontogeny of grammars (also known as grammar induction)

shall be described in closer detail in following sections, as well as in Volume 2.

END NATURAL LANGUAGE PROCESSING 10.3

10.4 SEMANTIC VECTOR ARCHITECTURES

It was already (9) mentioned, that natural language furnishes a communication channel for exchange of meanings. Meaning («signifié») is intentional, it refers to some external entity («referent») . Within the language L, meaning M can be denoted by a token («signifiant») and it is by exchange of physical (phonic, in case of spoken language, graphemic in case of written language etc.) manifestations of these tokens that producer (speaker | writer) and reciever (hearer | reader) communicate.

Signifier, signified, referent

Traditionally meaning of the word, i.e. its «semantics», was often considered as something almost «sacred» and not-to-be-formalized by mathematical means. Maximum which could be done - and had been done since Aristotle until middle of 20th century - was to define concept in terms of lists of «necessary and sufficient features».

Necessary and sufficient

Two types of features were considered to be both necessary and sufficient for definition of majority of concepts : first specifying concept's genus (or superordinated concept) and second specifying the particular property (differentia) which distinguished the concept from other members of the same genus. Thus, for example, «dog» could be defined as domesticated (differentia) canine (genus). Important property of such system of concepts was, that it allowed no ambigous or fuzzy border cases : the logical «law of excluded middle» guaranteed that all entities which were not both canines and domesticated at the same time (e.g. a chihuahua which passed all her life in wilderness) could not be called a dog.

Aristotelic paradigm of word meaning

Even in contemporary CL practice, projects like WordNet (Miller, 1995) incarnate such aristotelic view in form of datasets organizing items of human lexicon in what is principially an arborescent hierarchy of sub- and super- ordinated terms (i.e. of hyponyms and hyperonyms).

The change of the classical paradigm came slowly with works of late Wittgenstein [11] but especially with empirical studies of Eleanore Rosch. What these studies (e.g. Rosch (1999)) found out, was that not only are concepts often defined by bundles of features which are neither necessary not sufficient but also that the degree with which a feature can be associated with a concept often varies. Subsequently, Rosch has proposed a «prototype theory» of semantic categories whose basic postulate is, that some members of the category (or some instances of the concept) can be more «central» in relation to the category (resp. concept) than others. Thus, in some cultures "rose" is more "flower" than "daisy", in other cultures contrary is the case.

Non-aristotelic paradigm

10.4.1 CATEGORY PROTOTYPE (DEF)

A prototype P of the category C is a member of C, which shall be retrieved with highest probability whenever one queries C for its most salient concrete representative.

Such a member of C is to be as similar as possible to all other members of C and as dissimilar as possible from members or prototypes of other categories. CATEGORY PROTOTYPE END 10.4.1

Prototypical theory as well as other both theoretic and empirical advances like formalization of notion of similarity, in combination with development of information-processing technologies, have paved the way to operationalization of semantics which allows to transform meanings of words into mathematically commesurable entities.

In modern semantics, concepts are operationalized as geometric entities. Thus, meaning of a token X observable within language corpus C is often characterized as a vector of relations which X holds with other tokens observable within the corpus. The set of such vectors associated to all tokens observable in C yields a «semantic space» which is a vector space within which one can effectuate diverse numeric and | or geometric operations.

Geometrization of meaning

Since a methodological objective of this disertation is to bridge developmental psycholinguistics with the computational one, we consider it to important to underscore that in NLP practice, transformation of corpus C into semantic feature space S is practically always based on empirical validity of "distributional hypothesis" (DH) which states that « a word is characterized by the company it keeps» (Harris, 1954) [12]

Distributional hypothesis

11 « For a large class of cases of the employment of the word 'meaning'—though not for all—this way can be explained in this way: the meaning of a word is its use in the language» (Wittgenstein, 1953)

12 DH can be also restated in somewhat more algebraic terms:« In the most simple case can be the vector which denotes concept X calculated as a normalized linear combination of vectors of concepts in context of which X occurs.» (Hromada, 2014d)

Practical usefulness of DH in practically all models of geometric operationalization of meaning is undisputable. But DH has also non-negligable theoretical importance. For stated as it is, it supports «associationist» theories based on the notion that the essence of mind is somehow related to mind's ability to create relations, i.e. associations, between successive states.

In addition to what was said in (9.4.1, we suggest that both mind's faculty to create associations, as well as the distributional hypothesis "*meaning of symbol X can be defined in terms of meanings of symbols with which X co-occurs*", can be neurologically explained in terms of already-mentioned Hebb's postulate:

Hebb's law

« The general idea is an old one, that any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated', so that activity in one facilitates activity in the other» (Hebb, 1964)

One can assume that IF

1. Hebb's rule govern activity of not only single neurons but also of neural ensembles

2. if distinct words W_x and W_y are somehow processed and represented by distinct neural ensembles N_x and N_y

THEN it shall follow that whenever a hearer shall hear (or speaker shall speak) the two-word phrase W_xW_y, the ensemble of material (synaptic?) relations between N_x and N_y shall get reinforced. In more geometrical terms, on a more « mental » level, such a « rapprochement » of N_x and N_y would be characterized by convergence of the geometrical representations of both circuits to their common geometrical centroid. Thus, after processing the phrase W_xW_y, the vectorial representations of both N_x and N_y will be closer to each other than before hearing (or generating) the phrase.

Associanist geometry

10.4.2 HEBB-HARRIS ANALOGY (APH)

For a corpus linguist, distributional hypothesis means, mutatis mutandi, the same thing as Hebb's law for a neuroscientist.

END H-H APHORISM 10.4.2

We conjecture that an associationist principle, similar to the one described above, is indeed at work whenever a mind projects stimuli percieved from the external world unto an internally represented semantic space. Such «semantic vector space» can be subsequently divided, partitioned or tesselated into diverse subspaces each of which represents diverse semantic categories, classes or concepts. Or maybe even more than *just* represent: such partitions *are* concepts.

Of concepts and subspaces

The big advantage of approaches modelling the « geometry of thought» (Gärdenfors, 2004) is that they allow, among other things, to measure and assess similarities and distances between two or more concepts. By doing so, they seem to be much more closer to actual human experience with meanings than other computational methods (expert systems, ontologies, RDF etc.), based principially on application of logical rules of inference. For programs which work with concepts as if they were geometrical entities have no problem whatsoever to answer questions like

Conceptual similarity

"what is more similar to a dog - a cat or a wolf?".

Such questions -which any child would love to answer- couldn't be answered by an expert system without intervention of human operator who would explicitely declare the criterium of similarity according to which the similarity is assessed. But a system considering all three terms -"dog", "cat", "wolf"- as being just labels denoting geometrical points, would not have problem to do so if ever it was already confronted with corpus in which the three terms occured.

And given the fact that these geometric models make it possible to calculate, evaluate or compere similarities between meanings, it is of no surprise that these very models make it quite easy to create artificial simulations for such cognitively salient phenomena as analogies, metaphors Lakoff (1990) and intuitions.

Let's now glance on few such NLP models which process meanings as it they were geometric entities.

10.4.3 BAG-OF-TERMS

Bag-of-Terms (BoT) distinguish contained and containing entities. Most often, words are understood as the contained entities and sentences or whole documents are the containing ones. What is important for such Bag-of-Words (BoW) models is that the document D_1 contains certain set of words while the document D_2 contains another set of words.

Such quantitative information about the number of occurences of diverse words in diverse documents can be used to construct vectorial representations of such documents. This is done by representing every distinct document with a row vector whose specific elements denote specific words. Table 7 illustrates this for three sentences[13], considered as individual documents.

The order of words or other aspects (e.g. morphosyntax, phonology, prosody) are considered as irrelevant: in pure BoW, it is only the occurence of the word that counts. This, however, is not necessary the case in BoTs which implement another definition of the "contained

13 Sentences like these (meaning "mama has ema", "ema has mama" and "mama has mama") are often among the first used in Slovak language primers.

	mama	má	emu	ema	mamu
mama má emu	1	1	1	0	0
ema má mamu	0	1	0	1	1
mama má mamu	1	1	0	0	1

Table 7: Vectorial representations of three sentence-sized documents. Every distinct word yields a distinct column.

entity" - i.e. of component term by means of which one characterizes the "containing" document. For one can also work with terms which are either smaller, bigger or utterly different from words. One can look for occurence of syllables or, simpler yet, a distinct sequence of N characters (an N-gram). Construction of vectorial representations based on occurence of 3-gram terms is presented on Table 8.

	"mam"	"ama"	"ma "	"a m"	"má "	"á e"	" em"	"emu"	"á m"	" ma"	"amu"
D_1	1	1	1	1	1	1	1	1	0	0	0
D_2	2	1	1	1	1	0	0	0	1	1	1

Table 8: Vectorial representations of sentence-sized documents D_1 = "mama má emu" and D_2 = "mama má mamu". Every distinct character trigram yields a distinct column.

In this case, one can see that some information about the word order is also included into the vectorial representation. This is so, because the word-dividing empty space character " " is also taken into account which was not the case in pure BoW presented in Table 7. On the other hand, by focusing on trigram features and not on whole words, one may observe a feature "mam" to occur twice in document D_2. Hence $X_{2,1} = 2$.

No matter what definition of documents and term one uses, one obtains, at the end, a list of N D-dimensional row vectors where N is the number of documents in the corpus and D is the number of distinct tokens observed in the corpus. One thus obtains a term-document matrix X. In NLP practice, it is common and recommendable to process the resulting values of such matrix to so-called *term frequency–inverse document frequency* (tf-idf) weighting scheme.

TF-IDF

Let $t_f(t, d)$ denote the **term frequency**, i.e. number of times the term t occurs in document d and let $idf(t, D)$ denoting the **inverse document frequency** be obtained as follows:

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in D\}|}$$

where N denotes the total number of documents in the corpus and $|\{d \in D : t \in D\}|$ denotes the number of documents in which t occurs.

Then term frequency–inverse document frequency (tf-idf) is to be calculated as follows:

$$\mathtt{tfidf}(t, d, D) = \mathtt{tf}(t, d) * \mathtt{idf}(t, D)$$

in order to yield a numerical weight reflecting how important a word is to to a document contained in a corpus.

TF-IDF END 10.4.3.0

Verily is tf-idf a very simple yet very effective means how an NLP engineer can increase the accuracy of one's vectorial model. But it has also some disadvantages. Primo, it adds a second pass to construction of term-document matrices which can, especially in case of BigData NLP, bring about certain computational and memory costs. Secundo, the cognitive plausibility of tf-idf models is still to be demonstrated. In other terms: while practically whole history of NLP empirically demonstrates that tf-idf represents an information-processing component wherein statistical properties of the whole influence weights of individual associations, current psycho-linguistic knowledge seems to fail to identify a cerebral mechanism functioning as tf-idf's neural correlate.

Be it as it may, tf-idf brings even more order and information into the metric space given by the entities represented by the term-document matrix. And given that these entities are already of numeric, quantified nature, they can be commesurated. Distance between words can be obtained by measuring distances between two column vectors; distance between documents can be obtained by assessing distances between two row vectors. Multiple metrics (e.g. Jaccard index, Euclidean distance, cosine for real-valued vectors, Hamming distance for binary etc.) are used in order to do so.

BAG-OF-TERMS END 10.4.3.0

10.4.4 Latent Semantic Analysis (TXT)

A major disadvantage of term-document occurrence matrices, as generated by BoW models, is their sparsity. Given, for example, a corpus containing N=1 million documents and M=50000 distinct terms, BoW postulates existence of a rectangular term-document matrix with fifty billion elements. And given that only a relatively small subset of distinct words shall occurs in any specific document, vaste majority of values in such a matrix shall be zero.

Latent Semantic Analysis (LSA) was one among the first solutions aiming to address this sparsity problem in NLP scenario. By unfolding the formula, known in algebra as singular value decomposition (SVD):

$$X = U\Sigma V^T$$

it transorms the original term-document matrix X into orthogonal matrices U and V and a diagonal matrix Σ. By selecting D values from Σ and vectors of U and V associated with these values, one can reduce the dimensionality of original matrix X to only D dimensions, with minimal smallest error.

Algebraic and dimensionality-reduction aspects aside, LSA was, in its time, revolutionary for one principal reason: it allowed to compare not only documents with documents and terms with terms, but also terms with documents. It also allowed for a means of optimization: one could tune model's performance by modifying the dimensionality [14]. Feats furnished by LSA were, at the time of its conception, so astounding that LSA's conceptors considered their model to be the answer to the problem of category induction and antique problem concerning the essence of knowledge in general, hence promoting their computational model to status of « a solution to Plato's problem: latent semantic analysis theory of knowledge» (Landauer and Dumais, 1997).

LSA is indeed able to furnish dense, low-dimensional vector spaces of semantic categories and concepts. It seems to yield interesting solutions for dozens of other problems, let's mention, as an example, the problem of grapheme-to-phoneme in speech synthesis (Bellegarda, 2005). And it is also true that transition through the site http://lsa.colorado.edu has been and is - for at least one generation of all sorts of cognitive science students - an important, useful, and potentially obligatory *rite of passage* of their academic parcours.

But it is also true that LSA has certain drawbacks. Computationally speaking, LSA is costly because SVD is costly. And cognitively speaking, it is somewhat difficult to see how human brain could perform such a precise deterministic operation like SVO, let alone the dimensionality optimization which should precede it[15]. As LSA's conceptors put it: « It still remains to understand how a mind or brain could or would perform operations equivalent in effect to the linear matrix decomposition of SVD and how it would choose the optimal dimensionality for its representations, whether by biology or an adaptive computational process.» (Landauer and Dumais, 1997)

We propose to address the problem by simply ignoring the SVD altogether and rather focusing on another means of dimensionality reduction: the random projection.

14 According to (Landauer and Dumais, 1997), an optimal dimensionality for problem of concept induction from English language corpora is approximately 300

15 Note that the dimensionality optimization could have occured during development, either phylogenetic or ontogenetic, or both.

LATENT SEMANTIC ANALYSIS END 10.4.4

10.4.5 RANDOM INDEXING (TXT)

Random Indexing (RI) is a method of representation of textual corpora with dense, low-dimensional vector spaces. In theory, RI is justified by a lemma of Johnson-Lindenstrauss whose corollary « states that if we project points in a vector space into a randomly selected subspace of sufficiently high dimensionality, the distances between the points are approximately preserved» (Sahlgren, 2005). In more formal terms, dimensionality of rxc-dimensional term-document occurence matrix X can be reduced by projection through rxd-dimensional random matrix R, whereby the target number of dimensions (d) is the parameter of the projection, and is smaller than the initial number of columns (i.e. $d \ll c$):

$$X'_{rxd} = X_{rxc} R_{cxd}$$

In NLP practice, the simplest yet quite efficient variant of creation of such slighthly distorted d-dimensional matrix X' is implemented by a following procedure: « Given the set of N objects (e.g. documents) which can be described in terms of F features (e.g. occurence of the string in the document), to which one initially associates a randomly generated d-dimensional vector, one can obtain d-dimensional vectorial representation of any object X by summing up the vectors associated to all features F_1, F_2 observable within X. The original random feature vectors are generated in a way that out of d elements of vector, only S among them are set to either -1 or 1 value. Other values contain zero. Since the "seed" parameter S is much smaller than the total number of elements in the vector (d), i.e. S «d, initial feature vectors are very sparse, containing mostly zeroes, with occasional value of -1 or 1.» (Hromada, 2014c). The PERL Data Language (PDL)-compliant source code of the procedure is presented in Listing 5.

Listing 5: Random Indexing Source Code

```
my %doc_vectors;
my %term_vectors;
sub generate_initvector {
        my $value;
        my %set;
        my $vec=zeroes $dimensions;
        for (0..$seed) {
                (rand >0.5) ? $value=1 : $value=-1;
                my $offset=round(($dimensions-1)*rand);
                while (exists $set{$offset}) {
                        $offset=round(($dimensions-1)*rand);
                }
```

```
                $set{$offset}=$value;
                index($vec,$offset).=$value;
        }
        return $vec;
}
for $document (@document_list) {
        my @words=split(/[^\w]/,$document);
        for my $word (@words) {
                if (!exists $tvectorz{$word}) {
                        $term_vectors{$word}=generate_initvector;
                }
                $doc_vectors{$document}=zeroes $dimensions if  !
                    exists $doc_vectorz{$document};
                $doc_vectors{$document}+=$term_vectors{$word};
}
```

Simply stated, vectorial representation of documents A is obtained as simple linear combination[16] of initial vectors associated to terms T_1, T_2, T_3 observable in A. For any such term, an $d - dimensional$ initial vector is randomly generated and contains $d - S$ zero elements and S elements whose value is either -1 or 1.

The output of this simple variant of RI is a set of $d - dimensional$ document vectors which can be used to calculate similarity among the documents. Normalization of these vectors is needed when one uses the cosine metrics. But one can go further: for one can additionaly "reflect" the whole process, forget the random vectors (initially attributed to individual terms) and now calculate the vectorial representation of the term T_x as a linear combinations of documents in which T_x occurs. After 2 or 3 iterations[17] of such "reflection of information" from documents to terms and vice versa, one obtains numeric representations of both documents and terms both projected into one holistic metric space. Thus, in the spaces generated by Reflective Random Indexing (RRI) (Cohen et al., 2010), *there is no distinction of essence* between words and documents or, more generally, *between objects and the context of their use*. All can be understood as points or vectors of the same $d - dimensional$ space. Not only that, such geometric entities can be also interpreted in terms of subspaces: one can speak about the region whose centroid is the entity, E or one can speak about subspaces orthogonal to E's vector. The world of meanings once thus geometrized, verily many are applications of such "vector symbolic architectures" (Widdows and Cohen, 2014).

16 Weighting the term vectors with related tf-idf values is strongly recommended.

17 Note that due to convergence properties of random projection, more than 2 or 3 iterations of the reflective process often tend to degrade the accuracy of RI's semantic discrimination. On the other hand, multi-iterative convergence of associanist matrices yields highly useful results in other NLP tasks, including the estimation of the "importance of the sign" (Hromada, 2009) commonly known as PageRank (Hromada, 2010a).

RANDOM INDEXING END 10.4.5

10.4.6 LIGHT STOCHASTIC BINARIZATION

Raison-d'etre of all semantic space architectures is information | knowledge retrieval. No matter whether one encodes one's dataset in form bag-of-words, LSA, RI or RRI vectors, the objective is often the same: to implement the model in real-life applications which are able to identify members of the dataset which are semantically closest to some user-specified query. And to do so in reasonable time.

Thus, the computational complexity of retrieval phase at least as important as the computational complexity of the indexing (encoding) phase. Moreso in the BigData scenario where one aims to find a needle in a haystack of billions of documents. In case of data of very low dimensionality ($d < 10$), the solution is quite straightforward: one can one's data and create indices for it, by use of binary trees or other indexing techniques[18].

Unfortunately, because of a so-called "curse of dimensionality", it is practically impossible to create retrieval indices for entities with higher dimensionality. In layman's terms this is so because two entities close to each other in many dimensions can still be considered far from each other (because being really far from each other in just few dimensions); or because two entities far from each other in some dimensions can still be considered relatively close to each other (because they are quite close in many other dimensions). Thus, in huge-dimensional spaces, usage of indices (e.g. k-trees) in retrieval can sometimes turn out to be more costly than a simple "linear" search in which one compares one's query with all vectores stored in the dataset.

Given that the complexity of such linear search is $N * d$ and given that one cannot reduce the size of one's dataset (i.e. N) and given that one accepts that the "curse of dimensionality" is inevitable in semantic spaces, one can still fasten, *in silico* the retrieval by at least two possible means:

1. construct semantic spaces smallest possible (yet still sufficiently high to encode semantically relevant distinctions) dimensionality d
2. execute operations with binary vectors (instead of integer, float or complex ones)

Combination of these two means into one algorithm yields Light Stochastic Binarization (LSB).

18 Dataset indexing is often explained in terms of a huge library with one shelf containing a sorted cartotheque of cards which specify book's position in the libary

The idea behind LSB is fairly trivial and is inspired by approaches like Locality Sensitive Hashing (LSH, Datar et al. (2004)) or Semantic Hashing (SH, Salakhutdinov and Hinton (2009)). In these hashing approaches, the objective is to use a "hashing function" able to attribute a short and concise binary vector (i.e. "a hash") to any document in the dataset in a way that if two documents are similar (or identical) their hashes will also be similar (or identical). In this sense, LSB can also be understood as a sort of hashing algorithm which simply uses the Reflected Random Indexing (10.4.5) as its hashing function.

Once RRI transforms document (or a query) Q is transformed into its vectorial representation $\vec{q}$ and whose n−th element we denote with $\vec{q}_n$, one obtains the resulting binary hash $\vec{h}$ by trivial thresholding:

$$\vec{h}_n = \begin{cases} 0 & \vec{q}_n < 0 \\ 1 & \vec{q}_n >= 0 \end{cases}$$

Expressed verbally, when value generated by RRI is bigger than 0, one shall put 1 into respective position of the binary hash, otherwise one puts 0[19]. At its very core, it is nothing else than mapping of RRI's output integer | float range onto the binary range. A mapping which exploits a mathematically beautiful intuition of Sahlgren (2005) that the random projection - as performed by RI and RRI - should be seeded solely with values of -1 and 1.

The study (Hromada, 2014c) has indicated that in case of classification scenarios where low recall is allowed if high precision is attained, LSB yields results comparable (or better) than both binarized LSA and renowned deep-learning technique proposed by Salakhutdinov and Hinton (2009). Figure 19 displays these results for the problem of multiclass classification (C=20). All models thereby represented used dimensionality $d = 128$ and the size of a document hash was thus exactly 16 bytes.

LIGHT STOCHASTIC BINARIZATION END 10.4.6

10.4.7 EVOLUTIONARY LOCALIZATION OF SEMANTIC ATTRACTORS

Reflective procedure asides, LSB involves neither optimization nor machine learning components. But given that it produces simplest datastructures possible - id est, low-dimensional binary vectors - it can be easily embedded into more complex frameworks. Evolutionary Localization of Semantic Attractors (ELSA, Hromada (2015)) aims to change it.

19 This trivial thresholding is applicable only in case of huge (BigData) corpora where law of big number applies. C.f. Hromada (2014c) for LSB's variant usable in cases of smaller corpora.

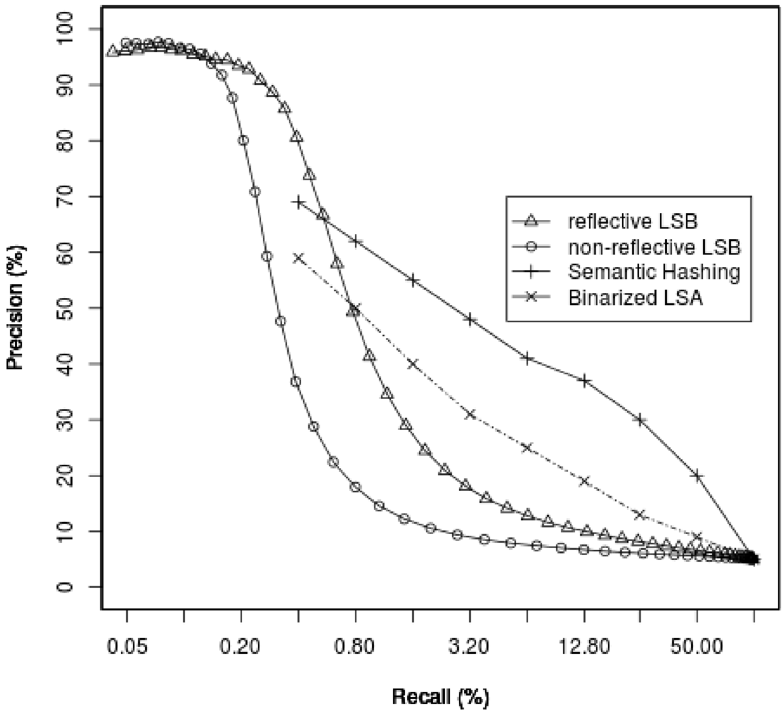


Figure 19: Comparison of reflective LSB (with I=2 iterations) and unreflective LSB (I=0) LSB with Semantic Hashing and binarized Latent Semantic Analysis. Reproduced from Hromada (2014c).

ELSA is a result of embedding the LSB into an evolutionary computation framework. More concretely, ELSA uses canonic genetic algorithms (8.7.1) to localize a set of category prototypes (10.4.1) best adapted to document classes encoded in the training corpus. LESB thus aims to address the problem of supervised document classification abd as such expects to be trained with corpus containing documents and associated category labels. It first processes whole corpus by LSB algorithm and once documents are transformed into binary vectors, it starts to look for the most optimal set of category prototypes.

In ELSA, the search for category prototypes is equivalent to discovery of such a set of prototypes which minimize the function:

$$F(P) = \alpha \sum_{t \in c_P} H(t, P) - \beta \sum_{f \not\subset c_P} H(f, P) \quad (1)$$

whereby P denotes the vector representation of the prototype in the binary space, H denotes the Hamming distance[20], t denotes the vector representation of the "true" document belonging to same class (c_P) as the prototype, f is the vector of the "false" document belonging to some other class of the training corpus and α and β are weighting parameters.

Thus, a candidate prototype P of category c_x is considered to be most fit if it is as close as possible (i.e. has smallest Hamming distance) to all documents which are attributed to c_x in the training corpus; and as far as possible from documents which are not attributed to c_x in the training corpus.

In ELSA, solution to multiclass classification problem is formalized as such group of prototypes which minimize the distance to members of categories they should represent and maximize the distance to others. Given that the training corpus divides its documents into $|C|$ classes, and given that every document and every prototype can be represented as a $d-$dimensional binary vector, chromoses which are to be optimized by ELSA are binary vectors of $|C| * d$.

The rest is in work in progress. C.f. Hromada (2015) for comparison of ELSA with binarized LSA, non-optimized LSB, or Semantic Hashing. Given that ELSA introduces in a one unified framework three components which we claim to be cognitively plausible, id est:

1. dimensionality reduction by means of random projection
2. theory of semantic prototypes
3. evolutionary computation

and given that its binary nature predestines it to execute very fast on any transistor-based computer, we shall use aim to implement ELSA,

20 Hamming distance of two binary vectors h_1 and h_2 is the smallest number of bits of h_1 which one has to flip in order to obtain h_2.

in one way or another, in majority of simulations described in volume 2.

ELSA END 10.4.7

In this section we have presented multiple architectures which have all one thing in common: they succeed to transform textual documents into geometric and | or mathematical entities. To keep the overview as simple and concise as possible, only scalars, vectors and matrices were discussed; the reader is to be reminded that other mathematical models of semantics exists which also involve tensors of higher order. Even a very introduction of these, however, surpasses by far the objectives of this Thesis.

Thus, instead of closer discussion of fascinating topics like interrelations between "binding operators", "circular convolution", "complex numbers" and "quantum logic" (Widdows and Cohen, 2014), we have preferred to acquaint the reader with the idea that *meanings are subspaces of* d*—dimensional semantic spaces*. Departing from simple word-document occurence matrices of first bag-of-words models, passing through LSA's ambitions to answer perennial questions:

What are ideas, how are they stored and how are they accessed ?

and discussing other more natural means of dimensionality reduction, we finally approach the Point where multiple divergent streams converge into one. But before exploring it somewhat further, let's see whether the realms of semantic and syntactic categories does not have something in common. In computational sense, for example.

SEMANTIC VECTOR ARCHITECTURES END 10.4

10.5 PART-OF-SPEECH INDUCTION

POS-i

The term Part-of-speech-induction (POS-i) designates the process which endows the human or an artificial agent with the competence to attribute the POS-labels (like "verb", "noun", "adjective") to any linguistic token observable in agent's linguistic environment. POS-i can be understood as a « partitioning problem » since one's objective is to partition the initial set of all tokens occuring in corpus C (which represent agent's linguistic environment E) into N subsets (partitions, clusters) whose members would correspond to grammatical categories as defined by the gold standard. Because one does not use any information about « ideal » gold standard grammatical categories during the training phase and uses it only for final evaluation of the performance of the model, POS-i is considered to be an « unsupervised » machine learning problem.

POS-t

POS-i's « supervised » counterpart is the problem of POS-tagging. In POS-tagging, one trains the system by serving it, during the training phase, sequence of couples (word W, tag T) where tag T is the label denoting the grammatical category into which the word W belongs. POS-tagging is thus simpler than POS-i where no information about ideal labels is furnished during the learning. Training of POS-tagging systems is of particular importance especially for languages where many word forms can potentially belong to many part-of-speech categories (in English, for example, can almost any noun play also role of the verb; token like « still » can be intepreted as substantive, verb, adjective and even adverb (Páleš, 1994), its POS-category being determined by its context). On the contrary, in morphologically rich languages where such a « homonymy of forms » is present in lesser degrees and relations between word types and classes are less ambigous, one can often simply train the POS-tagging system by simply memorizing an exhaustive list of (W, T) couples.

10.5.1 NON-EVOLUTIONARY POS-I

The paradigm currently dominating the POS-i domain was fully born with article published by Brown and his colleagues in 1992 Brown et al. (1992). Brown and his colleagues have applied the information theoretic notion of « mutual information » M :

$$M(w_1, w_2) = \log \frac{P(w_1, w_2)}{P(w_1)P(w_2)}$$

upon all word bigrams (i.e. sequences of two tokens w_1, w_2 which co-occur with probability $P(w_1, 2_2)$ and had subsequently devised a merging algorithm able to group words into classes in a way that the mutual information within a class would be maximized.

In two decades since publication of study of Brown's study Brown et al. (1992), the word n-gram co-occurence approach has inspired hundreds of studies : be it hidden Markov Models tweaked with variational Bayes, Gibbs sampling morphological features, or graph-oriented methods – all such approaches and many others consider co-occurence of words with n-gram sequences to be the primary source of relevant information for subsequent creation of part-of-speech clusters. In all these models, one aims to discover the ideal parameters of Markovian statistical models, often employing a so-called Expectation-Maximization (EM) algorithm to discover the optimal partitioning. Unfortunately, EM is unable to quit locally optimal states once they were discovered. Notwithstanding this disadvantage, comparative study Christodoulopoulos et al. (2010) suggests that probabilistic models of part-of-speech induction can be indeed very performant.

POS-i induction can be also realized by means of k-means clustering algorithm, or one of its variants. K-means algorithm MacQueen et al. (1967); Karypis (2002) partitions N observations, described as vectors in D-dimensional space, into K clusters by attributing every observation into the cluster with the nearest centroid (i.e. mean). If one considers these centroids to denote prototypes of the categories in center of which they are located, then one can consider the k-means algorithm to be consistent with « prototype theory of categorization », as proposed by Rosch. Table 9 illustrates simple K-mean partitioning of tokens present in English version of Orwell's 1984, as contained in Multext-East Erjavec (2004).

CLUSTER	NOUNS	VERBS
0	10	3
1	568	67
2	97	668
3	13	**1011**
4	**1173**	67
5	608	958
6	**1977**	97

Table 9: K-means clustering of tokens according both suffixal and co-occurence informations. Table partially reproduced from Hromada (2014b).

In this example case we have clustered all tokens observable in the corpus into 7 clusters according to features both internal to the token – i.e. suffixes[21] – and external – i.e. co-occurrence with other tokens. Note that even such a simple model where no machine learning or optimization were performed, K-means algorithm somehow succeeds to distinguish verbs from nouns. As is shown in the Table 9, whose columns represent the "gold standard" tags and rows denote the artificially induced clusters, even such a naïve computational model has assigned 83.6% of nouns to clusters 1, 4 and 6 while assigning 91.8% of verbs into clusters 2, 3 and 5.

NON-EVOLUTIONARY POS-I END 10.5.1

10.5.2 EVOLUTIONARY

Usage of evolutionary computing in NLP is - in comparison to other methods like neural networks, Hidden Markov Models, Conditional Random Fields or SVMs – still very rare. This is also the case to NLP's sub-problem of part-of-speech tagging and thus we are not aware any tentative resolve the POS-i problmem with evolutionary means, and of only one tentative to use genetic algorithms to train a part-of-speech tagger:

21 That suffixes are of particular importance for POS-induction is more closely demonstrated in our article Hromada (2014a).

In his Araujo (2002) proposal, Araujo describes a system of POS-t involving crossover and mutation operators. What is particularly interesting about Araujo's system is that separate evolution process is run for every separate sentence of the test corpus. Training corpus, on the other hand, serves mainly as a source of statistical information concerning co-occurrences of diverse words and tags in diverse word & tag contexts. This information concerning the « global » statistic properties of the training corpus is later exploited in computation of fitness.

Let's take, for example, the phrase « Ring the bell ». Since words like « ring » and « bell » are in English sometimes used as verbs, and sometimes used as nouns, such a sentence can be tagged at least in 4 different ways :

N D[22] N
V D V
N D V
V D N

Such sequences of tags yiels individual members of Araujo's initial population of chromosomes. In languages like English where almost every word can be attributed to more than one POS category & the number of possible tag sequences therefore increases with length of the phrase-to-be-tagged, one will be most probably obliged to randomly choose such initial individuals. Fitness of every individual possibly tagging the sentence of n words is subsequently calculated as a sum of accuracies of tags (genes) on position i :

$$\sum_{i=0}^{n} f(g_i)$$

Accuracy gi of an individual gene is calculated as :

$$f(g_i) = \log\left(\frac{context_i}{all_i}\right)$$

whereby values of $context_i$ and all_i are extracted from the training table which was constructed during the training phase and represent the overall frequency of occurrence of word wi within specific ($context_i$) and all (all_i) contexts.

Once fitness is evaluated, fitness-proportional crossing-over (50%) and mutation (5%) is realized. Notwithstanding the fact that Araunjo doesn't seem to have used any other selection mechanism, in less than 100 generations, populations seemed to converge into sequence of tags which were more than 95% correct in regards to gold standard. This is a result comparable to other POS-tagging systems but

22 The non-terminal symbol D denotes the category of determiners containing such elements as articles "the", "a / an" etc.

with lesser computational cost. It is also worth noting that Araujo's experiments indicate that working solely with contextual window W_L, W, W_R , i.e. just looking one word to the left and one word to the right, seems to yield, in case of POS-tagging of English language higher scores than extracting data from larger contextual spans.

When it comes to the «unsupervised» variant of the POS-t problem, id est the problem of Part-of-speech induction, up to this date there have been -as far as we know - no tentatives to address the POS-i problem by means of evolutionary computing. For this reason, we shall aim propose our own solution in volume 2.

EVOLUTIONARY POS-I AND POS-T END 10.5.2

POS-I AND POS-T END 10.5

10.6 GRAMMAR INDUCTION

Input of Grammar Induction (GI) process is a corpus of sentences written in language L, its output is, ideally a grammar (i.e. a tuplet G=S,N,T,P as defined in 10.2) or a language model able to generate sentences of L, including such sentences that were not present in the initial training corpus.

The nature of resulting grammar is closely associated to the content of the initial corpus as well as to the nature of the inductive (learning) process. According to their « expressive power », all grammars can be located somewhere on a « specificity – generality » spectrum. On one extreme of the spectrum lies the grammar having following production rules :

$$1 \rightarrow 2*$$

$$2 \rightarrow a|b|c \ldots Z$$

whereby * means « repeat as many times as You Want ». This very compact grammar can potentially generate any text of any size and as such is very general. But exactly because it can accept any alphabetic sequence and thus does not have any « discriminatory power » whatsoever, is such a grammar completely useless as an explication of system of any natural language. On the other extreme lies a completely specific grammar which has just one rule :

$$1 \rightarrow < \text{corpus} >$$

This grammar contains exactly what corpus C contains and is thus not compact at all (it is even two symbols longer than C). Such a grammar is not able to encode anything else than the sequence which was literally present in the training corpus and is therefore also useless for any scenario were novel sentences are to be generated (or accepted).

The objective of GI process is to discover, departing solely from corpus C (which is written in language L), a grammar which is neither too specific, nor too general. If it is too general, it shall «overregularize» (9.2.4), i.e. shall be able to generate (or accept) sentences which the common speaker of L wouldn't consider as grammatical. If it is too specific, it shan't be able to represent all sentences contained in C or, if it shall, it shan't be able to generate (or accept) any sentence which is considered to be sentence of L but was not present in the initial training corpus C.

10.6.1 EXISTING NON-EVOLUTIONARY APPROACHES

One of the first serious computational models of GI is Wolff's «Syntagmatic – Paradigmatic» (SNPR) model Wolff (1988). Its core algorithm is presented in Listing 6.

Listing 6: Outline of Processing in the SNPR Model (reproduced from Wolff (1988)

```
1. Read in a sample of language.
2. Set up a data structure of elements (grammatical rules)
    containing, at this stage, only the primitive elements of the
     system.
3. WHILE there are not enough elements formed, do the following
    sequence of operations repeatedly:
        BEGIN
        3.1 Using the current structure of elements, parse the
            language sample, recording the frequencies of all
            pairs of contiguous elements and the frequencies of
            individual elements. During the parsing, monitor the
            use of PAR elements to gather data for later us in
            rebuilding of elements.
        3.2 When the sample has been parsed, rebuild any elements
             that require it.
        3.3 Search amongst the current set of elements for shared
             contexts and fold the data structures in the way
            explained in the text.
        3.4 Generalize the grammatical rules.
        3.5 The most frequent pair of contiguous elements
            recorded under 3.1 is formed into a single new SYN
            element and added to the data structure. All
            frequency information is then discarded.
        END
```

We consider the SNPR model to be of particular importance because of its aim to explain the process of Grammar Induction as a sort of cognitive optimization : « The central idea in the theory is that language acquisition and other areas of cognitive development are, in large part, processes of building cognitive structures which are in some sense optimal for the several functions they have to perform» (Wolff, 1988). Wolff also associates his « cognitive optimization hypothesis » with Brown's «law of cumulative complexity » (c.f. RE-

FREF) which Wolff paraphrases in statement: « if one structure contains everything that another structure contains and more then it will be acquired later than that other structure» (Wolff, 1988).

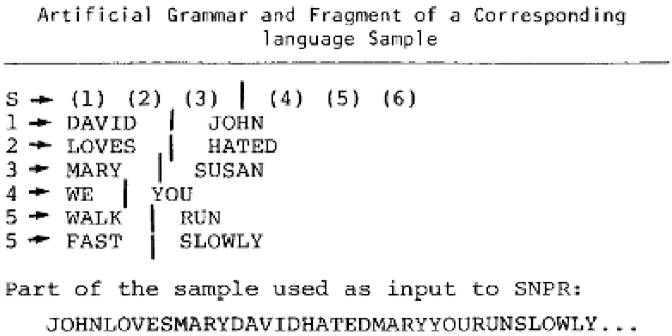


Figure 20: Equivalence classes and production rules induced from English language samples by ADIOS algorithm. Fig. reproduced from Wolff (1988).

Grammar resulting from such a contact between language sample and SNPR inducing mechanism is displayed on Figure 20.

In Wolff's theory optimalization is further understood as compression. Within the SNPR model is such compression realized in part 3.5 of his algorithm, where the most frequent pair of contiguous elements (either terminals or non-terminals) is substituted for a new non-terminal symbol. For this reason, the size of grammar able to generate the initial language sample ideally decreases with every cycle of model's « while » loop until the process converges to state where there is no redundancy to « compress ».

Wolff proposes that Grammar Induction is a process which should maximize the coding capacity (CC) of the resulting grammar while minimizing its size, i.e. its *Minimal Description Length* (MDL). He defines the ratio between grammar's CC/MDL to denote grammar's efficiency and it may be the case that within a more evolutionary framework where one would work with populations of grammars, a very similarly defined notion of efficiency could be used as the core component of the fitness function. Unfortunately, Wolff's 1988 SNPR model is not evolutionary since it does not involve any stochastic factors nor notion of multiple candidate solutions. SNPR is simply confronted with the language sample, deterministically compresses redundancies in a way that can sometimes ressembles human grammar (and sometimes not), gets subsequently stuck in local optimum and there's no way how to get out of it.

Another famous model of GI is that of Elman Elman (1993). Contrary to Wolff's algorithm which is principially « symbolic », is Elman's model « connectionist » one. More concretely, Elman had succeeded to train a simple recurrent neural network which was « trained to take one word at a time and predict what the next word would be. Because the predictions depend on the grammatical structure (which may involve multiple embeddings), the prediction task forces the network to develop internal representations which encode the relevant grammatical information.» (Elman, 1993).

The most important finding of Elman's study seems to be the evidence for a so-called «less is more hypothesis» which Elman himselfs labels with terms «importance of starting small»: « Put simply, the network was unable to learn the complex grammar when trained from the outset with the full "adult" language. However, when the training data were selected such that simple sentences were presented first, the network succeeded not only in mastering these, but then going on to master the complex sentences as welli» (Elman, 1993). Something similar occured also when he tuned the capacity of « internal memory » of his networks rather than the corpus itself. Elman observed: « If the learning mechanism itself was allowed to undergo "maturational changes" (in this case, increasing its memory capacity) during learning, then outcome was just as good as if the environment itself had been gradually complicated» (Elman, 1993).

Less is more hypothesis

Thus, not only results of Elman's computational model point in the same direction as many developmental and psycholinguistic studies of « motherese » (c.f. Section 9.3) ; they also show the importance of gradual physiological changes for ultimate mastering of maternal language. He goes even so far to state that prolonged infancy of human children can possibly go hand in hand with the fact that only humans develop language in an extent we do : « In isolation, we see that both learning and prolonged development have characteristics which appear to be undesirable. Working together, they result in a combination which is highly adaptive» (Elman, 1993).

Notwithstanding these interesting results which are not to be underestimated, we see two disadvantages of Elman's approach. Primo, as is often the case for connectionist neural networks, his resulting model is somewhat difficult to interpret : given the training constraints mentioned above, the network seems to predict quite well the next word in the phrase, but it is not evident why it does what it does. Elman himself dedicates major part of his article to descriptions of his tentatives to understand how his « blackbox » functions. Secundo, Elman confronted his model only with artificial corpora, i.e. corpora generated from manually created grammars. Thus, his model accounts only for a limited subset of properties of one language (English) and as such is still quite far from full-fledged solution to problem natural language's GI.

The model called « Automatic Distillation of Structure » (ADIOS) seems to be in lesser extent touched by this second disadvantage since, as Solan and his colleagues state: « In grammar induction from large-scale raw corpora, our method achieves precision and recall performance unrivaled by any other unsupervised algorithm. It exhibits good performance in grammaticality judgment tests (including standard tests routinely taken by students of English as a second language) and replicates the behavior of human subjects in certain psycholinguistic tests of artificial language acquisition. Finally, the very same algorithmic approach also is proving effective in other settings where knowledge discovery from sequential data is called for, such as bioinformatics.» (Solan et al., 2005)

ADIOS is a graph-based model. It considers the sentences to be a path in the directed pseudograph (i.e. loops and multiple edges are allowed), each sentence being delimited by special « begin » and « end » vertices. Every lexical entry (i.e. a word type) is also a vertex of the graph, thus if more than two sentences share the same word X, they cross themselves in the vertex V_X ; if they contain the same subsequence X_Y, their paths share the common subpath (edge) $V_X V_Y$ etc.

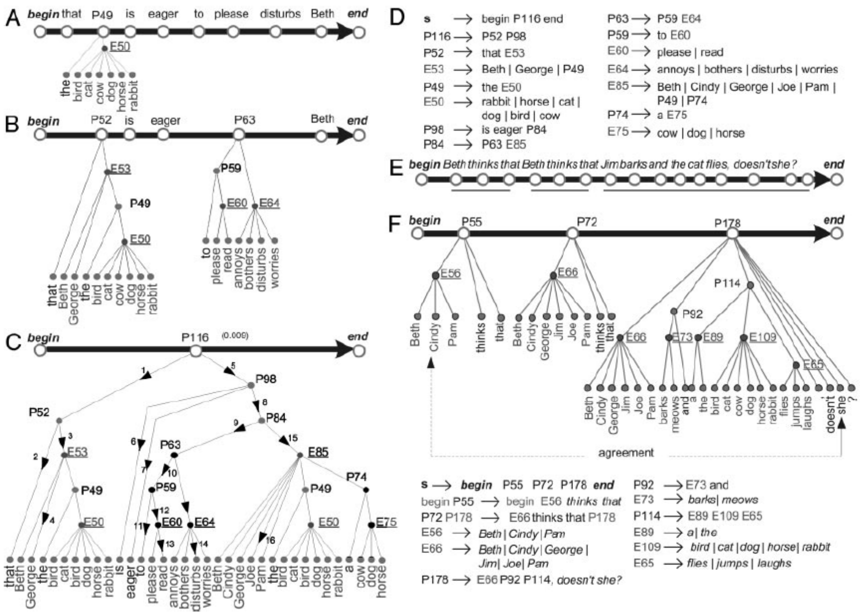


Figure 21: Equivalence classes and production rules induced from English language samples by ADIOS algorithm. Reproduced from Solan et al. (2005).

Authors of ADIOS describe their algorithm as follows : « The algorithm generates candidate patterns by traversing in each iteration a different search path (initially coinciding with one of the original corpus sentences), seeking subpaths that are shared by a significant number of partially aligned paths. The significant patterns (P) are selected according to a context-sensitive probabilistic criterion defined in terms of local flow quantities in the graph...Generalizing the search path, the algorithm looks for an optional equivalence class (E) of units that are interchangeable in the given context [i.e., are in complementary distribution]. At the end of each iteration, the most significant pattern is added to the lexicon as a new unit, the subpaths it subsumes are merged into a new vertex, and the graph is rewired accordingly... The search for patterns and equivalence classes and their incorporation into the graph are repeated until no new significant patterns are found» (Solan et al., 2005).

In other terms, ADIOS starts with a so-called Motif Extraction (MEX) procedure which looks for bundles of graph's subpaths which obey certain conditions. Once such « patterns » are found, they are subsequently « substituted » for non-terminal symbols and a graph is « rewired » to incorporate such newly constructed non-terminals. Such a « pattern distillation » procedure of generalization bootstraps itself until no further rewiring is possible. Output of the whole process is a rule grammar combining patterns (P) and their equivalence classes (E) into rules, able to generate even phrases which weren't present in the initial corpus. Example of how ADIOS progressively discovers more and more abstract combinatorial patterns is presented on Figure 10.6.1.

ADIOS is undoubtably one of the most performant GI systems which currently exist. It combines both statistic, probabilistic and graph-theory notions with notion of rule-based grammar and as such is also of great theoretical interest. On the other hand, ADIOS does not involve any source of stochasticity, seems to be purely deterministic and as such incapable to deal with highly probable convergence towards locally optimal grammars. In confrontation with some partial corpora this may possibly not cause any problems but, we predict, without any stochastic variation whatsoever, ADIOS could not account for more than few « advanced » & real-life properties of natural languages and as such shall possibly share the destiny of SNPR model.

END NON-EVOLUTIONARY GI 10.6

10.6.2 EXISTING EVOLUTIONARY APPROACHES

Multiple authors have proposed to solve the GI problem with different variants of evolutionary computinng - in following paragraphs we shall describe five different approaches:

1. hill-climbing induction of finite state automata Tomita (1982)
2. GIG method for inference of regular languages Dupont (1994)
3. Evolution of stochastic Context-Free Grammars Keller and Lutz (1997)
4. Evolutionary method of inducing grammars from POS tags of nine different English language corpora Aycinena et al. (2003)
5. Genetic algorithm of Smith & Witten Smith and Witten (1995) for inducing a LISP s-expression grammar from a simple corpus of English sentences

Tomita's 1982 paper can be considered to be one of the first empiric studies of grammatical inference. The study focused on inference of grammars of 14 different regular languages – which are often called « Tomita languages » in subsequent litterature – by means of deteministic finite state automata. Tomita had first encoded any possible finite state machine with n states in a following manner :

$$((A_1, B_1, F_1)(A_2, B_2, F_2)....(A_n, B_n, F_n))$$

whereby every block « (A_i, B_i, F_i) corresponds to the state i, and A_i and B_i indicate the destination states of the 0-arrow and the 1-arrow from the state i, respectively. If A or B is zero, then there is no 0-arrow or 1-arrow from the state i, respectively. F_i indicates whether state i is one of the final states or not. If F_i is equal to 1, the state i is one of the final states. The initial state is always state 1.» (Tomita, 1982)

Thus, for example, the string ((1 2 1) (3 1 1) (4 0 0) (3 4 1)) encodes the finite state automaton illustrated on figure item 10.6.2.

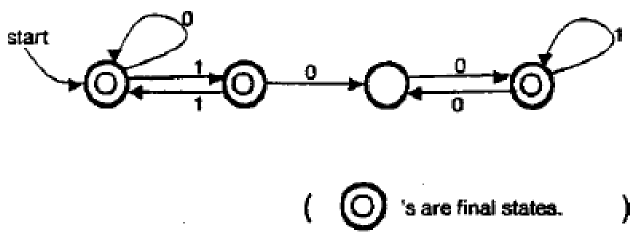


Figure 22: Finite state automaton matching all strings over (1 + 0)* without an odd number of consecutive 0's after an odd number of consecutive 1's. Reproduced from Tomita (1982).

Such encoding allowed Tomita to subsequently apply his hill-climbing approach. Hill-climbing can be considered to be a precursor to more extended genetic programming, since it employs both random mutations to explore surounding search-space and sort of selection algorithm which always prefers to use, in following iteration of the algorithm, such individual solutions for which the value of evaluation function E increases. Tomita's definition of E is very simple:

$$E = r - w$$

« where r is the number of strings in the right-list accepted by the machine, and w is the number of strings in the wrong-list accepted by the machine» (Tomita, 1982). Right-list is a positive sample corpus while wrong-list is the negative sample. Thus, if a random mutation transforms an individual X_n into individual $X_n + 1$ so that $E(X_n + 1) > E(X_n)$, - i.e. if an automaton is discovered which matches more positive sequences, or less negative sequences, or both - it will be $X_n + 1$ which will be mutated in the next cycle of the algorithm.

Tomita's approach cannot be considered to be fully evolutionary because he haven't used populations nor did he employed any kind of cross-over operator. For this reason, Tomita's regular grammar-infering algorithm did sometimes got stuck in local maxima from which there was no way out. Notwithstanding this small imperfection – of which Tomita himself was well aware – his work served, and still serves, the role of an important hallmark on the path to full-fledged GI.

Dupont Dupont (1994), for example, has also focused his study on induction of 15 different regular Tomita languages. In his formally very sound work, he defines the problem of inference of regular languages as a problem of finding of optimal partition of a state space of a finite « maximal canonical automaton » (MCA) able to accept the sentences from positive sample. Fitness function takes into account also the system's tendency to reject the sentences contained in the negative sample. By using a so-called « left-to-right canonical group encoding », Dupont succeeds to represent diverse individuals automata in a very concise way which allows him to subsequently evolve them by means of structural mutation (« the structural mutation consists of a random selection of a state in some block of a given partition followed by the random assignment of this state to a block» (Dupont, 1994), e.g. MUTATE$(((1,3,5),(2),(4)) \rightarrow ((1,5),(2,3),(4)))$ and structural crossover (« the structural crossover consists of the union in both parent partitions of a randomly selected block» (Dupont, 1994), for example $(((1,4),(2,3,5)) \otimes ((1,3),(2),(4),(5)) \rightarrow ((1,3,4),(2,5)),(1,3,4),(2),(5))$.

Because « the search space size dramatically increases with the size of the positive sample, making the correct identification more difficult when we have a larger positive information on the language» (Dupont, 1994), Dupont has also proposed an incremental procedure allowing to start the search process from smaller yet pertinent region of the search space. Procedure unfolds as follows « first sort the positive sample I+ in lexicographical order. Consequently, the shortest strings are first taken into account. Starting with the first sentence of I+, we construct the associated MCA(I+) and we search for the optimal partition of its state set under the control of the whole negative sample I. Let A_1 denote the derived automaton with respect to this optimal partition. Let snext denote the next string in I+. If snext is already accepted by A1, we skip it.» (Dupont, 1994) Otherwise, the aumaton A_1 is be extended so that it can cover also snext. The search under the control of whole negative sample is then restarted and whole process is repeated until all sentences from positive sample have been considered.

With population size of 100 individuals, maximum number of 2000 evaluations, crossover rate 0.2, mutation rate/bit 0.01 and semi incremental procedure implemented, Dupont's approach have attained, in average, classification rate of 94.4%. For five among fifteen Tomita's languages, grammars were constructed which attained 100% accuracy (i.e. accepted all sentences from positive sample and rejected all strings from negatives sample). Results have also indicated that if ever the semi-incremental procedure is applied, the sample size has positive influence upon the accuracy of infered grammars – bigger sample yields more accurate grammars.

While Tomita's results indicate and Dupont's results further confirm the belief that induction of grammars by means of evolutionary computing is a plausible thing to do, they do so only in regards to most similar type of grammars – the regular ones. Grammars of natural languages, however, are definitely not regular languages and models of GI of more expressive « context free » (CFG) or « context sensitive » grammars are needed.

Keller and Lutz Keller and Lutz (1997) employed a genetic algorithm to evolve parameters of stochastic context-free grammars (SCFG) of 6 different languages. SCFGs are similar to traditional CFGs (see item 10.2 for definition of CFGs), but extended with probability distribution, so that there is a probability value in the range $[0, 1]$ associated to every production rule of the grammar. These values are called SCFG's parameters and these are the values which the algorithm of Keller & Lutz aims to optimize by means of GAs. Their approach involves following steps :

1. Construct a covering grammar that generates the corpus as a (proper) subset.
2. Set up a population of individuals encoding parameter settings for the rules of the covering grammar.
3. Repeatedly apply genetic operations (cross-over, mutation) to selected individuals in the population until an optimal set of parameters is found.

Their fitness function F(G) is based on idea of Minimal Description Length (MDL). More formally, Keller & Lutz aimed to maximize:

$$F(G) = \frac{K_c}{L(C|G) + L(G)}$$

by minimizing the denominator which is defined as a sum of number of bits needed to encode the grammar G (L(G)) plus the number of bits needed to encode corpus G, given the grammar G (L(C | G)). Numerator Kc is just a corpus dependent normalization factor assuring that the value of fitness shall be in range $[0, 1]$. When confronted with positive samples of cca 16000 strings (typically of length 6 or 8) of 6 different context-free languages :

1. EQ : language of all strings consisting of equal numbers of as and bs

2. language $a^n b^n (n \geqslant 1)$
3. BRA1 : language of balanced brackets
4. BRA2 : balanced brackets with two sorts of bracketing symbols
5. PAL1 : palindromes over a,b
6. PAL2 : palindromes over a,b,c

their algorithms have converged, in majority of cases, to such combinations of parameters of their SCFGs which had allowed them to accept more than 95% of strings presented in the positive sample. Such results indicate that genetic algorithms can be used as a means for unsupervised inference of parameters of stochastic context-free grammars. Note that Keller & Lutz confronted, during both testing and training, their algorithm only with positive sample. While doing so for training is justifiable - since the objective of their study was to study whether grammars can be infered solely from positive evidence – not doing so during testing phase makes uncertain the extent to which their infered grammars overgeneralize.

Another huge disadvantage in regards to aims of our Thesis is the simple fact that their approach also seems to be very costly (« number of parses that must be considered increases exponentially with the number of non-terminals» (Keller and Lutz, 1997)). And since they confronted their algorithms only with corpora composed of sentences of artificial and not natural languages, we shall not aim to imitate their approach of «tuning SCFG parameters» in our simulations.

By being context-free and not simply regular, the grammars studied in Keller and Lutz (1997) or (Choubey and Kharat 2009) could be considered to be more similar to grammars of natural languages. Nonetheless, languages composed of palindromes and sequences of balanced brackets are still far way off from natural languages and the question « in what extent are results concerning GI of artificial languages applicable to GI of natural languages ? » is far from being answered. Rather than trying to answer it, we proceed now to discussion of two approaches where evolutionary GIs have been applied upon natural language sentences :

The first method, proposed by Aycinena et al. in Aycinena et al. (2003) focuses on induction of CFG grammars from nine different part-of-speech tagged natural language corpora. Sentences contained in these corpora, composed thus of sequences of part-of-speech tags (see Section 10.5)) were used as positive examples, while randomly generated sequences of POS-tags have yielded negative examples.

Initial population was composed of linear encodings of randomly generated context-free grammars, for example the string SABABCBCDCAE would represent this CFG :

$$S \rightarrow AB$$

$$A \rightarrow BC$$

$$B \rightarrow CD$$

$$C \rightarrow AE$$

During the evaluation of individual grammar G, one would first try to parse both positive and negative corpora with the grammar G and subsequently calculate the final fitness by applying the following formula :

$$F = \gamma^{max(0,|\alpha|-|P|)} C(\alpha) - \delta I \alpha$$

« where P is the set of preterminals, $C(\alpha)$ is the number of parsed sentences from the corpus, $I(\alpha)$ is the number of sentences parsed from the randomly generated corpus, δ is the penalty associated with parsing each sentence in the randomly generated corpus, and γ is the discount factor used for discouraging long grammars.» (Aycinena et al., 2003)

In their study, Aycinena and her colleagues had placed randomly generated population of 100 individual grammars on a two-dimensional 10 x 10 torus grid. Subsequently, they had applied a following select-breed-replace strategy :

1. Select and individual randomly from the grid
2. Breed that individual with its most fit neighbor to produce two children
3. Replace the weakest parent by the fittest child

In their framework, « cross-over is accomplished by selecting a random production in each parent. Then a random point in these productions is selected and cross-over is performed, swapping the remainder of the strings after the cross-over points» (Aycinena et al., 2003). Every symbol of a resulting string can be subsequently mutated (mutation rate=0.01). « A mutation is simply the swapping of a non-terminal or pre-terminal with another non-terminal or pre-terminal» (Aycinena et al., 2003)

Figure 10.6.2 shows the number of generations each run was able to complete, the grammar G that last evolved, the percentage of positive examples parsed by G, the percentage of negative examples parsed by G and G's fitness.

While results displayed above may seem encouraging authors, have noticed that in majority of cases, their approach « gives a grammar that is very capable of detecting whether a sentence is valid in English, but it has not learned much English structure» (Aycinena et al., 2003). In other terms, Aycinena et al. have succeeded to breed grammars which have certain discriminatory power but are practically useless as models of English language. They go even so far as to state, in

Corpus	Number of generations completed	grammar	% positive examples parsed	% negative examples parsed	fitness
aliceinwonderland	200000	0 V 0 0 R 0 0 0 J 0 V N 0 R J 0 P N 0 P J 0 0 N 0 0 V 0 V J 0 P 0 0 N N 0 N 0 0 T N 0 J 0 0 T 0 0 T J	92.50%	8.40%	657.364
brown1_a	48500	0 N 4 4 2 N 0 T N 0 J 4 2 0 V 0 P 0 3 R N 0 N 0 4 0 N 0 N V 0 R 2 7 T 0 0 T 4 0 0 T 2 V 0 0 0 P 4 0 V 0 0 V 0 0 R 0 0 J 0 V 4 0 0 N 0 J N	94.10%	6.10%	1667.85
brown1_b	200000	0 J J 0 N J 0 R 0 0 N 0 0 T 0 0 0 N 0 J 0 0 V 0 0 V N 0 P 0 0 T N 0 T J 0 P N 0 R J 0 P J 0 V J	94.70%	6.70%	1227.17
brown1_c	15500	0 4 4 7 R 4 7 T T 1 V J 2 2 P 4 0 J 5 P 7 2 J N 4 0 T 7 1 T 4 V 4 2 O 1 4 0 V 1 0 4 4 P T 0 3 P 7 0 6 0 0 N 0 R 7 7 J N 6 J T 4 P 1 1 T 6 5 7 5 6 N 2 4 J P 7 0 N 0 7 J 4 P 5 0 0 R 4 N N 2 V V 5 2 T 4 P N 3 4 N 6 1 5 0 4 J 0 N V 7 7 J 0 N 4 4 0 5 6 N 2 4 J N 0 R 6 2 2 N 5 5 7 4 R P 0 N T 4 T 0 0 4 P 0 N P 3 V 0 1 1 5 4 R O	80.50%	4.70%	302.996
brown1_d	45000	3 V 0 0 0 J 5 T V 0 0 V 0 R 2 3 0 N 0 N V 0 T N 1 T 3 3 N N 0 0 T 3 T V 0 0 N 0 T 3 3 N 0 0 J 3 0 0 P 0 3 0 2 0 R 2 0 P 0 0 R 3 0 T 0 P 3 0 N P	88.20%	5.60%	583.596
brown1_e	122000	0 P J 0 P 0 0 V 0 0 N 0 0 T N 0 T 0 0 T J 0 V J 0 J J 0 R J 0 J 0 0 R 0 0 P N 0 0 N 0 N J	93.90%	5.90%	1762.67
children	200000	0 P J 0 V 0 0 V N 0 J J 0 N J 0 0 N 0 N 0 0 T J 0 V J 0 J 0 0 R 0 0 R J 0 P 0 0 T N 0 T 0	91.80%	5.70%	677.211
tomsawyer	200000	0 P J 0 V N 0 N N 0 0 V 0 T 0 0 T N 0 T J 0 0 N 0 0 J 0 N J 0 P 0 0 P N 0 V J 0 V 0 0 R 0 0 J 0 0 N 0 0 R J	92.70%	8.60%	2292.89
wizardofoz	200000	O 0 0 O P J O V O O R O O P O O R J O V N O T N 0 O O O O 0 J O T O O P N O N V O V J O T J O N O O N N O J N	89.50%	9.20%	920.852

Figure 23: Grammars induced from nine different POS-tagged corpora. Reproduced from Aycinena et al. (2003).

the ultimate paragraph of their work that « It is still possible that English grammar is too complex to be learned from a corpus of words» (Aycinena et al., 2003) and that other external clues are necessary for successful GI of English.

The big disadvantage of above-mentioned algorithm was also the fact that its input were sequences of already attributed POS-tags and not sequences of words themselves. Thus, even if the approach would discover some interesting grammars, a reproach could be made and justified that in fact it only re-discovered the rules of the tagging system which was used in the first place. From perspective of our Thesis, another disadvantage of Aycinena et al.'s approach is related to the fact that their approach is anything but model of grammar development in human child. For it is evident 9 that children learn the grammar of their language in an incremental fashion – they are not confronted with whole corpus from the very beginning. Nor does the corpus stay identic after each iteration of the learning process. On the contrary : as child grows, its linguistic environment - the corpus – also grows. Both in length and complexity.

An interesting evolutionary approach of GI which both tries to create own non-terminal categories and also takes such « incrementality » into account is presented in the work of Smith & Witten Smith and Witten (1995). In their scenario, candidate grammars are evolved after presentation of every new sentence. Grammars have form of LISP s-expressions whereby AND represets a concatenation of two symbols (i.e. a syntagmatic node) and OR represents a disjunction (i.e. a paradigmatic node). Whole process is started as follows : « The GA proceeds from the creation of a random population of diverse grammars based on the first sample string. The vocabulary of the expression is added to an initially empty lexicon of terminal symbols, and these are combined with randomly chosen operators in a construction of a candidate grammar...If the candidate grammar can parse the first string, it is parsed into the initial population.» (Smith and Witten, 1995)

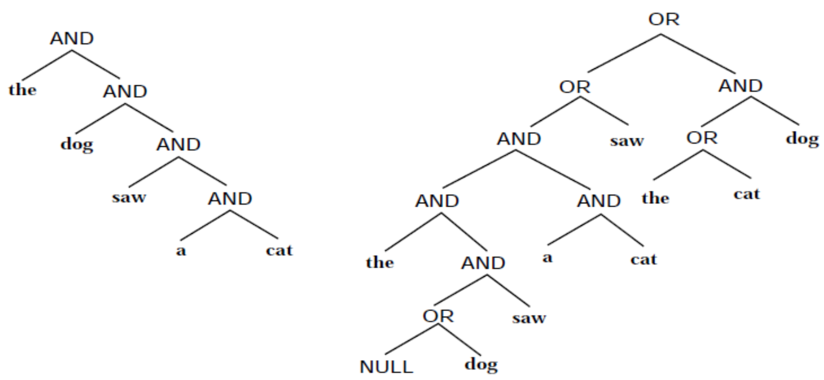


Figure 24: Two simple grammars covering the sentence "the dog saw a cat". Fig. reproduced from Smith and Witten (1995).

Figure 10.6.2 displays two sample grammars for the sentence « the dog saw a cat ».

S-expression sequences representing individual grammars are subsequently mutated. Couple of parent grammars can also switch their nodes – probability of being chosen for such cross-over is inversely proportional to grammar's size : shorter grammars are prefered. Cross-over is non-destructive, parents thus also persist. The events of reproductions are grouped in cycles, at the end of each cycle, population of candidate grammars is confronted with new sentence from sample of positive evidence.

In their article, Smith & Witten demonstrate, how after presentation of sentences : «the dog saw a cat », « a dog saw a cat », « the dog bit a cat », « the cat saw a cat », « the dog saw a mouse » and « a cat chased the mouse » their system naturally converged to a grammar which had quite correctly subsumed determiners like « a », « the » under one group of OR nodes, verbs like « chased », « saw », « bit » under another, and nouns like « dog », « cat », « mouse » under yet another. The grammar which they finally obtain is not ideal but, as they argue, it could get better if confronted with new sentences. « It is an adaptive process whereby the model is graudally conditioned by the training set. Recurring patterns help to reinforce partial inferences, but intermediate states of the model may include incorrect generalizations that can only be eradicated by continued evolution. This is not unlike the developing grammar of a child which includes mistakes and overgeneralisations that are slowly eliminated as their weaknenesses are made apparent by increasing positive evidence.» (Smith and Witten, 1995)

While strongly agreeing with above citation, we nonetheless cannot ignore certain drawbacks of Smith & Witten's approach. Most importantly, by using LISP's s-expressions as a way of representing their grammars, they ultimately have to end up with highly bifurcated binary trees (since arity of AND|OR operators is 2). Thus, one can easily subordinate two non-terminals to one terminal (e.g. OR(cat,dog)), but in case of three subordinated terminals, one is obliged to use complex expression involving three non-terminals (e.g. OR(OR(cat,dog),OR(mouse,NULL)). Therefore, in such an s-expression based representation, is any class having more than two members neccessarily represented by a longer sequence → is more prone to mutation → is highly « handicapped » in regards to much shorter expressions subordinating just two nodes.

Another drawback of Smith & Witten's work which cannot be ignored is related to the fact that while they used English language sentences to train their system, the sentences were very simple and the relevance of their findings to GI of « natural » English is more than disputable. In fact, they seem to achieve, with quite complex evolutionary machinery, even less than Wolff's deterministic SNPR model have achieved almost a decade before. Notwithstanding these two drawbacks we nonetheless consider as particularly inspiring their approach aiming to solve the problem of GI of natural languages by uniting, in one framework, the notions adaptability, evolvability and statistical sensitivity to recurring patterns.

We summarize : all five above-mentioned approaches indicate that evolutionary computing can potentially yield useful solutions to the problem of Grammar Induction of both artificial (regular, context-free) and natural language grammars. The length of the candidate grammar is frequently used as an input argument of the fitness function. Note also that both solutions of Dupont and Smith & Witten also use a sort of « incremental » procedure whereby individual solutions gradually adapt to every new sentence. Especially Dupont's findings are reminiscent of what was already told about « importance of starting small » when discussing computational model of Elman (Section 10.6.1).

On the other hand, none of the above mentioned models was confronted with corpus of child-directed (i.e. « motherese ») or child-originated utterances. The objective of our Thesis shall be to fill this gap.

END EVOLUTIONARY MODELS OF GI 10.6.2

Asides these non-evolutionary and evolutionary algorithms for grammar induction, there exist also first tentatives to solve the GI problem by means of Grammar Systems (10.2.3). The pioneer work in this regard is the study of (Sosík and Štỳbnar, 1997). Contrary to majority of GS-inspired authors who focus on productive (i.e. generative) aspects of GS, Sosik & Štýbnar have focused on GS's language-accepting properties. In a hybrid connectionist-symbolic architecture, they have used a «neural pushdown automaton» to infer a language colony (10.2.3) able to cover some simple artificial context-free grammars able to cover balanced parenthesis or palindrom languages. While their results demonstrate that it is indeed viable to perform grammatical inference by means of grammar systems, the artificial nature of the input languages makes it difficult to see whether their approach could be of any use in modeling acquisition of natural language.

This being said, we conclude with statement that as of 2015, ADIOS (Solan et al., 2005) seems to be the only full-fledged computational model of unsupervised grammar induction which is

- publicly available (at least partially[23])
- capable of inducing grammars even from child-speech transcript input data (Brodsky et al., 2007)

For this reason we shall compare, in second volume of this Thesis, results of our ELSA-based simulations with those, induced by ADIOS.

END GRAMMAR INDUCTION 10.6.2

As of 2015, NLP is one of the most "hottest" and active sub-disciplines of not only computational linguistics, but also computer and potentially cognitive sciences in general. Without being aware of it, lifes of billions of people are influenced on a daily basis by client platforms, applications, marketing bots or search engines which implement some kind of NLP technique.

In NLP, accuracy - defined, for example, in terms of precision and recall (10.3.1) - is always important because it is easier for human users to interact with more accurate systems. But in real-life applications, accuracy is not the only constraint which has to be taken into account: speed and computational complexity of the task are also crucial.

23 Demo version of ADIOS can be downloaded from http://adios.tau.ac.il/download.html

To support our point, let's take a Turing test as an example: question-answering systems which need hours to generate the most accurate and valid answer shall not pass the test; the test shall be passed by machines which offer an approximate answer in few seconds. Hence, even in case of the challenge from which whole discipline of NLP originates, accuracy of one's model is not a goal *per se* and is, in fact, useless if one forgets that expression "natural language" does not mean a piece of dead, static corpus stored on one's disk, but rather a set of sequences of symbols always expressed in a context, and alway expressed with an intention.

END NATURAL LANGUAGE PROCESSING 10.6.2

Computational linguistics is a symbiont of computer science and linguistics. In this chapter, we have explored its three principal components:

1. Quantitative and Corpus Linguistics (QCL) devoted to discovery of patterns and laws within linguistic corpora

2. Formal Language Theory (FLT) devoted to formalization of principles of syntax in terms of set theory and algebra

3. Natural Languauge Processing (NLP) devoted to amelioration of machine's faculty of processing of information which machines exchange with human beings

During introduction to QCL Zip's law "the frequency of the word is inversely proportional to its rank" and logistic law "the increase is first slow, than fast, shan slow again" were discussed in somewhat closer detail. It was noted that both of these laws are relevant descriptive mechanisms for diverse diachronic processes, both in linguistic ethnogeny as well as linguistic ontogeny. The fact that both of these laws yield very successful models for description of *ecological* phenomena was also brought to attention.

Brief overview dedicated FLT to has offered only the very basic definitions: language L was defined as a potentially infinite set of strings of symbols chosen from a finite alphabet; grammar G_L was defined as a formal system containing rules of production able to generate, as its theorems, exactly all and only strings of L. Classes of regular, context-free, context-sensitive and unrestricted grammars were described and usefulness of such hierarchical view of things was mentioned, notably in relation to artificially (e.g. programming) languages. Brief excursion to multi-agent, non-monolithic, parallelized and modular "Grammar Systems" have illustrated that "miraculous" things - like ability to generate infinite language obtained by interlock of two finite grammars - can happen whenever individual component grammars share their input/output string environments.

The major part of the chapter was dedicated to NLP. Methodological aspects which NLP shares with machine learning field of artificial intelligence were first pointed out. Subsequently, three classes of problems were addressed:

1. problem of geometrization of meaning was principially presented as projection of semantic features into N-dimensional metric spaces
2. problem of part-of-speech induction was principially presented as projection of morphosyntactic features into N-dimensional spaces + subsequent attribution of specific partitions of morphosyntactic metric spaces with specific non-terminal labels
3. problem of grammatical induction was principially presented as a problem of part-of-speech induction + gradual optimalization of content and order of substitution rules

Few exemplar solutions to these problems were mentioned, both deterministic and, if existing, also non-deterministic and evolutionary. It was noted that some encouraging results were already attained but that there is still plenty of work to be done.

So let's do it.

END COMPUTATIONAL LINGUISTICS 10.6.2

11 SUMMA II

Different paradigms have been presented in preceding chapters:

1. universal darwinism
2. developmental psycholinguistics
3. computational linguistics

1^{st} offering a theoretical framework; 2^{nd} offering the data, the materia, the object of interest; 3^{rd} offering the method how the validity of the theory in relation to materia is to be ultimately demonstrated.

The framework: a theory of intramental evolution. Id est, a theory stipulating that not only genes or memes evolve, but that there exists yet another, 3rd kind of evolutionary force which moulds man's destiny. An evolutionary force which is neither phylogenetic like unceasing development of DNA-molecule, nor ethnogenetic and cultural like the memetic evolution occurent between mutually communicating minds. An evolutionary force which is profoundly ontogenetic: a sort of process limited by a life span of the individual in whose mind the process occurs.

The materia, the object of interest: a mind of a child. Id est, a mind in constant change, an exploring mind, a playful mind. A mind that masters, in less than three years of existence and practically completely *ex nihilo*, the most fundamental structures of her mother language. Indeed in less than three years do the representations encoding the universally perturbing cry of a newborn into our world evolve into evermore precise, robust and well-adapted prosodic, phonologic, phonetic, morphosyntactic, semantic and pragmatic representations. Being unafraid of commiting an error and feeling no shame nor guilt when doing so, a soul of an infant, previously so alien to our world, gradually and swiftly learns how to live in it. Gets grounded in it, gets informed how to live in it with us.

The method: a computational simulation. Id est, a simulation aiming to reproduce, *in silico*, at least few key processes through which a child learns its mother language. A simulation that would succeed to partition the world of its representations into categories or clusters similar to those which an organic child would construct, if ever presented with the same data. A simulation able to discover and provide grammars whose products would be undistinguishable to utterances produced by normal human children in course of their daily interactions. If such goal were to be attained by means of evolutionary computation, success of such simulation could be used as a non-invasive, indirect proof that a sort of ontogenetic, intramental evolutionary process governs the process of language acquisition in human children.

A theory of intramental evolution, a mind of a child and a computational simulation: among this trinity of cornerpoints embedded in a semantic space representing our current knowledge, one can observe overlapping regions, one can observe common topics. To start with, note the notion of a gradual yet continuous change: no matter whether in subdisciplines of UD, DP or CL, outputs of phase T_N serve as inputs for the next phase T_{N+1}. There exist an analogy between successive stages of a developing child and successive iterations of an NLP algorithm: both invest present energy into processing of knowledge attained in the past so that more accurate performance can be attained in the future. Cognitive representations continuously change but the processes which make the change possible are always present.

Note that such gradually changing continuity does not exclude that from time to time, paradigm shifting, phase-transiting phenomena shall be observed. On the contrary, such moments of global equilibriation of the whole psycholinguistic system are necessarily implied by any theory that consideres child's linguistic faculty in moment T to be a nexus of parallel activity of many modular entities whose means of interaction are complex and potentially non-deterministic.

The notion of "parallel activity" is thus equally crucial both for the theory, as well as for correct understanding of observations and simulations that shall follow. That human brain is a device which processes information is a well-known fact; the sequential nature of language can, however, lead one to a conclusion that language is processed in a monolithic, serial fashion. To a somewhat "monotheistic" conclusion that to every language utterance (in production) or to its understanding (in comprehension) there leads only one correct sequence of applications of rules extracted from one correct grammar.

We consider such conclusions as fallacious. Knowing how the nature usually tends to proceed, we do not consider as necessary to postulate cold, fixed, static, formal, universal and omnipresent order there, where much more local notions of dynamism, variation, interaction, exchange and convergence clearly suffice. Given that the notion of "convergence" is flexible enough to account for the fact that, in course of time, completely different species (e.g. humans and cephalopodes) "obtained" the organ with identic function (e.g. eye) by following two completely different evolutionary trajectories, we believe that it should also be flexible enough to explain the "mystery" of language acquistion:

Children learn language by converging to it.

And as we shall now proceed to demonstrate, it is through interaction with peers and parents that the point of convergence is to be discovered.

END SYNTHESIS OF PART II 11

Part III

OBSERVATIONS

A child's spontaneous remark is more valuable than all questioning in the world. Jean Piaget

This part shall describe certain observations related to ontogeny of linguistic structures and interpret in terms of theory of intramental evolution.

Its first chapter is principially a longitudinal qualitative study of one particular human child. At its beginning, a non-invasive, phenomenological, observational data-collecting method shall be described and few salient moment of subject's prenatal and postnatal development shall be mentioned. The major part of the study shall be devoted to subject's linguistic development during the toddler period, id est between 10 and 30 months of age. Among others, some among subject's first words, first phrases, first pivot grammars and first variation sets shall be presented.

A set of "evolutionary" notions shall be developped and defined in order to facilitate the interpretation of the obtained data in evolutionary terms. Notions like intralexical | intraphrastic | interlinguistic crossover shall be thus introduced and multiple real-life cases shall be furnished for each notions

These notions will play important role in the following chapter devoted to quantitative observations. When possible, they would be transcribed into a form of PERL-compatible regular expression. The corpus of child-language transcriptions (CHILDES) shall be subsequently processed by such regexps in a series of simple and reproducible data-mining, pattern-extracting experiments. Ideally, patterns and statistical regularities shall be discovered which are not only language-specific but also language-indepedent. That is, occurent in not only English but ideally in all languages attested in CHILDES corpus.

12 QUALITATIVE

12.1 METHOD AND DATA COLLECTION

Limits of traditional method

In no domain of scientific endeavour are limits of Gallileo-Cartesian dubitating yet experimental method as visible and problematic as in studies of subtle mental and psychic layers of human subjects. And in case of studies of human children, these problematic situation is marked to the very extreme: due to a sort of psychosocial *uncertainity principle*, the very act of observation significantly modifies the properties of the observed subject. Trying to fox a healthy, curious, vivid human child in an artificial experimental setting is plainly and simply contradictory to any tentative of evaluation of child's *natural* behaviour.

Significance levels are arbitrary

Neither is reassuring a traditional quantitative "psychological" paradigm in which one proves one's hypothesis through statistic comparison of a study-group with a control-group. Even if all went well and one would succeed to solve the unsolvable and limit the influence of external and hidden variables to a very minimum and even if all children would behave as expected during the experiment (a very improbable "if" indeed), and even if all subsequent statistical evaluation was sound and solid, one would end up with one null hypothesis, few coefficients and a p-value. "So You state that those kids cross-over such linguistic structures and those other don't. And that the difference is significant because the p-value is 0.045. But, You know, our community has decided not to bow in front of the Fisher-defined $p<0.05$ significance level threshold (Fisher, 1925)" could be a provocative, yet valable denial of such a result.

Problem of experimental invasivity

Asides and above all such criticizms thrones the ethical problem of invasivity of one's experiments. One cannot have a theory that postulates that any stimulus - *no matter how small and ephemere* - can influence child's lifelong trajectory and still aim to prove such theory by means of putting a child with artificial, non-human, mentally perturbing experimental conditions. Of course, in mental world of experimentators who depart from an axiom that children neither feel nor reason, such methodology is still allowed. Others can also somehow bridge the cognitive dissonance which necessarily follows. But for a scientist who departs from a belief that child feel and reason much more than adults shall ever will - and such was, indeed, our bias of departure - an experimental invasivity is an important κριτήριον which significantly constraints one's ways of doing responsible and sustainable science (Hromada, 2010b).

Given that our objectives were not (medi | clini)cal but rather those of *recherche fondamentale*, we have not found any reason which could potentially justify use of any kind of invasivity. All these considerations + practically zero funding taken into account a traditional quantitative methodology of experimental psycholinguistics has been discarded as inappropriate cca in 2nd year of our doctoral studies.

Such a methodological design choice was further motivated by information announcing the "good news" that a child is to be born, in whose closest presence we could spent years to come. This have put us into positions of savants like Piaget, Braine, Labows or Tomasello who had all honor and luck to confront their theories with years-lasting, longitudinal observations of their own children. Thus, our rejection of purely cartesian attitude seems not to have disastrous consequences neither for validity nor for reproducibility of observations which have followed.

parvuli deûm regnum

And what have followed is this: from the moment of subject's birth (0;0;0) author of this dissertation has kept a journal. Journal was first written as an objective "observation log" but quite soon (0;7) it obtained a form of personal monologue addressed, in 2nd person singular, to adult person which shall, ideally, become from the child herself. Entries in the journal have been written down according to a sort of biased, random sampling procedure: that is, whenever subject generated an event which was sufficiently salient and whenever all among other conditions were fulfilled (i.e. father observed the event or mother told about it to the father; journal was in the proximity; pen or pencil was in the proximity; observer had enough time to note the observation down, etc.), then and only then was the entry written down.

Given such a relaxed methodology, 123 hand-written journal pages have been filled with 167 records (14 recorded by mother; 153 by father) before the subject have attained the upper bound of the toddler period (2;6;0).[1] What shall follow in this chapter are principially biased descriptions and biased intepretations of suched biased observations.

12.1.1 BIASES

In retrospective analysis of the observation journal, which started at (2;6;0) and ended at (2;6;14) observer is struck by omnipresence of following biases:

1. observer consider the subject to be endowed with consciousness

1 Asides the hand-written journal, cca 20 gigabytes of audiovisual material were collected, often in situations when the subject played, ate, was in REM-phase, danced or simply toddled and babbled. With two or three minor exceptions, this data shall not published in the present work.

2. observer considers the subject to be somebody, who shall evolve into a conscious adult

3. observer is the parent of the child

4. observer and vaste majority of other personnae mentioned in the journal seem to be strongly attached to the child with a bound which is difficult to describe without referring to the meaning of the word "love" (14.4.2)

5. observer focused on noting down the observations which match his theory and was, in fact, unable to note down observations which do not match the theory

Disadvantage of such biases is that they distort the objective state-of-affairs. But this disadvantage can be reduced if such biases are known. And in case of biases 1-4, the disadvantage can even turn out to be advantageous: for these biases are well-known to vaste majority of those, who were ever blessed with having a child. Thus, instead of making our observations more subjective they help us to establish a common prism through which our communicative intention could be potentially understood.

The 5th bias, of course, is problematic. From our current perspective there is little which can be done to combat such a sort of cognitive blindness which have made us ignore practically all data which does not fit our theory. Thus, instead of solving the situation by pretending that we have observed all that was to be observed, we prefer to honestly admit that in regards to all that could have been noted down, but wasn't, observers has often acted as strongly biased, cognitively blind, hormonally reprogrammed *fachidioten*.

END BIASES 12.1.1

It is known since time immemorial that a conscious, reflected, *sattvic* awareness of one's biases is a condition sine qua non of a viable and valable methodology. But it was Husserl and his followers who gave the method its western name by calling it **the phenomenological method**.

It is, indeed, a sort of phenomenological methodology which can be understood as *the method* behind words to come.

END METHOD 12.1

12.2 SUBJECT

The subject was conceived as a result of emotionally charged yet fully conscious decision of two adult individuals. In the prenatal period, mother have included consumption of magnesium-rich mineral waters and iron-containing supplements into her otherwise healthy, dairy&vegetable&fish-based diet. Pregnancy progressed without any major complications and in 6-th month (0;-3), father could have felt, during a week-lasting music festival, that the child was already able to atune its kicking with musical beats.

Birth occured approximately three weeks before expected term and was probably caused by mother's passing in the proximity of an active asphalt-drilling machine. Birth itself lasted exhausting 28 hours: the mother asked for an epidural injection after 23 hours of tentatives to mentally influence the extent of cervical dilation. From now on, the initial letters I.M. of her two names shall be used to refer to healthy girl thus born.

Given that the first-born came to world approximately two months before winter solstice, its first tentatives to move corresponded with increased luminosity of longer days.

Standard unfoldment of the *universal* sensori-motor algorithm followed: rotation from back to belly at (0;5), first unsuceessful crawling tentatives at (0;5;25), sitting on chair at (0;7;14), crawling on four at (0;8), autonomous standing at (0;10) and first step at (0;11;20). Lateralization expressed by right-handed object manipulation preference was noted down at (0;6;25).

Eruption of first teeth noted at (0;9;17). In spite of this fact had the breast-feeding continued until the material bond between mother and daughter was broken - after multiple unsuccessful tentatives - at (1;10) by a more or less bilateral agreement of both participants involved.

In the toddler period, neither IM nor the members of have closest social surroundings suffered any serious ilness or traumatic experience. IM can thus to be considered as what is often known in developmental literature as "normal child".

END SUBJECT 12.2

12.3 LINGUISTIC ENVIRONMENT

IM's linguistic competence developed in multilingual environment. Both parents are of slovak (western slavic) origin. However, since mother spend more than half of her life in Germany, and since the child was born and raised in Germany, IM-directed "motherese" was at least 60% german-based.

Father migrated to Germany just few months before the child was born and was thus struggling with problem of secondary language acquisition practically in the same period as the child was struggling with first language acquisition. Between themselves, parents spoke mostly slovak. Father's IM-oriented language was also mostly slovak.

But in majority of other regular daily interactions, IM was mostly exposed to german. In non-negligeable amount of cases, IM could observe one or both of her parents verbally interact in czech, english, french, spanish and, in much lesser extent, polish, ukrorussian, sanskrit and tibetan (sorted in descending order according to structural exposure frequency).

IM started going to creche two days after her first birthday (1;0;2). There, she was mainly surrounded by peers verbally interacting by means of german-ressembling idioglossias.

END LINGUISTIC ENVIRONMENT 12.3

12.4 CRYING AND BABBLING

After few months of more&more differentiated crying forms, "happy cooing" was, along with smiling, noted down at (0;2;18). Three months later, as soon as of (0;5;11) mother had noted down the presence of canonical babbling sequences *bäh, bäh, bäh; dwn dwn dwn; mama-mama*. In the same record the mother conjectures that the sequence *hop hop hop* corresponds to knee-bending and *tou tou tou* corresponds to stretching of hands.

Being more sceptical about IM's ability to verbally communicate, paternal record from (0;5;25) observed in child's production the presence of vocalizations with occlusive labial, velar, glottal and laryngal features. Glide-like *dwndwn* like and trill-like *drndrn* were also noted during the period. Paternal scepsis notwithstanding, a synchronicity between the overall context and child's communicative intention had made the father to note down, already at (0;7;14), the hypothesis that *bwí* could potentially mean porridge [Breie].

First sequences composed of different syllables were observed at (0;7;23). At (0;10;13), babbling sequence of a sort *tititatatetededededidi* was recorded and a week later, syllables *ma;pa;ba;ta;da;te;ti;ne;de; me; pe;be;we;bwe* were enumerated as most salient.

As late as (1;8;7) such canonical babbling was listed as one among multiple modes of communication:

1. crying of a hungry newborn
2. squeling disapproval of a pampered child
3. "mentor mode" (observable especially when IM communicated with smaller children, acommpanied with vivid gests)
4. melody singing (especially when in stroller or in bike sit)
5. canonical babbling

In spite of the fact that both bursts of cry as well as production of expressions **highly repetitive yet gently variating** syllabic streams was observed as far as the end of toddler period[2], we conclude that the babbling schemas had lost their dominant position not later than at (1;6). For at this period it became evident that at least certain forms of IM's language had lost their private, idioglottic character. Convergence of IM's neurolinguistic structures towards an optimal communicative system was on its way.

END BABBLING 12.4

12.5 FIRST WORDS

As was indicated in the previous section, mother had detected the sequence "mama mama" as early as of (0;2;18). Father had noted down the marked repetition of sequence *məmə* at (0;7;19) and few days later, at (0;8) had noted down that *məmədenotes disagreement*. However, it was only at when (0;9;17) father had noted down that "*it is possible that the term məmə- which becomes more and more phonetically similar to MAMA*[3] *- already denotes the mother not only as a source of food, but also as a person whom You love and whose presence makes You happy*". Given that IM had often used, in following months, the word "mama" in contexts as diverse as

1. request for food
2. call for help
3. declaration of joy
4. looking at father's photo (c.f. below)
5. approaching "home"[4]

it seems to be the case that even the meaning of such a fundamental signifiant is not completely fixed and varies in time. But given that IM's mother have practically always interpreted such a term as a signal which made her personally and immediately concerned, the term got potentially quite fixed and served as a sort of label denoting IM's

2 C.f., for example, a sequence recorded between 30th and 70th second of the video downloadable at http://wizzion.com/im/latebabbling.avi. Recorded at (2;5;8).

3 We shall use upper case letters to mark such signifiers which most probably already encoded a specific meaning. Lower case transcriptions shall represent sequences whose meaning, at the moment of production, seemed to be absent or highly ambigous.

4 This was noticed at (1;7;12) when the father used "Google streetview" application to perform a small experiment. IM could see, on the monitor, the streets she already know from the real life. Once the walk ended in front of entrance to house where IM lives, IM pointed to monitor and cried "MAMA!".

mother. Later, already in two-word phase and after she has "discovered" that every peer in the creche has his own distinct "mama", IM started to denote her one and only mother with the term "MAJNE MAMA".

When it comes to paternal term which is "tato" in slovak, father had noted down the production of sequence "tata" as soon as (0;7;23), mother at (0;8;21). The first indication that IM's brain associates the term with the father was furnished by the mother who, during the trip to seacoast where father was absent (0;9;9) saw IM looking at father's photo, uttering "TATO" and than observing the sea for a long time, silent. Three months later, at (1;0;23), such a romantic view was somewhat perturbed by father's observation that IM used the term "mama" when looking at photo on which only father was depicted. Thus, it was only after months-lasting experimentation with pronounciation with dental occlusives in sequences like "ata" (0;7, 1;0;1), "ada; dada" (0;8;21) "toto; tete; tata" that the father noted, at (1;3;16) that *the most popular word is currently TATO and it is quite possible that it also means what it is supposed to mean, since You often say it either when I disappear from Your view, or when You want something from me*".

A first non-parental term whose repetitive usage was considered as worth recording both by mother (0;8;21) and soon afterwards by father (0;8;29), was ENTE. Since at such an early age, IM had used this term -which meaning "duck" in german- in an exclusive and strongly repetitive fashion when she was confronted with books with ducks, bathtub ducks as well as real organic instances of species *Anas platyrhynchos* it can be stated that IM succeeded to create a cognitive representation of the word ENTE whose extension strongly overlaps with the one held by IM's social surroundings. This can potentially be explained as a consequence of "duck-feeding and duck observation" rituals in which IM participated on a regular weekly basis since third week of her life. But given that ducks are often mentioned in "lists of first words" (c.f. Table 3) or "first word combinations" (c.f. for example (Braine and Bowerman, 1976, pp23,32,44,49)) presented by other western authors, one is tempted to state that IM's obsession with the form ENTE is to be explained not only as a sort of caprice of ontogeny of individual psyche, but can also have cultural or even phylogenetic roots.

At (1;0;23) father noted down that IM often uses the term BABA when speaking to and/or demanding the presence of her grandmother. This is consistent with the fact that the term is a slovak coloquial denoting grand-mother, or old woman in general[5]. Later, the term was often used as a part of fixed construction "HALO BABA" (1;2;21,1;4;10)

5 Note that in many languages, the term "baba" is often associated to meanings which C.G.Jung would most probably understand as instances of the archetype of "old and wise authority". Thus, asides its well-known use as a honorific in sanskrit, persian, turkish or arab, the term baba denotes old and wise man among Shona people of Zimbabwe or Yoruba people of Nigeria, and potentially in other ethnics as well.

People: MAMA (0;9), TATO (M0;9), BABA (1;3)
Food: BAJA/ANAN [banana] (F1;5), MI [milk] (F1;5), BROT [bread](F1;4)
Body parts: NENE [F1;0;23], HÁE [hair](1;5)
Places: KITA [creche](F1;4), ŠPIPA[playground](F1;5),
Animals: ENTE [duck] (MF0;8), uau-uau [dog] (F1;5), mjau [cat] (F1;5)
Toys: BAJ [ball] (1;6), TEDY [teddy-bear] (1;6)
Household objects: KE [keys] (1;5)
Routines: halo (F1;4), e-e [refusal] (M1;4), najn [no] (M1;5)
Activities: papa [to eat](1;2), hají [sleep](1;5), daj [give!](F1;5), auke [sway!](F1;5)

Table 10: IM's productive lexicon before attainment of 18 months. Words in the brackets denote most plausible meaning, as decoded by either father (F) or mother (M). Compare with Table 3.

potentially imitated by rote from mother's telephone talks to her mother.

Table 10 contains the list of words noted down before IM attained one and half year of age. The list is fairly standard and resembles other such lists reported in litterature. Food and game-related imperatives were common, as well as animal-like onomatopeias. In majority of cases, an initially idioglotic, private sound-form of produced word developped in a sense which would ideally match the "ideal" sound-form of the parents. IM's C- and P- structures adapted to her surroundings.

There occured, however, multiple cases where C- structures of parents adapted to private P-structures of the child. Most salient among these was the case of a word NENE, noted down quite early (F[6] 1;0;23), referring to mother's "breast". Mother swiftly included the paedologism into her own productive lexicon, as documents her (M1;5;24) journal entry where she used the term as a component of a wider declinated expression "meine nene".

12.5.1 nene & taboo (aph)

Humans are essentially mammals. In a healthy normal situation, first communicative channel between the child and the world passes through *mamelles de mama*. And indeed many are indices that bond created by and during breast-feeding can significantly influence ontogeny of child's cognitive and linguistic structures (Hromada, 2009).

It is thus somehow surprising to see that the topic of breast and breast-feeding is either ignored or tacitly cast aside by major figures of contemporary DP. Indeed, **one shall not find a single occurence of**

6 From now on, all references to the observation log shall be preceded by the consonant specifying the author of the enter, e.g. (f)ather or (m)other.

the word "breast" in (Tomasello, 2009) or (Karmiloff and Karmiloff-Smith, 2009). Also in Pinker's Language Instinct which pretends to introduce *The New Science of Language and Mind*, the breast is mentioned only once in context quite unrelated to ontogeny (« Proto-Indo-European melg "to milk" resembles Proto-Uralic malge "breast" and Arabic mlg "to suckle» (Pinker, 1994)). Thus the only monography, which somehow saves the score and mentions breast in developmental context, is (Clark, 2003) where, in table 4.2 on page 83 is the term "nenin", produced by a french child translated as breast.[7]

END NENE & TABOO 12.5.1

The fact that the term breast seems to be taboo for contemporary psycholinguists is even more striking when one realizes that it was already one of the father of the discipline, Roman Jakobson, who pointed out that « often the sucking activities of a child are accompanied by a slight nasal murmur, the only phonation which can be produced when the lips are pressed to mother's breast or to feeding bottle and the mouth is full. Later, this phonatory reaction to nursing is reproduced as an anticipatory signal at the mere sight of food and finally as a manifestation of a desire to eat, or more generally, as an expression of discontent and impatient longing for missing food or absent nurser, and any ungranted wish. Since the mother is la grande dispensatrice, most of the infant's longings are addressed to her, and children gradually turn the nasal interjection into a parental term, and adapt its expressive make-up to their regular phonemic pattern.» (Jakobson, 1960)

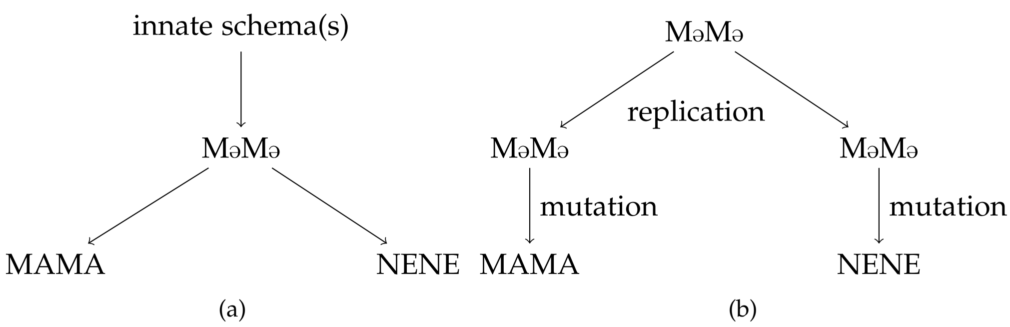
Asserting that our observations of IM's interactions confirmed Jakobson's insight, we propose following developmental analysis of IM's πρώτα ονόματα:

Left part of Figure 25 suggests that the development of structures MAMA and NENE can be understood in terms of a general process during which the succion reflex extends into vocalized labial P-structure (MəMə). Subsequently, this "centroid" schema differentiates into two schemas MAMA and NENE. Right part of the figure conjectures that such a differentiation can be explained in terms of replication, variation and selection:

1. first the initial structure (MəMə) gets reproduced

7 Note that none of IM's parents was aware that the term NENE means "breast" in french argot. This was "discovered" only post hoc, after the term NENE was already unambigously used and understood by all family members. Given that the same signifiant was found out to denote the same referent in two independent language systems (i.e. IM's idioglossia and french argot) the theory of "arbitrairness of sign" (de Saussure, 1916) is to be partially revisited.

Figure 25: First differentiation between the whole and its part (a) and its evolutionary explanation (b).



2. some of resulting replicas are subject to mutation (shift towards open vowels in case of emergence of MAMA, shift towards alveolar nasals in case of NENE)

3. structures which turn out to be useful (e.g. they increase probability of being breast-fed) get reinforced, fixed and succeed to survive to time (contrary to "less fit" structures not resulting in fulfilment of child's communicative intention)

This being said, we shall now focus on other phenomena of IM's linguistic development which seem to fit into such evolutionary framework.

END FIRST WORDS 12.5

12.6 REPETITIONS AND REPLICATIONS

Repetitio est mater studiorum et repetitio replicatio est. Repetition is a form of replication (3). It may be argued, of course, that this formula is not always valid: take as an example an agent without any memory whatsoever which just executes random movements and by sheer caprice of hasard repeats the same movement as it has already executed sometimes in the past. But in case of agents with mnemonic substrate powerful enough to project the temporal onto spatial (e.g. human brain) we see no a priori reason why the formula should be rejected. Hence, repetition of information is a form of replication of information.

By repeating information, children brains replicate information. We distinguish two major types of processes behind replications:

1. intersubjective replications

2. intrasubjective replications

As everything in human mind, these processes mutually interact. But in early development, so we argue, they can be discriminated as independent.

In an intersubjective replication, a structure S is articulated, performed and|or expressed by two or more distinct subjects. Thus, when mother's saying of the word "TATO" is followed by child's utterance of the same word, one observes a minute intersubjective replication. Thus, intersubjective repetition can be understood as equivalent to imitation.

One observes intrasubjective replication whenever a structure S is articulated, performed and|or expressed by one subject in two distinct moments. A replication of a syllable MA in the word MAMA can be thus understood as one amongst its most simple cases. Canonical babbling or many among Piagetian "circular reactions" can be also understood as expressions of such a general cognitive process.

As is always true in case of a healthy human mind, the intrasubjective and the intersubjective mutually interact. But in early development, so we argue, the two can be discriminated as independent.

In IM's case, it was around her first birthday when the interplay between these major processes started to express itself in observable forms of verbal interaction. More concretely, at (f1;0;7), IM produced a bi-syllabic MAMA after hearing bi-syllabic MAMA and tri-syllabic MAMAMA after hearing tri-syllabic MAMAMA. Her internal and potentially innate tendency to repeat was exposed to parsable and reproducible stimuli: result was one among first bipartite micro-dialogues noted down.

The interplay between the two processes became more salient half a year later when IM started to consistently use her private words in recurrent contexts. Parents could therefore quite easily decode "meanings" of such intralexically repetitive terms as BIBIBIBI (f1;6;12) [in presence of a "baby"]; ANAN (f1;10;15) [when requesting a "banana"]; VAVA (f1;8;0) [when playing with "water" or NANA (f1;8;6) [when looking into mirror][8]. Given that these words do not exists *per se* in neither in German or Slovak (but, as will be shown in 12.10.1, some can be understand as cross-over forms between the two languages), they had gradually disappeared from IM's lexicon.

Disappearance of these pre-syntactic, protolexical structures notwithstanding, intrasubjective replication did not cease to play important role in development of IM's linguistic faculty. Consistently with what is known in litterature, such repetitions prevailed whenever IM became aware of existence of a new form, whose articulation was to be perfectioned and mastered. For example, after having understood that a difficult-to-pronounce form AUTOBUS refers to instances of

8 Later, at (f1;11;16) it was noted down that IM tended to use the term ICH when she was agent of the action and NANA when she was receptor or benefactor of the action.

large, noisy, useful yet dangerous species, IM produced (f2;0;2) the term 63 times in less than 30 minutes. Given that during this time interval there was sometimes no autobus in sight and given that the articulatory sequences where sometimes interrupted by minutes-lasting pauses or sequences dedicated to other topics, one is obliged to explain such loops in terms of structures and processes whose temporal span extends well-beyond the milisecond- or second- span of the standard Millerian short-term memory.

At the end of IM's toddler period, we constate that plain intrasubjective replications are more and more rare. Sometimes, they still occur when the child is playing alone, especially with water or her child-ressembling puppets. Or they occur in situations where the term is too difficult to pronounce on its own (i.e. IM's pronounciation of SAMBASAMBhAVA when exposed to the picture of the buddhist saint Padmasambhava (f2;9;3)). And some intrasubjective repetitions are still observable in communative scenarios (e.g. when saying MEINE, MEINE in order to emphasize that a certain toy or food should not be taken away). Whether these cases still represent a rudiment of a subjacent cognitive processus, or whether they are simply expressions of structures which were culturally acquired[9] opens an argument which we have no intention to enter.

END REPETITIONS 12.6

12.7 FIRST CONSTRUCTIONS

While repetitive sequences can be rightfully considered as "constructions" because they contain multiple juxtaposed (con-) elements (-structions), we label as "constructions" only such expressions which fulfill following conditions:

1. they contain sequences of two or more elements which, when taken alone, are distinct from each other
2. basic elements are at least as complex as a morpheme or a syllable

Under such definition, the sequence "ma" is not a construction because it does not fulfill the second condition ("m" and "a" are neither morphemes nor syllables); and the sequence "mama" is not a construction because its basic elements (two "ma" syllables), when taken alone, are not distinct from each other.

9 Note that rhetorical figures as diverse as antanaclasis, epizeuxis, conduplicatio, anadiplosis, anaphora, epistrophe, mesodiplosis, diaphora, epanalepsis, diacope or chiasm all exploit, in one way or another, the impact of repetition upone one's C-structures. Note also that "reduplication" is a phenomenon observed in practically all major language families of the world.

Under such definition, products of plain intrasubjective replication are not to be considered as "constructions". What is needed in order to obtain "constructions" thus defined, is not only replication, but also variation.

12.7.1 FIRST WORD COMBINATIONS

The first observed, decoded and registered multi-word combination which IM had uttered was: MAMA NENE (F1;4;25). Construction was uttered in context of a request for breast-feeding.

Note that without previous knowledge of what NENE (12.5.1) means, it would be impossible to decode the signal as a legitimate phrase on its own and given its $C_1V_1C_1V_1C_2V_2C_2V_2$ structure, it could be even considered, by an external observer unable to parse IM's idioglossia, to be a meaningless babbling fragment.

A month later, at (f1;4;30) IM uttered the expression TATO MAMA BABA ALA. Given that IM called her paternal grandmotherwith the nickname ALA, IM's mother immediately understood the 4-word (!) utterance as a nominal phrase meaning "father's mother is grandmother Alena". Three weeks later, at (f1;5;23) IM pronounced the utterance MAMATATO immediately after waking up, potentially requesting attention of (or greeting?) both parents by means of concatenation of both parental terms. At (f1;8;6) it was suspected that TATOTUTO means "here, father" since "tuto" is a standard local demonstrative of Slovak language. But the advent of full-fledged two-word stage was noted down at (f1;8;9) when IM had said in the interphone "HALO TATO" while usually she was saying *either* HALO *or* TATO.

Only two weeks later, at (f1;8;23), the mother has immediately decoded the expression AJs NANA MAMA AKUKE as ditransitive construction meaning "ice-cream me mother buy". Given that the utterance was indeed produced in the proximity of an ice-cream stand, and given that it was accepted by both parents at least since (m1;8;23) that AKUKE means "einkaufen" (e: to shop, s: nakupit) and taken as granted that IM uses the term NANA to refer to herself (c.f. previous section), the father would be obliged to set aside his scepticism and buy both girls an ice-cream if ever the younger one would not fall asleep in the meantime.

END FIRST WORD COMBINATIONS 12.7

12.7.2 FIRST PIVOT(S)

"Names" like MAMA, TATO or NANA were sometimes used as components of longer and more complex constructions. They were also partially productive in a sense that some of these constructions (like

MAMATATO) were never uttered by the parents and thus could not have been learnt by rote. But before introduction of pivot words, productivity of such nominal terms was highly restricted: they never occured in less than a handful of constructions.

Things changed with arrival of the first pivot term, and in case of IM, it was the term AUCH (meaning "too", "also"). Thus, an (f1;10;0) entry mentions following constructions: TATO AUCH (as a father-addressed request to eat as IM does); MAMA AUCH; NANA AUCH (when requesting to eat same food as parents eat); ENTE AUCH (when feeding ducks). The pivot AUCH could be thus understood as a productive "seed" of a following micro-grammar:

MAMA TATO NANA ENTE	AUCH

Table 11: IM's seeding grammar: AUCH at ultimate position.

Depending from the context and agent-term of the construction, the pivot carried meanings as diverse as imperative "You (father) do that (eat) as I do", declarative "I do it (put clothes) as You do" or "They (ducks) also want to eat". In general, it seems that the term was quite closely related to the fact of imitation and/or to the fact of intending that activities of two distinct agents should be aligned. Thus, the next recorded constructions were:

ICH AUCH	NACH HAUZE (f1;10;15 - When wanting to go home) AKE, NANA AKE (f1;10;17 When seeing father swinging on a seesaw; AKE=g:schaukeln, e:to swing) YOGA (f1;10;30)

Table 12: Seeding grammar extended: AUCH in the central position.

Note, however, that the term AUCH allowed IM to articulate thoughts encompassing realities well beyond here&now. For example, when watching the scene of her favorite animated movie in which the benevolent mole bottle-feeds an orphaned eagle, IM declared: ICH AUCH MI (f1;11;1) meaning something like "I am also used to drink milk". Or putting the milk-bottle on mother's breasts and saying NENE AUCH (f1;11;5). Or, when reading book about Babar the elephant (f1;11;5) IM stated ICH AUCH AM (f1;11;5 AM=tram) when observing picture on which which Babar the elephant takes the tram, potentially intending to declare that she also takes the tram; few pages later, ICH AUCH LIEALO (F1;11;5 LIEALO=s:lietadlo, e:airplane)[10]) was

10 In a sort of congitive and phonotactic economy par excellence, IM had consistently used the slovak signifier LIEtAdLO when mentioning airplanes in her otherwise germanophone constructions. Cognitive: airplanes were strongly associated with departures and arrivals of slovak-speaking TATO. Phonotactic: it is definitely easier for a child to pronounce a word full of laterals and vowels than the german "flugzeug" con-

declared when observing picture on which Babar exercises yoga on the airport. During the same evening reading session it was, however IM's act of uttering ICH AUCH accompanied with pointing to the image of the Eiffel tower which made both parents to feel utterly perplexed. Not only because IM had indeed, visited the Eiffel tower more than 2 months before, but also because no "ICH AUCH" was uttered during the lecture of subsequent pages, on which Babar exercises yoga in Yosemite park, near Golden Gate bridge etc...

Given the recurrence of the construction ICH AUCH, one would be tempted to state that it was this longer complex and not the simple AUCH which was the true pivot. But this was no the case since more than often, AUCH agglutinated to and with other agential terms than simple AUCH. Thus, TATO AUCH LIEtAdLO (f1;11;2) was uttered when observing airplanes on the sky; expression TATO AUCH UHE (UHE=g:schuhe,e:shoes) ordered father to put on his shoes. What's more, in her pre-sleep monologue of (f1;11;4), IM had spontaneously generated all utterances given by the paradigm: and did so in a repetitive and combinatorial fashion (i.e. produced all 6 combinations) normally common to scholastic methods or text-books in secondary language acquisition.

ICH MAMA TATO	AUCH	AJA (e: egg) KUCHEN (e: cake)

Table 13: Another AUCH-centered paradigm.

This being said, both parents unanimously agree that IM's first pivot "strong" enough to structure around itself whole system of constructions, was the intersubjective term AUCH. This pivot was only slightly antecedent to gain of force of other pivot, namely the egocentric MAJnE (d:meine,s:moje,e:my) expressed at (m1;11;2) in such utterances as MAJnE MAMA or MAJnE MIAU. Soon after, these phrases were also cried out from sleep: MAJnE MIAU at (f2;0;0, f2;0;21) MAJnE MAMA at (f2;0;21). But it was already at (f1;11;21) that this "pivot of personal property" was already strong enough to cause IM to cry out the expression MAJnE UHE (my shoes!) amidst the REM-phase of one of her sleeping cycles.

Somewhat contrary to other children reported in litterature, IM had started to use only relatively late her term MEA (meaning d:mehr,e:more) as a productive pivot. Often, she had simply used other means (including the usage of AUCH or NOCH) to express longing for bigger quantities of food or for reproduction of certain action.

END FIRST PIVOT(S) 12.7.2

taining such phenomena as voiced velar occlusive juxtaposed with an affricate. Being reassured that she masters the syllable "LIE" well, IM had later consistently prefered to use the term LIEnKA (meaning "ladybird") instead of German "Marienkafer".

12.7.3 FIRST MICRO-GRAMMARS

Once pivot words had helped IM to "understand" the meaning-specifying expressive force behind the act of juxtaposition of specific tokens, IM had swiftly and naturally proceeded to the application of such "combinatorial trick" in other contexts and for other uses. Asides protoislands of order structure around AUCH and MAJnE, instances acceptable by following micro-grammars were noted down (f2;0;7) as most salient and recurrent:

$$\begin{aligned} A_{gent} &\rightarrow \text{MAMA} \mid \text{TATO} \mid \text{NANA} \mid \text{ICH} \mid \text{BABA} \mid \text{BEJBY} \\ G1 &\rightarrow A_{gent}\ \text{AUCH} \\ P_{atient} &\rightarrow \text{MIAU} \mid \text{METE} \\ G2 &\rightarrow \text{MAJNE}\ P_{atient} \\ F_{ood} &\rightarrow \text{BROT} \mid \text{AJA} \mid \text{ANAN} \\ D_{rink} &\rightarrow \text{MI} \mid \text{VAVA} \\ A_{ction} &\rightarrow \text{HAJI} \mid \text{ESSEN} \mid \text{TRINKEN} \\ G3 &\rightarrow F_{ood}\ \text{ESSEN} \\ G4 &\rightarrow \text{Drink TRINKEN} \\ G5 &\rightarrow \text{Action MACHEN} \\ G6 &\rightarrow A_{gent}\ \text{KOM}\ A_{ction} \end{aligned} \tag{2}$$

Grosso modo, this proto-grammar already includes references to those actions (eating, drinking) and agents (parents, self) which are most vital for IM's survival. But in rules G5 and G6, one can already observe "the seed" of much more general a knowledge, a knowledge that certain precise actions can be "made" (G5) and, in a sort of half-imperative, half-causative fashion, other agents can be incited to "come" and actualize them (G6). From such knowledge, child is only one cognitive step far from the reflected and conscious meta-knowledge of the fact that it is by language and language alone that such precise incitations can be made.

From there on, whole evolution of IM's syntactic P-structures has become complex, filled with non-monotonic returns, iasynchonic detours, parallel developments and both intra- and inter- insular population dynamics. Since sufficient accounting of such development would demand a book on its own, let's know shift away from terminology of "grammars" towards more dynamic a terminology speaking about "mutations", "crossovers" and "life".

END FIRST MICRO-GRAMMARS 12.7.3

Many constructions hereby presented would not have been successfully decoded if the parents had not accepted as granted that at least some among unintelligible productions of their daughter are, in fact, complex utterances. Almost a year later, at the very end of the toddler period, none of the parents is able to correctly understand the meaning behind all child's utterances. Many are still unintelligible and very often, the child must still take recourse in other means of information-passing (e.g. pointing, gests, facial expressions) to make herself understood.

What is, however, understood by both parents as well as by wider social surrounding is that at 2;6, IM diposes of rich internal world of dreams, intentions and playful tendencies which strongly influence what|when|how she interacts with her environment, both verbally and not verbally. Perhaps there is meaning, perhaps there is communicative intention behind any sequence which the child utters, no matter how unintelligible the sequence may sound. Or perhaps not and child simply explores the limits of language[11] by joyful playing of the most fundamental among all the language games (Wittgenstein, 1953).

Be it as it may, the process of language ontogeny brings into the world an unprecedented amount of novelty. True, novelty having the form of an unstoppable tantrum can sometime destroy one's day. But luckily for all the parents of the Earth, the positives seem to outweigh the negatives. Thus, most of the time, verbal interaction with children is simply beautiful, comforting and -let's not forget the another important aspect motivating all parties involved - are child's first linguistics constructions perceived and *felt* as cutely and adorably funny.

END FIRST CONSTRUCTIONS 12.7

12.8 MUTATIONS

Mutations (from lat. *mutare* "to change") are basic atomic units of change. Mutations occur in time; in informatic terms mutations are events caused by transition of information-encoding substrate from one state into another. Given that the physical nature of substrate of linguistic representations is still speculative, and in great extent unknown (8.6), we shall present, in the following paragraphs, just a handful of illustrations of such transitions occuring in ontogeny of IM's linguistic structures and processes.

12.8.1 CONTEXT-FREE SUBSTITUTIONS

Context-free substitutions are mutations characterized by substitution (replacement) of each occurence of the original symbol $S_o rigin$

11 If the statement « The limits of my language are the limits of my world» (Wittgenstein, 1922) is true, than agent's exploration of limits of her language *equivauts* the exploration of limits of her world.

with exactly one instance of the target symbol S_target. Given that all occurences of S_origin are substituted, CSM operators are, so to say, agnostic of substituents position.

A first example was already given: transition MəMə → MAMA (c.f. 25) can be explained as a substitution of a central vocalization ə for a more marked A, i.e. as a result of application of a rule ə → a.

Other particularly illustrative example of a CSM was given by IM on three **consecutive days**, during which she was observed to utter sequences of a form

BABIJÁ (f1;4;16)
MAMIJÁ (f1;4;17)
PAPIJÁ (f1;4;18)

Within the framework of the theory hereby proposed, such transitions could be explained by mutation of the content attributed to non-terminal $C_{lab,occ}$ within the template:

$$C_{lab,occ}aC_{lab,occ}ij\acute{a}$$

which is equivalent, at certain level of abstraction, to substitutions $b \rightarrow m$ and $m \rightarrow p$ which most probably occured in IM's mind during the first (resp. second) night between the observations.

Note, however, that in spite of being labeled as context-free, even these mutations are not "global". It would be utterly false to believe that the fact that every B within the construction BABIJÁ was substituted by P resulted in the situation whereby IM ceased to pronounce the sound B alltogether. This was, of course, not the case and the sound "B" did not disappear from IM's repertoire. Thus, in regard to the local "template" in which it occured, the substitution could be considered as context-free. But not more: the mutation had practically no impact beyond the local micro-grammar within which it took place.

Or, to come back to the example of the primary differentiation Figure 25, the fact that ə was replaced by A in case of insula slowly converging to meaning of "mother" did not have any impact whatsoever upon the fact that within the insula slowly converging to meaning of "breasts" another mutation (i.e. MA → NE) took place. This is so because in moment of the mutation, both insulae were already materially encoded in at least partially distinct neural loci.

To summarize: context-free mutations are mutations which alter all instances of a certain symbol. But the scope of their action is still constrained to only a specific template | insula | micro-grammar [12]. Or a restricted group of these.

END CONTEXT-FREE MUTATIONS 12.8

12 In following sections we shall use terms *template*, *micro-grammar* and *insula* in a mutually interchangeable, synonymic fashion to mark the fact these notions are computationally equivalent

12.8.2 CONTEXT-SENSITIVE SUBSTITUTION

The scope of impact of a context-sensitive mutation is also constrained to a specific template or to a strongly restricted group of these. But in addition to this constraint, scope of applicability of CS-mutation is also limited by the context | position | neighborhood within the template itself. To illustrate with first well-documented CS-substitution: during her stay by IM's czech-speaking BABA, the mother has documented IM's production of expressions MAMI and BABI (m1;4;9). Emergence of these forms, which are completely correct vocatives in czech, could be explained by a context-sensitive mutation A$ → I$ [13] occuring in IM's mind. In the observation journal, mother had commented the phenomenon: *I suppose these came because of my calls "Babi" tu my own grand-mother and "Mami" to my mother.*. Further analysis can unveil, however, that acquisition of such vocatives could have been synergetically catalyzed by the presence of a dog called DEXI and a cat JESI in grand-mother's appartement. Since it was one among first IM's exposures to animal life and since IM did not hesitate to establish not only visual, but also verbal (by production of onomatopees like HAU-HAU and MIAU-MIAU) and haptic communicative interlock, it is undoubtable that representations (i.e. signifiees) of both pets attained a highly salient status within IM's mind. And given that in czech language vocative forms of I-terminated animal pet names are identic to the nominative forms, it cannot be excluded that the very presence and saliency of pet-denoting −I$ protonominals had stimulated IM's nominative-to-vocative transition within more general a class of living beings.

First vocatives

Thus, IM's success in mastering of vocatives seems to be result of interplay of three mechanisms:

1. an endogenous mutation which caused the −A$ → −I$ transition within certain among IM's private P-structures
2. exogenous gold-standard structures (i.e. persons in IM's social environment which use −I$ nominals within certain contexts)
3. a cerebral mechanism reinforcing or even replicating such private representation which match public structures

Nature of these three mechanisms correctly understood, one can see how development of practically any expressions - from initial babbling all the way through infantese, toddlerese, pupilese to the "correct" adult-like pronounciation - can be characterized as a sequence of such CSSs. In IM's case, for example, one can see the trajectories along which the words for "milk", "water", "baloon" podded out of the initial babbling:

13 Consistently with the syntax of Perl Compatible Regular Expressions (PCREs) we shall denote the "ultimate position" with the dollar sign $.

Context-sensitive substitutions (EXT)

MiMi*[14] → MI (f1;5;12) → MICH (f2;0;13) → MILCH
UaUa* → VAVA (f1;8;10) → VASA (f2;1;8) → VAS
BALaL* → BALOL (f1;10;30) → BALOND (f2;0;13) → BALON (f2;4;19)

END CONTEXT-SENSITIVE SUBSTITUTIONS 12.8.2.0

In these cases, mutations had often counteracted child's tendency for elision, assimilation or fronting of certain phonemes at certain positions (9.2.1). In each example the symbol → tends to denote a moment, or a group of moments whereby IM's linguistic structures underwent a structural change, i.e. mutation.

In reality the situation is, of course, much more continuous and much less discrete than in our transcriptions. To describe whole phonic development more closely one would have to use a more refined transcription alphabet (e.g. International Phonetic Alphabet) but even this one could be criticized as too coarse-grained for the task at hand. But no matter what transcription system would one choose, independently even of the fact whether one stays faithful to continuous reality or discretize the phenomena in already existing boxes, one thing stays certain: IM's interiorization of any individual linguistic structure consisted of multiple intermediate steps.

END CONTEXT-SENSITIVE MUTATIONS 12.8

By stating that *development of any individual linguistic structure consists of multiple intermediate steps* we want to focus reader's attention to the fact that not only P-structures and articulated signifiers develop, but - and this is important - also any C-structure (i.e. conceptual signifié) as well as structures relating the two do so as well. In preceding paragraphs, we have focused mainly on development of P-structures because their development is easier to assess. But this does not mean that the world of C-structures does not develop i.e. that it is not subject to mutations.

Contrary, in fact, is the case: in course of her development, IM's innermost structures had been constantly modified by multitude of events of exogenous origin. By myriads of minute interactions and couplings of linguistic inputs with other auditive, visual, haptic, olphactoric, gustative, vestibular, nociceptive or placiceptive inputs. By parental questions and parental corrections and by facts that a certain question and a certain correction were given in one context but

14 We use the star sign * to denote expressions which were not recorded in the observation journal but are considered as plausible protoforms. This is similar to comparativist tradition in which the *-sign denotes hypothetic forms postulated by the theory but not attested by any existing corpus.

not in another. But other, more endogenous factors related to playing, dreaming and φαντασία well beyond traditional adult notions of "abstraction and generalization" had to be active as well, in order to account for emergence as well as correction of such cases of *poietic* over-generalization as:

1. at (f1;11;8) saying ZONE (e: sun; d: Sonne) when seeing a full-moon in the evening sky
2. at (f2;0;19) naming the circle of light projected by the lamp upon the bedrooms's ceiling with the term BALONd
3. at (f2;3;19) saying LIENKA (e: ladybug) when seeing, on a picture in a picture book, a red ball with white dots
4. at (f2;5;8) using the term KUGEL (e:sphere) to describe ping-pong ball (correctly called "BAL" a year before)
5. at (f2;6;15) answering NENE when asked to describe what is on a swimming-pool tile with two concentric circles
6. and the DING-DONG mystery

12.9 CASE STUDY OF SEMANTIC MUTATIONS: THE DING-DONG MYSTERY (APH)

To demonstrate the arbitrairness of any system of categorization or even any epistemology, both Michel Foucault as well as Eleanore Rosch fondly cite the taxonomy fictitiously attributed, by Jose-Luis Borges, to an ancient Chinese encyclopedia entitled the Celestial Emporium of Benevolent Knowledge:

« On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance.» (Borges, 1952)

Such a Borgesian account is something which has to come, willy-nilly, to one's mind when confronted with the case of the DING-DONG mystery (DDM). Contrary to Borges, however, is DDM not fictitious but rooted in reality of facts. These are as follows:

First mention of DING-DONG (f2;0;7) clearly mentions the term in context of church bells. Indeed had used the expression to express *Cathedral* her will to be in the proximity of Bratislava St.Martin cathedral exactly at 18:00 when cathedral's bells ring the most. The same record mentions, however, that the term cannot refer to "church" in general since another church building was labeled as OKOL (s: kostol).

That the concept develops started to be evident a month later (f2;1;4), when it was noted down, during the visit to the library, that IM had picked from a bookshelf a book about European history and labeled the building depicted on the front cover as DING-DONG. A week later, the (f2;1;8) record continues: *You are still occupied with the DING-DONG concept. It seems to denote all big buildings, today, for example, have You seen the picture of **the skyscraper** and called it a DING-DONG.*

Big building

A later (f2;2;1) log indicates when things started to get somewhat more complex. Thus, during a simple walk between Berlin's central station and Hackesheer Markt, IM used the term DING-DONG when labeling following objects:

- stone sculptures on the bridge
- tower in the distance
- fluttering German wing atop the Bundestag
- a building just next to Bundestag
- synagogue's golden dome
- cross atop the Berlin's cathedral
- buildings of Marienkirche and Boden museum

which indicates that at that period, the DING-DONG concept still overlapped with something similar to an adult concept of a "fancy piece of masonry" or "building's top". The fact that the later (f2;2;11) log states "*Still occupied by DING-DONG, You were completely fascinated by youtube videos of St.Martin's cathedral.*" seems to support the hypothesis.

While (f2;2;18) log entry stated that "*DING-DONG fades into background*" it also stated that "*from time to time, You still label something with that term: picture of castle in the book, **two noodles stuck together**...*".

Castle and noodles

The entry logged at (f2;5;7), i.e. 5 months after initial use of the term, states *with the word DING-DONG You have labeled the picture on the "ace of staffs" Crowley's tarot card as well as **a flute**)*. And two weeks later (f2;5;21), it was written that "*You are still occupied with DING-DONG. It seems that You use it especially to denote spiky things, for example **green buoys on the Elbe river** are DING-DONG. But the red ones, without the spike, are not*". Approximately in the same period, father also considered as quite plausible the hypothesis that the term can also denote the property of being long (d: lang).

Tarot card and a flute

Spiky green buoys

Given the importance of the term within IM's world, a small constructional island coalesced around it. At (f2;3;19) a recurrent usage of the construction dING-dONG mACHEN was noted down when building towery lego churches[15], at (f2;5;8) intense repetitive production of expression ING ONG OJTET (d: lautet e: rings) was noted down and the (f2;5;23) recorded following playful variations:

dING dONG lOJTET
dING dONG lOJTET
dING dONG lOJTET
lOJTET dING dONG
lOJTET dING dONG
lOJTET dING dONG

which repeatedly transgress even the most primitive subject-precedes-verb syntactic rule of german language.

Such indeed is the mystery of DING-DONG: unconcerned by the "correct word order", unconcerned even by appropriate, adequate and optimal conceptual boxes, infantine mind plays the poetic game. Shamelessly, joyfully and naturally plays the poetic game, and do so at all levels. Isn't that Borgesian?

END THE DING-DONG MYSTERY 12.9

The above aphorism indicates that asides context-free and context-sensitive substitutions, yet another variation operators are at play in a developing mind. Not only formal but also semantic substitutions, not only replacement of symbol by an empty one, but also diminution (or expansion) of extension of a concept C (when C is understood as a set).

Or, when concepts are understood in more geometric terms, mutations consisting of either increase or decrease in volume of the C-localizing subspace or translation of C's centroid to some other position.

But as was already indicated not only by IM's playful switch from grammatical dINGdONG lOJTET to agrammatical lOJTET dINGdONG-during, but firstly the enumeration of different intralexical metatheses (12.9.1), yet another class of mutation operators seem to act within the developing mind: switching of position within the sequence, permutations within the temporal order.

12.9.1 FIRST TRANSPOSITIONS

A transposition occurs when two or more elements of a bigger whole (e.g. phonemes within a word or words within a phrase) exchange

15 However, a general term for other constructions built from lego or wooden cubes was BAUT (f2;2;18, f2;3;19), potentially derived from past participle of to build (d: gebaut). What's more, when asked to label diverse lego blocks, IM had consistently used the term BAUK (f2;2;1). Note that such a term does not exist neither in slovak nor in german.

their position. Frequency of occurence of elements within the sequence thus does not change, relative positions, however, do.

Already a relatively early transcription (f1;5;30) of IM's sometimes babbling, sometimes one-word "stream of consciousness" improvisations, produced at the breakfast table contains sequences like ÁU,UÁ and ÍTÁ, ÉTÍ. Such were indeed IM's first tentatives to switch positions of two protophonemes in her protowords.

IM's later productions indicated activity of more complex (i.e. involving more than 2 transposed elements) metathesis-like reorganizations as:

Context-sensitive metatheses (EXT)

APUK (f2;1;10) → "kaput"
IPEK (f2;1;12) → Wipke
UKAKS (f2;1;24) → "Rucksack"
MAKTA (f2;6;0) → "matka"
etc. END CONTEXT-SENSITIVE METATHESES 12.9.1.0

Given the prominence of such "errors" and mis-productions[16] in IM's speech, we are tempted to state that a non-negligeable amount of transpositions commonly studied in evolutionary linguistics (e.g. e: "fog", s: "hmla", czech: "mlha") had their origin in slips-of-tongue of individual toddlers which were subsequently accepted and spread through wider community.

At the end of this development, IM started to permutate position of not only individual phonemes, but also of phonetic clusters or even words within phrases. Thus, (f2;4;5) mentions a permutation

HUNDUNDMIAU
MIAUUNDHUND

undoubtably stimulated by a symmetric, non-preferential binary coordinative UND (e: and). But as demonstrates already the following entry (f2;4;6) noted down during a game whose objective was to put diverse wooden animals into correspondent slots:

DA IST KKO, IST DA KKO

as well as an entry noted down a day later, during the marble game

KUGEL EINE, DA IST KUGEL EINE

IM's propensity to permutate the order often didn't worry much about even the most fundamental among the syntactic contraints of germanic languages. That is, the constraint that the article (EINE)

16 From adult-like point of view

should precede the noun (KUGEL) and definitely not the other way around.

In the following chapter, dedicated to quantitative analyses of the CHILDES corpus, we shall aim to shed somewhat more light upon the question whether this situation - in which IM's urge to permutate word order was stronger than the most fundamental among the syntactic constraints - was peculiar to IM who, as partially slavic person potentially feels less bounded by the need to correctly prefix substantives with determinants; or more general a trend present even among germanic and anglosaxxon toddlers.

END FIRST TRANSPOSITIONS 12.9.1

In above subsections we have presented multiple variation operators which, we believe, could be rightfully labeled as "mutations". Substitution of nothing with something, something with nothing, something with something else; expansion or diminution of extension; switches in positions, fillings of empty slots: in one way or another, all this was already known not later than after Aristotle and his followers (8.5). It was evident to Godel and Turing as it is evident to proponents of FLG: any computable number (resp. construable string of symbols) can be obtained by means of insertions, deletions, substitutions and transpositions.

Until now, practically nothing new in comparison with traditional cognitivist symbolic architectures. But everything changes when the most noble among all variation operators is introduced: the crossover.

END MUTATIONS 12.9

12.10 CROSSOVERS

Boldly speaking, crossover is the operator of *unita in diversitate*. As indicated by the figures attached to tour brief discussion of biological evolution (Figure 4) and fitness landscapes (Figure 8), the power of crossover consists in its ability to:

1. let two (or more) parent structures to project their features upon one (or more) child structures

2. allow the evolving system to get out of locally optimal states (i.e. to fly away from the peaks into unkown realms in between)

The second point systems implies that systems involving a crossover operator are able to continue evolving even there and then where other, gradient-following approaches are doomed to get stuck. Computationally speaking, crossover's ability to direct the search into regions of *useful unknown* yields the ultimate *coup-de-grace* which the model involving the operator apdodictcically gives to any model which does not do so.

When it comes to the first point, consider, for example, a sort of semantic crossover:

$$\frac{\begin{array}{c}(\text{HUMAN})\\ \times\,(\text{WINGS})\end{array}}{(\text{ANGEL})}$$

Without resorting to cross-over mechanism and without falling into a trap of pseudoscientific divinatory explanations, we consider it as very difficult - if not impossible - to offer a scientific account of a cognitive mechanism by means of which all angels and chimeres, all centaurs and mermads, all mythological visions as well as technoscientific insights, how *representations of entities without a material* referent could have ever entered the mind of a primordial man.

And by whom else if not by children could one see such process act?

12.10.1 MULTILINGUAL CROSSOVERS

Given the lucky coincidence (12.3) due to the fact that structures in IM's linguistic environment primarily consisted of structures coming from two distinct sub-branches (i.e. germanic and balto-slavic) of the same language tree (Figure 5), IM was exposed, on a regular basis and in analogous contexts, with instances of constructions which were both similar and distinct in the same time.

What resulted is a phenomenon of interlingual "mixing" which is well known to practically every parent of a healthy bilingual child. Let's now focus on two most prominent types of such mixing.

Intralexical crossovers

A multilingual intralexical crossover is a mixing of two word-representing schemas S_L and S_J intersubjectivly replicated from exogenous oracles using diverse languages L and J. Given that the schemes-to-be-combined are defined as word-representing, and given that their sources are oracles (e.g. parents, grandparents, teachers), they are expected to occur during:

1. cases of bi- or multi- lingual language acquisition (hence "intralexical")

2. acquisition of base-level terms (i.e. signifiers) for base-level meanings and referents

In the following exposure, crossovers shall be presented consistently with following formula:

referent
german
slovak
TODDLERESE

whereby the first row shall contain the English term for the referent R, second row shall contain the R-denoting term most frequently used by IM's mother and third row the R-denoting term used by IM's father. The last row of every example shall contain the transcript of IM's idioglottic productions consistently (i.e. more than once) uttered in R-cooccurrent context.

First three multilingual intralexical crossovers all noted down at (f1;7;30) were:

eyes
augen
oči
OGE

and

water
vasə
voda
VAVA

and

shoes
šúhe
boty
$\text{OG}_{\text{H}}\text{E}$

Another couple of salient crossovers was noted down during game with animal picture books, at (f1;8;11)

monkey
afe
opica
API

and at (f2;1;9)

elephant
elefant
slon
OLOOND

Asides substantives, other parts-of-speech were mixed as well. Verbs, for example (f1;8;24)

$$\frac{\text{buy}}{\begin{array}{c}\text{ajnkaufen}\\ \text{nakúpiť}\end{array}}\\ \text{AKUKE}$$

as well as possessives (f2;2;22)

$$\frac{\text{my}}{\begin{array}{c}\text{majne}\\ \text{moje}\end{array}}\\ \text{MAJE}$$

Many other cases were noted down where IM had opted for a form which shared as many features as possible with forms in both ambient languages. Thus the (f2;0;14) entry recorded

$$\frac{\text{stick}}{\begin{array}{c}\text{štok}\\ \text{papek}\end{array}}\\ \text{AK}$$

and it has to be added that in following weeks to come IM had used the term AK and | or əK to denote practically any piece of wood she could easily carry and manipulate. This was done in spite of initial parental tentatives to correct her and resulted, in fact, to parental adaptation whereby parents resigned and used the convenient term AK es well.[17]

We believe that these examples illustrate that in many cases of production of new words, IM tended to:

1. produce forms with known characteristics
2. produce forms which are as close as possible to both parental forms

The first tendency seems to be the case for all healthy children, no matter whether they are raised in monolingual or multilingual environment (c.f. Table 2 and the associated discussion of "preference" and "avoidance"). When it comes to second, centroid-form-seeking tendency, it is definitely most easily assessed in case of multilingual acquistion wherein the forms to be crossed-over are distinct.

To prove our point, we conclude this brief enumeration of IM's multilingual intralexical productions with the final example (f2;0;17)

$$\frac{\text{drink}}{\begin{array}{c}\text{trinken}\\ \text{pije}\end{array}}\\ \text{PIJEN}$$

17 The form əK had withdrew into the background once IM mastered the correct pronounciation of more correct forms OK (f2;6;13), TOK and ŠTOK. The form əK however soon reappeared in order to mean "wolf" (sk: vlk).

as well as with a link http://wizzion.com/thesis/videos/pijen.mp4 which, between 4:20 and 4:34 (as well as at 5:47), demonstrates our case.

END INTRALEXICAL CROSSOVERS 12.10.1.0

Intraphrastic crossovers

Multilingual intraphrastic crossovers are crossovers which mix, within one construction, morphemes originating from multiple languages. They are well known to practically any person subjected to second language acquisition: one wants to form construction in language 1 but somehow "unvoluntarily" populates it with certain items proper to language 2.

IM started to produce her first intraphrastic mixes when she was still in the "pivot grammar" stage. It had thus often been the case that certain slots within constructions pivoted by german AUCH were filled with an item of slovak origin:

```
(f1;11;3) TATO AUCH LIEtAdLO (nom. sg. sk: "airplane")
(f1;11;4) ICH AUCH LIEtAdLO
(f2;0;0) NANA AUCH rUKY (acc.pl. sk: "hands")
(F2;0;7) LIENKA (nom. sg. sk: "ladybug") AUCH
```

Alongside these AUCH-pivoted intraphrastic crossovers, IM's production was also full of utterances composed of slovak noun and a german predicate. For example, during the period dedicated to story of mole and eagle, following utterances were very common

```
(f1;11;12) OLOL (nom. sg. sk: "eagle") šAUEn (inf. de: "to watch
    ")
KKO (nom. sg. sk: "mole") šAUEn
(f1;2;18) OLOL ÍflGt (3p. sg. pres. de: "to fly")
```

It is, however, discutable, whether one could count such constructions as "interlexical" crossovers . This is so because in concrete cases of usage IM had used the term KKO, OLOL etc. similiarly to terms like TATO, MAMA, BABA, i.e. as personal names. Not knowing other instance of eagle or mole than the one which was presented to her it seems more plausible to state that IM has juxtaposes language-agnostic *names* and not language-specific *nouns* asides her german-originated predicates.

But 6 months later, with her toddler period coming to an end, a sudden phase transition in both amount and diversity of intraphrastic crossovers had occured. Thus, during interval of three days only, production of following germanoslavic structures has been observed:

```
(f2;5;13) TIETO (sk: "these") ČpANuCHy (sk: "sock pants")
    AJNgekAUFT (de: "bought")
2 (f2;5;15) WO (de: "where") IST (de: "is") mOTYl (sk: "butterfly")
     ?
(f2;5;18) TO (sk: "that") JE (sk: "is") MAJNS (de: "mine")
(f2;5;18) NANDA tAM (sk: "there") BYVA (sk: "lives") , MAMA AUCH
    tAM BYVA
(f2;5;18) DA (de: "there") IST (de: "is") mUCHA (sk: "a fly")
```

Closer inspection of these examples may reveal that sometimes IM had used the slovak (ex. 3. and 4.) and sometimes the german (ex. 5.) forms to express the meaning "there is". On the very same day even a small multilingual grammar was noted down:

NICHt	chrOBÁČIK (sk: "beetle") mrAVČEK (sk: "ant") ČMELJAK (sk: "bumblebee"	)

Table 14: Interlinguistic micro-grammar.

Both parents are unaware that they had produced such "negation in german + slovak animate substantive" constructions. Given that it is highly unlikely that someone in IM's wider environment would expose her to such constructions, the sole explanation of their existence has to be sought for among IM's endogenous cognitive processes. We agree with Piaget that at this stage, one among such processes can be the child's egocentricity and her tendency to playfully negate any information that comes from exogenous oracles. And this had been, in IM's case, expressed notably by means of german pivots NAJN and NICHt whose productive affinity at this period was such, that they succeeded to form constructions even with non-germanic words.

end intraphrastic crossovers 12.10.1.0

On preceding pages we have presented few cases of multilingual crossovers, i.e. crossovers between schemas embedded in distinct languages. Two main groups - intralexical and intraphrastic - were introduced in order to organize the presentation. We consider as highly plausible that asides these two types, the super-group of interlinguistic cross-over contains other types of operators as well.

But instead of studying into detail each one of them, let's just close this brief discussion of bilingual acquisition with the aphorism stating that

Of crossover and calques (APH)

If the reader has understood that operators which we have labeled as "interlingual crossovers" could elucidate phenomena which traditional linguistics call as "calques" or even "faux amis", then the reader has understood us well.

END OF CROSSOVER AND CALQUES 12.10.1.0

and now focus upon crossovers occurent not among elements of multiple languages, but among elements of one sole language.

END MULTILINGUAL CROSSOVERS 12.10.1

12.11 MONOLINGUAL CROSSOVERS

A monoglingual crossover is a crossover between two or more input schemas which all originate and are extracted from the same language L.

A schema is the most fundamental element of the theory hereby introduced. It is a template, a pattern a sort of micro-grammar which, when embedded within human brain or within computational agent, can be useful for both comprehension and production. In comprehension, schema's role is to "match" an external stimuli (e.g. linguistic utterance). In production, e.g. when coupled with articulatory circuitry, a schema determines the process of generation and execution of a specific action (e.g. pronouncing of a word or a phrase).

Schemas themselves are composed of atomic features and it is important to realize that, in theory, one individual schema can integrate in itself features of different types: conceptual, semantic, syntactic or morphophonologic features can be considered as a constitutive elements of one individual schema S. In theory.

In spite of the fact that certain schemas S_X, S_Y can integrate in themselves "semantic" (i.e. signified) and "morphophonologic" (i.e. signifier) components in an extent which strongly ressembles entities W_X, W_Y commonly known as "words", it would be a mistake to simply state that words are schemas and schemas are words. For it may be the case that a certain word can be encoded by multiple schemas.

Let's know glance at few crossover types which indicate that it can be, indeed, the case.

12.11.1 INTRALEXICAL

Given that porridge with bananas was her favorite breakfast, the word denoting banana (de: banáne, sk banán) was one amongst the first items in her lexical repertoire. Thus, at (f1;5;12) it was noted down that IM consistently used the P-scheme BAJA to denote the fruit. Few months later, however, at (f1;10;8) it was observed that IM uses the P-scheme ANÁN to denote the same referent.

A month later, as IM had consistently used the incorrect pronounciation to ask for the fruit, the father had tried to exogenously induce the correct pronounciation:

```
IM: ANAN, ANAN
F: banan, ba, ba, baba, banan
IM: ANAN
```

without success since IM was still responding with pronounciation of ANÁN. But knowing that few months ago, there used to be a period where IM labeled the fruit with the schema correctly beginning with B, the dialogue continued:

```
F: baja
IM: BANAN
```

id est, IM had pronounced a correct form which she was unable to pronounce otherwise.

This pedagogical "success story" can be quite easily explained in terms of a monolingual crossover. Thus, knowing that IM used to produce the P-schema BAJA before, father had simply uttered the token which had reactivated the **latent schema**. Subsequently, during a moment of practically instantenous cognitive crossover, the latent schema mixed with the dominant schema:

$$\frac{\begin{array}{c}\text{BAJA}\\ \text{ANÁN}\end{array}}{\text{BANÁN}}$$

and the correct "centroid form" of two protoforms was obtained. END INTRALEXICAL CROSSOVERS 12.11.1

12.11.2 INTERLEXICAL

It may be the case that mind sometimes mixes together even the schemas which encode different semantic contents. Thus the first case of cross-over recorded by the father (f1;4;25) was the spontaneous usage of the vocative MAMI minute or two after the lecture of the book about the cat called MIMI:

MAMA
MIMI
———
MAMI

c.f. 12.8.2 for description of other exogenous factors which have primed IM for acquisition of vocatives.

Another couple of quite interesting crossovers was observed amidst IM's "eagle period" [18]. As was already mentioned, the word which dominated IM's production during the period was OLOL (sk: orol, en: eagle) and given the frequency of occurence of the term in IM's production, it is undeniable that the P-schema OLOL$ was strongly activated.

It may be for this reason that at (f1;10;30), IM's term for "ballon" was BALOL, which could be explained in terms of a cross-over:

OLOL$
BALÓN
———
BALOL

But since it could be argued that the production of the word BALOL could be also explained as the assimilation of the lateral feature by the terminating nasal consonant, and since we want to evit confusion between causes and effects[19], let's just focus on the second case which we consider as particularly instructive.

Thus it happened that at (f1;11;20), during her pre-sleep oratory, IM had tried to list names of all her kindergarten friends. But given that she forget to mention her friend Nikol, IM's mother had turned monologue into a dialogue:

M: NIKOL

IM: KOLOL

thus producing a word which does not exist neither in german nor in slovak, and doing so in a context which undeniably indicates that her communicative intention was to say "Nikol". One could, of course, argue that the production of such term as a result of avoidance of using the syllable ni- in the initial position (demonstrated, for example, by calling one of her friends KITA instead of Nikita) combined with the reduplication. But if such was the whole explanation, one could hardly see why IM had opted for the term KOLOL and not for *KOKOL. Thus, another force had to be at play and we argue that it was the productive affinity of the scheme OLOL and the subsequent crossover:

18 During this period IM was exposed, on her own request, to dozens or potentially even hundreds of instances of the same narrative concerning the friendship of the benevolent mole and an orphaned eagle. The exposure was multimodal: thus IM had sometimes watch the movie without commentaries, sometimes it was commented. Sometimes the picture book was read, sometimes the story without any visual support whatsoever was narrated. C.f. http://wizzion.com/thesis/videos/olol.mp4

19 What was first? BALOL or OLOL?

$$\frac{\begin{array}{c}\text{OLOL\$}\\\text{NIKOL}\end{array}}{\text{KOLOL}}$$

which have taken their toll.

Another interesting interlexical crossover was observed during another pre-sleep dialogue (f2;1;14). When IM was asked to describe games she plays in her kindergarten with another kid, she had answered with the word MAUEN. Given that such a word does not exist in german language and given that frequent usage of terms "mahlen" (en: to draw, to paint) and "bauen" (en: to build) was observed noted down already a month before (f2;0;16), it cannot be excluded that the term was a result of a following crossover:

$$\frac{\begin{array}{c}\underline{\text{paint and build}}\\ \textit{mahlen}\\ \textit{bauen}\end{array}}{\text{MAUEN}}$$

and that, potentially, it had a meaning of both building (e.g. lego, wooden cubes etc.) and painting (a common activity in IM's kindergarten) in the same time. If that was the case, IM's answer by means of the term MAUEN could potentially suggest that crossover may be useful not only for explanation of development of surface morphophonologic signifiers but also in explanation of much more deeper semantics- and concept-related signifieds.

END INTERLEXICAL CROSSOVERS 12.11.2

12.11.3 INTRAPHRASTIC CROSSOVERS

Monolingual intraphrastic crossovers are operators which mix together components (e.g. morphemes) originating from different phrase-encoding schemes.

Let's look at just one video[20], recorded at (f2;5;15), to see what this could mean. The video shows IM and her mother during a creative session initiated by stone-painting and terminated by sticking small artificial eyes on the painted stones. Many interesting things happen in the video, including:

1. within 304 seconds, IM uses the fixed construction UNKəBLAU ("dark blue") 18 times in three "bursts"
2. at 4:21, IM produces a multilingual crossover VO (de: "where") ISt (de: "is") OKO (sk: "eye" nom. sg.), subsequently is corrected

20 http://wizzion.com/thesis/videos/augen.mp4

by her mother which she immitates and produces the full slovak construction ĎE jE OKO at 4:25

In regards to monolingual intraphrastic mixing, it is already the first phrase:

ICH MAHLEN

pronounced at 9th second which is of certain interest. This is so because this phrase - agrammatical on its own due to the the non-agreement of the pronoun (1p. singular) with the verb form (infinitive or 1p. plural) - can be understood as a result of crossover of two grammatically correct phrases:

$$\frac{\begin{array}{c}\textit{ich mahle}\\ \textit{wir mahlen}\end{array}}{\text{ICH MAHLEN}}$$

The same holds, mutatis mutandi, for incorrect pronounciations which came later, such as

1. 03:08 VO ISt AUGEN? (where is eyes?)
2. 04:02 VO ISt mAJN AUGEN? (where is my eyes?)
3. 04:14 mAJN AUGEN VEG (my eyes (is) away)

Thus, all such syntactically incorrect constructions can be easily explained as a consequence of a crossover between correct forms which the child could have easily heard in her environment. For example:

$$\frac{\begin{array}{c}\textit{wo ist auge?}\\ \textit{wo sind augen?}\end{array}}{\text{VO ISt AUGEN?}}$$

This being said, we feel no need to spam the reader with other instances of such "monolingual intraphrastic" crossovers, produced by IM aplenty since cca. 2 years of age. Instead we conclude with yet another aphorism:

Of crossover and overgeneralizations (APH)

If the reader has understood that operators hereby labeled as "monolingual intraphrastic crossovers" could elucidate phenomenona which developmental linguistics label as "overgeneralizations", then the reader has understood us well.

END OF CROSSOVER AND OVERGENERALIZATIONS 12.11.3.0

In other words, the notion of "monolingual intraphrastic crossover" can be a useful conceptual aid for anyone aiming to explain the problem of over-generalization or to construct a theory thereof.

END INTRAPHRASTIC CROSSOVERS 12.11.3.0

Many among above-mentioned cases of monolingual crossover were triggered, induced or even primed by an exogenous event (i.e. parent asking or saying something). Detection of crossover forms of purely endogenous origin is much more complicated: it is easier for a parent to believe that the child speaks agrammatical and meaningless gibberish than to admit that the toddler communicates meanings to which he (the parent) does not have access anymore. For this reason we had restricted, with exception of MAUEN, this introduction to purely surface crossovers between morphophonologic and syntactic P-schemas. To go deeper would be too speculative.

This being said, we conclude this introduction by a remark which states that hearing or seeing a child producing a monolingual crossover is, verily, a reveatory event: it is as if, for a brief moment, one had indirectly regained access to the realm of long-forgotten knowledge.

END MONOLINGUAL CROSSOVERS 12.11.3

On preceding pages, we had used the term "crossover" to denote operators acting within a cognitive system, which are able to yield a new child scheme by means of mixing of multiple parent schemes. It was tacitly indicated that existence of many phenomena, including linguistic calcs, creol languages or overgeneralizations could be explained in terms of activity of such operators in human brain. Following table recapitulates the basic distinctions

	Multilingual	**Monolingual**
Intralexical	PIJE + TRINKEN =PIJEN	BAJA + ANÁN = BANÁN
Interlexical	??? (difficult to assess)	BAUEN + MAHLEN = MAUEN
Intraphrastic	VO ISt OKO? (calques etc.)	VO ISt AUGEN? (overgeneralizations)

Table 15: Recapitulation of crossover types observed in IM's production.

Due to abstract nature of "operator" entities it aims to organize, can be this taxonomy rightfully criticized as both crude and arbitrary. Thus, for example, a distinction between *multi*lingual and *mono*lingual could be considered as arbitrary by anyone asserting that the child is exposed to a multilingual linguistic environment (composed of, for example, motherese, fatherese, teacherese etc.) even in case when all members of social environment speak the same dialect.

The distinction between intraphrastic and intralexical could be also attacked on the sole ground that in many morphosyntactically rich languages, the very nature or even existence of distinction between notion of lexeme and phraseme is not as straightforward as it may seem.

But be it as it may, such theoretical hassles are of little use for phenomenological objectives of this chapter. By aiming to stay as faithful as possible to our initial method of *describing but non-categorizing* we cast aside this taxonomy as secondary and precise, that all above-introduced înt-[21] terms were introduced and categorized not because we would be 100% sure that such operators indeed materially operate within the human brain, but because we hope that their introduction could potentially stimulate or even facilitate further discussions.

One such discussion, concerning the assessement of crossover-like phenomena in CHILDES corpus shall be soon introduced.

END CROSSOVERS 12.11

12.12 OTHER PHENOMENA

Many unexpected and surprising events occur during such a complex and years-lasting process as language development definitely is. But since many amonth these phenomena are already in exhaustively described in litterature, let's just briefly describe two observations which were in certain sense "salient":

12.12.1 MULTILINGUAL C-SCHEME MISMATCH

The journal log entry (f2;5;18) describes an interesting dialogue which happened one afternoon after the mother picked up IM in kindergarten:

```
FAT: ako bolo v (sk: "how was it in") Kite ? (sk. locative of
    german word meaning kindergarten)
IM: MAMA ABHOLEN (de: "mother pick up")
FAT: ako bolo v Kite?
IM: MAMA ABHOLEN
FAT: ako bolo v Kite?
IM: MAMA ABHOLEN
```

On first sight it is somewhat difficult to see why IM had responded, three times in a row, with an answer "mama picked me up" to a question "how was it in kindergarten"?. The thing however can get more

21 Consistently with the syntax of PCRE we shall use the symbol ˆto denote the initial position. An expression înt hence matches all expressions prefixed by the trigram int.

lucid when one realizes that a sequence "ako bolo" and sequence "abholen" have certain morphophonologic features in common. In other terms, both can be fact matched by a a following C-scheme[22]:

```
a.*?bh?olo?
```

Given that it is evident that notions associated to the event of "being picked up from kindergarten" are, within child's mind, definitely more important than smalltalk questions about past events; and given that the father question was terminated with the term "Kita" which was practically the only attribute of the term "abholen" to which IM was exposed on a frequent basis, IM's thrice repeated answer was neither non-sense nor surprising.

On the contrary, it was a meaningful and true answer to the question which her C-schemes processed as question meaning something like "who picked You up from kindergarten?". Hence, not the term "slip of the tongue" but rather "slip of the ear" could be used to describe such phenomenon.

We consider this case of multilingual perceptive parapraxis to be of particulary interest because it can potentially result in a method, or even set of experiments, allowing to elucidate the problem of development of C-schemas which is, contrary to development of directly observable P-schemas, quite difficult to empirically measure and assess.

END C-SCHEME MISMATCH 12.12

12.12.2 COMPRESSION OF INFORMATION

Another interesting phenomenon was observed at (f2;6;18) at onset of period of increasing phrasal productivity. During a trip through the forrest with another family which has following members: H = father, M = mother, J = older son, T = younger son; IM enumerated a list of people who should go home in a following manner:

ALLE NACH HAUZE (de: "all home")
AUCH T, M AUCH (de: "also T, M also")
PAPA AUCH J (de: "father too J")

What is striking is the last sentence which was uttered after few seconds of silence during which apparently tried to remember the name of J's father (i.e. H). Since she could not remember it (or avoided its pronounciation), she had ultimately found her way out by producing

22 The C-scheme is a valid Perl regular expression which matches both strings "ako bolo" and "abholen". In such regexps, symbol "." matches any possible symbol, symbol "*" means "match zero or more occurences of preceding element" and symbol "?" means "match zero or one occurence of preceding element".

PAPA AUCH J which, in correct german would have to be a 6-word "auch J und auch sein papa".

But not caring much about correct rules of grammar which would oblige her to articulate sentences twice as long as necessary, IM had expressed the same communicative intention with three words only. Hence, at least in this case, "optimizing" forces inviting her to express her intention with as little resources as possible were definitely stronger than socializing and normative forces obliging her to produce only grammatically correct constructions.

END COMPRESSION OF INFORMATION 12.12

In a following manner could we continue and discuss one entry of the observation log after another. For example, we could discuss not only IM's pre-sleep monologues, but also mention productions which she used to cry out of her sleep, or uttered immediately after waking up.

We could focus on one meaning and describe development of labels which IM used to denote it. Or, as was the case when discussing the DING-DONG mystery, we could focus on one label and describe development of its meanings. Or we could list IM's first adjectives, questions and sylogisms. Or publish digital versions of the observation log as well as all other recorded materials.

But given the momentanous lack of IM's conscious and reflected consent to publication of her personal data, we think it is now time to conclude this chapter dedicated to development of this particular child.

OTHER PHENOMENA END 12.12

f(2;4;6)
ICH HABE AJN HUND
HABE AJN HUND
HABE AJN HUNDI
HABE AJN HUND
LA LA LA
DA BAUEN
DA BAUEN JA
DA BAUEN TATO

END QUALITATIVE 12

13 QUANTITATIVE

13.1 METHOD

The method of previous chapter mainly consisted in observations and interpretations thematizing structures produced by one individual toddler. But knowing that science should always aim to unveil not only the individual and specific, but also and especially the universal and generic, a hard-core empiricist could rightfully reproach us that was presented until now was maybe cute, but it was not science.

Thus, in order to correct and complement the methodological gap, all the effort presented in this chapter shall be subordinated to two ultimate *virtues* of the cartesian method. They are, of course:

1. reproducibility
2. quantification

Reproducibility is to be attained by exact specification of the input data and by publication of computational machinery which transforms the data into information or even knowledge. More concretely, every analysis shall include a list of corpora which were analysed as well as the bash|PERL|R code which performed the analysis. Thus, instead of using traditional logicomathematical formalisms, other formalism - less theoretical and more practical one - shall be used: that of PERL and its regular expressons (regexps).

When it comes to quantification, it shall be exactly the use of regexps which shall allow us to transform texts into numbers. By using regexps which are, in their essence, nothing else than *strings of characters able to match sets of strings of characters*, it should be potentially possible to identify, detect and mesure frequencies of occurence of quite abstract patterns or schemas.

Summa summarum, the method of this chapter shall mix little bit of data-mining with little bit of statistics and information extraction in order to attain the goal commonly known as "knowledge extraction".

END METHOD 13.1

13.2 DATA

« Child Language Data Exchange System (CHILDES)» (MacWhinney, 2014) is undoubtedly the biggest publicly accessible collection of both recordings of child speech as well as their transcripts. Since its foundation in 1984 by Brian MacWhinney and Catherine Snow, CHILDES had attracted interest of thousands researchers from all over the world and thus became the most important dataset for the nascent DP discipline.

Given its open yet standardized design, CHILDES contains hundreds of megabytes of transcripts representing children verbal productions and interactions in more than two dozen world languages. What's more, some of these transcripts include morphosyntactic annotations and/or audiovisual recordings which allow a more thorough contextualization of otherwise pure-text transcripts.

Note, however, that not all transcripts downloaded from the site of CHILDES project[1] shall be analysed. Primo, both directory "Frogs" as well as "PhonBank-Phon" are to be removed from the workbench since they do not contain .CHA transcripts made "in vivo". Secundo, all transcripts of children whose age is higher than the upper-bound of toddlerese (i.e. >30 months) are also excluded from analysis. This can be done by running the agesort.pl[2] script whose main functionality, however, is to devide the transcripts into two datasets:

1. PROTOTODD - the "prototoddlerese" dataset contains transcripts of children not older than 16 months
2. TODDLER - the "toddlerese" dataset contains transcripts of children between 16-30 months

Transcripts contained in both datasets thus obtained follow the .CHA format which stipulates that:

1. lines with child-originated speech are marked with token *CHI
2. lines with mother-originated speech (motherese) are marked with token *MOT
3. lines with father-originated speech (fatherese) are marked with token *FAT

and in all tables which will follow, we shall apply the same CHI | MOT | FAT notation to denote child, resp. mother or father. Distribution of different line types is presented on Table 16

	CHI	MOT	FAT
PROTOTODD	224855[3]	320454	13974
TODDLER	1453931	893357	154964

Table 16: Activity of different speakers in two age groups.

Every line of the .CHA file roughly represents a distinct and unique utterance. Thus, Table 16 suggests first distinction between two age

1 $wget -r –no-parent http://childes.psy.cmu.edu/data/
2 http://wizzion.com/thesis/code/childes/agesort.pl

groups: in the protoddler period mothers in general produced 42% more utterances than children, the ratio was more than inversed in the later group[4] . In comparison to mothers, fathers seem to serve only marginal role within both datasets, their presence, however, seems to be significantly higher in case of the older group ($FAT_{PROT}/MOT_{PROT} \approx$ 4.3%, $FAT_{TODD}/MOT_{TODD} = 17.3\%$). END DATA 13.2

13.3 UNIVERSALS

This section offers analysis of CHILDES transcripts coming from different languages. Table 18 shows number of distinct .CHA files (transcript) which are to be analysed as well as languages in which they were spoken.

	ara	deu	eng	fra	jpn	rum	spa	tha	biling	other
PROTOTODD[5]	54	176	1026	152	42	53	56	142	31	107
TODDLER	25	591	2505	410	235	42	140	46	801	1063

Table 17: Repartition of languages in studied corpus.

It is thus evident that all in all, CHILDES is strongly biased towards indo-european languages in general and English in particular. This bias notwithstanding we shall, in following analyses, *throw all data into one bag* as if we were studying one sole language.

13.3.1 LETTERS

Let's now run the script [6] performing the most simple analysis possible: i.e. the measurement of frequencies of occurence of diverse graphemes (i.e. letters) within utterances produced by children and their parents. This yields distributions presented on Table 17.

It can be seen that no matter the speaker and no matter the age group, vowels A, E, and O are always among four most frequent entities. But closer inspection of the data can lead to discovery of certain interesting developmental phenomena occuring also between the groups whose contrast interest us the most, that is CHI_{PROTO} and CHI_{TODD}. It can thus be seen that the utterances of children younger not older than 15 months are dominated by occlusive consonants (H, M, T, D, N, P) and other types of consonants like fricatives (S), trills (R) or laterals (L) attain more dominant positions only in later period.

A particularly instructive case seems to be the decrease in ranking of the labionasal occlusive M. While this consonant is the 5th most

4 The ratio $1453931/893357 \approx 1.614$ is quite close to number $\phi \approx 1.618$ better known as "golden ratio" or "golden section". Sapienti sat.
6 http://wizzion.com/thesis/code/freq_1gram.pl

Table 18: 20 most frequent graphemes according to speakers and age groups.

CHI				MOT				FAT			
PROTO		TODDL		PROTO		TODDL		PROTO		TODDL	
a	32187	a	108278	e	400499	e	1448021	a	21516	a	208869
e	21151	e	103443	o	371032	a	1220454	o	12563	e	195921
o	19400	o	700249	a	344629	t	1101875	n	10394	o	155729
h	17569	n	665280	t	319301	o	1069481	e	10258	t	132600
m	12472	i	654456	h	252510	i	865696	i	9571	n	127344
t	11557	t	611047	n	233620	n	853893	u	8399	i	126082
d	10969	h	482184	i	228587	h	801306	t	8063	s	107667
u	10949	s	438493	s	194898	s	755952	h	7273	h	87652
n	10668	r	383941	u	173597	r	607236	r	5230	r	85586
i	10068	m	364861	r	160583	u	532238	k	5191	u	85416
b	8768	u	338297	y	142311	l	501064	s	4995	l	68172
p	7200	d	323376	l	137550	d	450710	m	4689	d	64320
y	6754	l	310237	m	109696	m	351217	j	4071	m	52445
l	6704	c	234078	d	106698	y	345154	d	3876	y	44005
r	6501	k	226683	w	94381	c	290880	p	3683	c	41410
c	6163	p	192730	g	85473	g	282157	l	3555	k	39026
s	5589	g	189097	k	77850	w	280615	w	2825	g	35873
â	5328	y	186999	c	76826	k	234218	y	2436	p	34373
g	5233	b	184783	p	72531	p	216647	c	2379	w	30227
k	4581	w	138086	b	67397	b	191091	b	1632	b	26719
w	3857	j	107099	f	33203	f	123552	g	1581	v	18730

frequent in the transcripts of younger children and is 2.58 times less frequent than the most frequent A, in case of older toddlers M is only 10th more important and 3.36 less frequent than A. Given that all four FAT and MOT distributions consistently place M at rank 12 or 13, the phenomenon of "decrease of importance of M" - and in lesser extent also of P and B - can be potentially explained in terms of divergence from certain potentially innate labiotactic schemata (c.f. 12.5.1) and gradual convergence towards more socially determined articulations.

We leave to readers's ingenuity detection and discussion of other phenomena presented by the table, including mother's preference for the vowel E and father's and children's preference for the vowel A.

END LETTERS 13.3.1

13.3.2 N-GRAMS

Let's now focus on distributions of N-grams, that is, the sequences of N letters. Since we have already presented the distribution of letters which is equivalent to distribution of 1-grams, Table 19 presents the distribution of 2-grams (bigrams) as assessed in 7697 transcripts which our script[7] had analysed.

Table 19: 20 most frequent bigrams according to speakers and age groups.

CHI				MOT				FAT			
PROTO		TODDL		PROTO		TODDL		PROTO		TODDL	
'a_'	6168	'e_'	341873	'e_'	167116	'e_'	562814	'a_'	6085	'e_'	72319
'e_'	5487	'a_'	273927	't_'	108290	't_'	383994	'n_'	4173	'a_'	53852
'^ a'	4688	'n_'	191424	'_t'	92762	's_'	335865	'e_'	3722	's_'	42541
'^ b'	4089	't_'	173103	's_'	89654	'_t'	311683	'aa'	3242	't_'	39194
'^ d'	4072	'o_'	157271	'th'	89035	'th'	266289	'oo'	2663	'n_'	36965
'^ m'	3744	's_'	153264	'he'	76868	'he'	246986	'j_'	2660	'_t'	34962
'y_'	3699	'_t'	123289	'ou'	73519	'n_'	243248	'i_'	2642	'o_'	34081
'h_'	3660	'er'	120991	'n_'	66505	'a_'	220702	'o_'	2448	'an'	25338
'ma'	3649	'an'	118561	're'	57772	'ha'	198677	'_t'	2233	' d'	24430
'ah'	3451	'i_'	116078	'ha'	57707	'ou'	186602	'_n'	2194	' a'	24018
'da'	3085	'he'	113844	'a_'	57042	'o_'	183667	'_m'	2022	'er'	23690
'n_'	3072	'in'	111541	'yo'	55805	'_d'	182416	'aj'	1990	' s'	23146
'oo'	3024	'th'	102953	'y_'	53855	'er'	180576	'an'	1981	'i_'	22763
'ba'	2716	'_a'	96622	'an'	51556	' a'	174639	'uu'	1933	'th'	22579
'h_'	263	'_d'	94794	'u_'	50789	'an'	170761	't_'	1880	'he'	22503
'an'	2361	'ch'	94477	'er'	49019	're'	170253	'_k'	1866	'r_'	22353
'o_'	2357	'_m'	94073	'at'	48593	'in'	167113	'u_'	1813	'ha'	21978
'^h'	2289	'h_'	93204	'o_'	46667	'at'	160793	'ha'	1758	'u_'	21764
'de'	2254	'ma'	92728	'_a'	44251	' i'	159637	' p'	1712	'ou'	20780
'^p'	2219	'en'	92239	'_y'	43725	'r_'	158742	'ii'	1574	're'	20519
't_'	2188	'ha'	91373	'on'	41773	'_s'	156205	'th'	1574	'en'	19847
'at'	2056	'r_'	89388	'_s'	41663	'u_'	147883	'na'	1555	'on'	17843

In our notation, symbol ^ means "beginning of utterance" and symbol _denotes the pause between the words (normally denoted by a simple blank space) and is understood as a symbol in its own right. In general, vowels A and E at the ultimate word position tend to dominate the lists but in case of the group which interests us most, i.e. $\text{CHI}_{\text{PROTO}}$ they are followed by a group of bigrams denoting either vowel A or occlusives B, D, and M (and somewhat later also H and P(occuring at the initial position of whole utterance.

7 http://wizzion.com/thesis/code/freq_2gram.pl

It is also worth noting that for this group, the most frequent bigrams having the consonent-vowel (CV) syllabic form are MA, DA and BA and bigrams following the VC form are AH, AN and AT. We consider these findings as consistent with both data commonly reported in DP litterature, as well as with qualitative observations of IM's first protowords (c.f. Table 10 like MAMA, DADA or BABA.

As usually, we set aside other potentially interesting questions like "is the predominance of long vowels AA, OO, UU, II in prototoddler-directed fatherese a sheer artefact of the corpus[8] or do these results point to somewhat more profound a phenomenon ?" and point hereby the reader's attention to Table 20 outputs of the scripts assessing the frequencies of 3-grams.

Table 20: 10 most frequent trigrams according to speakers and age groups.

CHI				MOT				FAT			
PROTO		TODDL		PROTO		TODDL		PROTO		TODDL	
'^ba'	1692	'_th'	57081	'_th'	58738	'the'	140513	'aa_'	1994	'_th'	18736
'^ma'	1629	'er_'	51736	'you'	54163	'you'	127201	'aj_'	1990	'the'	14051
'mam'	1619	'en_'	51386	'the'	41636	'hat'	112441	'_th'	1151	'you'	12720
'ah_'	1448	'the'	51022	'ou_'	40205	're_'	104532	'ii_'	1148	'hat'	11244
'^da'	1428	're_'	48700	'_yo'	38424	'_yo'	100048	'an_'	1074	're_'	10453
'ama'	1323	'^ja'	41974	're_'	35344	'he_'	98865	'oo_'	1016	'_yo'	10004
'det'	1145	'in_'	40925	'hat'	34041	'ou_'	97958	'on_'	1016	'he_'	98865
'dad'	1118	'no_'	40168	'at_'	30480	'at_'	96754	'_ma'	980	'ou_'	97958
'aa_'	1030	'^no'	39940	'he_'	28589	'is_'	76056	'_na'	823	'at_'	96754
'et.'	998	'her'	39541	'her'	28009	'her'	74624	'aw_'	819	'is_'	76056
'^ah'	971	'ne_'	38460	'ere'	25958	'er_'	74623	're_'	805	'her'	74624

In general it can be stated that the trigram-related phenomena seem to extend quite naturally the phenomena which were already observed and discussed in relation to bigrams. Word onset syllables BA, MA and DA thus dominate the list of prototoddlerese. But since these trigrams are not fully qualified (they contain the meta-character ^), it can be stated that the most frequent trigrams with equally trigramic phonemic correlates are MAM, AMA, DET and DAD.

In the later period, i.e. in $\text{CHI}_{\text{TODDL}}$ transcripts one can observe a bias towards distribution of standard english marked, of course, by the dominant position of the graphemic trigram (and phonemic bigram) denoting the most frequent word of english language, the determiner THE. This bias notwithstanding, word onset syllables ^JA

8 These long vowel sequences seem to originate, in great extent, from transcripts of japanese and tamil fatherese.

and ^NO appear at highest positions of the list for a reason which we can briefly elucidate only in the footnote[9].

Leaving again the question of fatherese aside as the problem of its own, let's now look at motherese. In general, both distributions indicate that the corpus was strongly biased towards English. Thus, the obligatory THE is present (as well as its fragment _TH preceded by the pause), as well as trigrams HAT and ERE owing their high ranks to highly frequent words like what/that resp. where/there/here, within which they occur.

What is striking is, however, the position of the trigram YOU. While in frequency lists generated from "standard English" corpora [10], the word You is 17th most frequent and occurs ≈ 9.3 less often than the most frequent word THE, in the speech directed to younger infants the trigram it is the trigram You which dominates the list of fully qualified trigrams, occuring 1.3 more often than THE !

Among all phenomena observed until now, we consider mother's tendency to say You, to be the most salient example of what we consider to be *the very essence of motherese*.

END N-GRAMS 13.3.2

13.3.3 INTRASUBJECTIVE REPLICATIONS

It has been already *repeated* on multiple places (5,12.6) that interpretation of "repetition of information" as a sort of "replication of information" is one among main tenets of the theory hereby presented. Thus, let's now try to assess the extent in which children repeat their own productions.

Intralocutory duplications

Intrasubjective duplications can be detected by searching for repetition of a sub-string X within the envelopping utterance-string U. If X is a bigram this can be easily done by matching the utterance with the regexp:

```
$U =~ /(.2)\1/g
```

9 Execution of grep -P "CHI:\tno" ./toddler/* indicates that within the corpus of later toddlerese, high frequency of ^no is principially caused by augmentation of child's egocentric tendency to answer questions in negative. In case of ^ja execution of the command grep -P "CHI:\tja" ./toddler/* indicates that the situation seems to complicated by the fact that the grapheme J denotes different phonemes in different languages (compare "jagen" in german and "Jack" in English or "jagami" in sanskrit). This complication notwithstanding, it seems that high frequency of ^ja can be in non-negligeable extent explained in terms of baltoslavogermanic "yes". Thus, for example, the sole 11312/c-00023045-1 transcript shows how small german boy Leo answered 104 times with the word JA

10 https://en.wiktionary.org/wiki/Wiktionary:Frequency _lists/PG/2006/04/1-10000

and for any duplicated sub-string at least two characters long, the matching pattern is

$$\$U =\sim /(.2,)\backslash 1/g \quad (3)$$

Note that these patterns match only adjacent repetitions, id est such cases when two instances of the repeated substring are juxtaposed side by side.

The script[11] confrontating the second (i.e. $length(X) \geqslant 2$) with child-produced utterances yields outputs presented in Table 21.

Table 21: Duplicated expressions and numbers of child-originated and child-directed utterances in which they occur.

CHI				MOT			
PROTO		TODDL		PROTO		TODDL	
ma	1117	ma	11756	ma	1266	ma	2633
pa	294	pa	4545	is_	733	is_	2408
da	290	ko	2970	bye	696	em	1904
bye	142	is_	1651	no_	547	mm	1338
an	140	ba	1412	it_	532	pa	1283
ba	136	la	908	na	523	it_	1177
ta	76	an	764	da	443	e_	963
na	75	da	730	mm	382	na	820
ah	65	no_	647	ba	336	a_	700
woof_	64	ta	616	an	254	an	692
uh	60	na	600	em	197	ba	505
open_	59	do	588	boo	177	to_	468
mommy_	57	be	580	ing	167	ko	435
cou	53	bye	552	uh	160	no_	384
vov	48	e_	536	nyan	158	in	374
mm	45	bo	535	man	157	ing	349
ga	41	pi	468	ha	147	bye	328
he	40	ca	430	pa	143	er_	319
no_	40	in	387	nai	135	cher	311
book_	39	cha	372	to_	134	li	293
ha	39	ka	344	cou	132	_we	290

Postponing the discussion[12] of specificites of the data hereby presented to the later date let's focus on scientifically more pertinent a fact: the overall statistics of duplications. This is shown on Table 24 whose values were calculated by normalization by means of a formula

$$P(duplication) = ALL_{matching}/ALL_{utterances}$$

11 http://wizzion.com/thesis/code/isipr.pl
12 "Mothers do not say woofwoofwoof as babies do, mothers say manmanman".

whereby $ALL_{matching}$ denotes the number of utterances produced by CHI (resp. MOT) matchable by regexp presented in Formula 3 and $ALL_{utterances}$ denotes the number of all utterances uttered by the person.

	CHI	MOT
PROTO	0.041	0.086
TODDL	0.066	0.058

Table 22: Probability that the utterance shall contain at least one ajdacently duplicated 2+gram.

This table indicates that the intralocutory duplication is to be most probably observed in motherese directed to younger children. Younger children, on the contrary, tend to produce less adjacently duplicated sequences[13]. In the later period, however, they tend to replicate, within one utterance, the fragments of their production more frequently than their mothers.

END INTRALOCUTORY REPLICATIONS 13.3.3.0

Translocutory replications

Let's now focus on reproduction not to be observed within one individual utterance, but between two adjacent utterances. Given the speaker S who utters U_1 before uttering U_2, one can look for replication of patterns between U_1 and U_2 by simply

1. creating a new datastructure, a "couplet" which concatenates two utterances and the divisor symbol #, i.e.

$$couplet = concatenate(U_1, \#, U_2)$$

2. matching the couplet with regex like

$$\$couplet =\sim /(.3,). * \#. * \backslash 1/g \qquad (4)$$

and this is exactly what is being done by a 3rd line of the[14] script whose outputs are in part presented in Table 23.

Note that in contrast to Formula 3, the regexp in Formula 4 contains expresion {3,} and not {2,}. This means that in this analysis, we have been looking for repeated strings of three or more characters (3+grams). This design choice was done in order not to pollute the results with repeated bigrams among which many (e.g. "th", "ch") represent in many languages just a sole phenome, and their repetition is thus highly probable. Other design choices are, of course, possible.

13 Or transcribers do not transcribe them as such.
14 http://wizzion.com/thesis/code/isitd.pl

Table 23: Most frequent translocutory 3+-grams.

CHI				MOT			
PROTO		TODDL		PROTO		TODDL	
det.	332	ja.	8160	you	8698	you	19229
kore.	223	no.	4461	the	3809	the	15302
maman.	210	the	3328	that	1554	what	557
mama.	182	da.	3071	here	1531	that	441
eh.	174	yeah.	2870	what	908	here	300
baby.	162	ein	1979	and	776	ing	2423
ball.	126	that	1670	t's	627	and	2088
no.	121	aa.	1655	look	550	t's	2056
daddy.	104	nein.	1536	ing	520	das	1612
aa.	102	en.	1535	there	393	there	1437
papa.	89	there.	1363	that's	382	she	1399
mommy.	84	here	1190	her	366	her	1323
ooh	81	das	1099	your	344	that's	1
up.	76	die	1088	where	338	tha	1179
dah	75	der	1006	come	325	ein	1093
ah.	73	this	973	this	323	ich	1091
da.	73	and	883	see	300	est	1069
dog.	67	you	830	one	272	der	1030
dada.	62	yes.	772	yeah	261	one	1020
uhoh.	61	ich	698	no.	249	n't	1005

Some interesting phenomena pop up here. In general the table is in general populated by deictic pronouns[15], determinants, answer particles and various form of positional adverbs. In motherese, an injunction to action appears from time to time in the form of a verb ("look", "come", "see"). And, of course, it is very probable that if the current motherese utterance contains the word "you", the next one utterance shall contain it as well.

The presence of motherese expressions like "t's" and "n't" also suggests the occurence of first variation sets (that's vs. it's, isn't vs. don't etc.)

What's more, one can see quite clearly a distinction between language of younger and older children. While the distribution of translocutory duplications of older children is quite similar to motherese[16] this is in no way the case for younger children. Repetition of "abstract"

15 DET is the danish deictic meaning "that" and KORE is japanese deictic meaning "this". Transcripts of danish children Anne (e.g. 11312/c-00021705-1), Jens (e.g. 11312/c-00021750-1) and japanese girl Hiromi (e.g. 11312/c-00009753-1) seem to be in great extent "responsible" for high scores of these words.

16 The most salient exception to this being the tendency to repeatedly utter "ja." or "no.".

deitics is quite rare and seems to be limited to few particular children like Jens and Hiromi. On the other hand, the list of repeated tokens is dominated by words denoting concrete persons ("maman", "mama", "baby", "ball", "daddy", "papa", "mommy", "dog") and particles with undefined content ("eh", "aa", "ooh", "ah", "uhoh") potentially referring to emotional states. Even the adverb/preposition "up" is present, sometimes probably serving the function of injunction "raise me up!" or "look up!".

This being said, let's now look at overall statistical properties of distributions thus obtained:

	CHI	MOT
PROTO	0.08	0.37
TODDL	0.28	0.38

Table 24: Probability that both parts of a utterance couplet shall contain at least one identic 3+gram.

An significant increase in amount of translocutory replications is observed when one compares data of younger and older children. This is consistent with what was observed in case of intralocutory duplications Table 24 but here, the phenomenon is even more marked. Motherese, on the contrary, seems to keep a property of *repeating a 3+gram slightly more often than once in three utterance couplets*.

END TRANSLOCUTORY REPLICATIONS 13.3.3.0

Many minor phenomena asides, preceding subsubsections have briefly shown:

1. a fast[17] and frugal[18] regexp-based method of extraction of repetitive patterns from huge corpora
2. that language of children younger than 15months contains less intralocutory resp. translocutory replications of 2+ resp. 3+gram sequences than language of older toddlers

This being said, let's now focus on replication of structures which is to be observed not in and/or between utterances produced by one speaker but in utterances produced my multiple speakers.

END INTRASUBJECTIVE REPLICATIONS 13.3.3

17 All presented analyses were performed in less than a minute on one single 2.5GHz core.

18 All scripts are shorter than 42 lines of pure PERL, including loading the corpora, cleaning it from metadata and most salient noise, parsing and printing the result.

13.3.4 INTERSUBJECTIVE REPLICATIONS

Intersubjective replication is equivalent to imitation. It is observed if and only if two distinct subjects produce the same construction in a very limited timespan. To make things simple, this section shall be concerned only with detection of the most trivial intersubjective replications: those which immediately follow each other.

PROTO				TODDL			
CHI_{INIT}		MOT_{INIT}		CHI_{INIT}		MOT_{INIT}	
ball	74	ball	42	the	2038	the	3058
baby	68	baby	40	that	1534	here	1731
daddy	50	here	33	here	1045	that	1175
up.	47	det	26	you	969	you	997
guh	40	apple	21	no.	895	ing	993
det	36	byebye.	21	what	764	what	539
dada	33	the	20	yeah.	528	one	502
more	33	daddy	20	ing	492	there	447
that	30	that	19	there	466	ein	436
byebye.	29	down	19	das	463	and	361
book	29	open	19	ein	444	t's	333
hi.	28	mommy	18	want	439	das	314
water	25	hi.	17	one	348	ich	285
car	25	dada	17	ja.	346	det	261
down.	24	book	16	t's	342	der	240
block	23	block	16	and	328	want	234
open	23	big	16	non	327	this	230
mama.	23	guh	15	yeah	326	que	216
bottle	22	you	15	daddy	324	est	209
big	20	water	14	hat's	303	can	203
non	20	car	14	det	303	die	184
uhoh.	19	boo	14	baby	275	c'est	161
no.	18	dad	13	oh.	274	see	158
agu	18	bye.	13	her	262	it's	155
backpack	17	okay	12	where	245	oh.	153
duck	17	bye	12	there.	230	den	149
doggie.	17	and	12	mhm.	226	they	144
apple	17	uhoh.	12	ich	219	baby	138
here	16	duck	12	that's	217	that's	138
dirty	16	sticky	11	can	213	car	138

Table 25: Most frequent words replicated from child to mother (CHI_{INIT}) and mother to child (MOT_{INIT}).

Technically, the extraction is performed by means very similar to those which extract intrasubjective translocutory replications (c.f. previous section). The only difference being, of course, the origin of UTT_1

and UTT_2. While in detection of intrasubjective replications UTT_1 and UTT_2 are uttered by the same person, in case of intersubjective replications it cannot be so and additional condition $speaker(UTT_1) \neq speaker(UTT_2)$ has to be implemented in the code.

Another thing which is to be carefully considered is identity of the person which initiated the replication (i.e. uttered UTT_1) in contrast to identity of the person which reacted (i.e. uttered UTT_2). On following pages these shall be distinguished by attributes INIT resp. REACT.

Listings generated by the[19] script implementing such considerations have been listed on table Table 25. They indicate, among other things, that

- entities intersubjectively replicated and shared between mothers and younger toddlers tend to denote concrete physical referents ("ball", "baby", "book", "water", "car", "mama", "bottle", "backpack", "doggie", "apple", "block"), their properties ("big", "dirty", "sticky") or directions along vertical axis ("down", "up")
- entities intersubjectively replicated and shared between mothers and older toddlers tend to encode more abstract linguistic entities (deictic pronouns, locative adverbs) as well as basic syntactic constructions ("that's", "c'est", "it's")
- children initiate exchanges about different "topics" than mothers do[20]

Overall statistic properties assessed by the script are presented on Table 26. These are: number of couplets in which child utterance preceeds or follows the mother utterance (N_C); number of couplets which have at least one 3+gram in common (N_R) and probability that a MOT-CHI or CHI-MOT couplet will have at least one 3+gram in common ($P_{R|C} = N_R/N_C$).

	CHI_{INIT}	CHI_{REACT}
PROTO	$N_C = 46795$ $N_R = 6005$ $P_{R\|C} = 0.128$	$N_C = 46923$ $N_R = 4167$ $P_{R\|C} = 0.088$
TODDL	$N_C = 378958$ $N_R = 130713$ $P_{R\|C} = 0.344$	$N_C = 378712$ $N_R = 92340$ $P_{R\|C} = 0.243$

Table 26: Basic statistics concerning the replication of 3+grams between mother and the child.

19 http://wizzion.com/thesis/code/tsr.pl

20 For example in younger toddler group had mothers repeated 33 times the word "more" uttered by their child. But only in 9 cases was the word "more" uttered by a mother repeated by her child. Or, in older group, mothers have reproduced "no." of their children in 895 cases; toddlers repeated "no." of their mothers only in 133 cases.

It may be thus seen that in both groups, toddlers initiate more intersubjective replications than they react to. Or, in other terms, that mothers tend to prefer reacting to topic changes than changing the topic themselves. It is as if mothers, not children, were adapting themselves to the currently addressed topic.

But it can be also seen that this asymmetry is less marked in exchange with older toddlers. For while mothers reproduce at least one trigram after approximately every third utterannce their children repeat at least one trigram approximately after every fourth utterance. In younger group this is not so: child reproduces the fragment of mother's talk only once in every 12th utterances and mother do so only once in 8 utterances.

These distinctions notwithstanding, we consider it worth mentioning that there seems to be, in fact, one thing common to both age groups: the ratio between probabilities. That is, Table 26 indicates that, statistically speaking, it is ≈ 1.4[21] times more probable that the replication-containing couplet was initiated by the child and not by the mother.

END INTERSUBJECTIVE REPLICATIONS 13.3.4

Thus ends our brief excursion through the realm of linguistic universalia. It could undoubtably continue, for example by following the direction indicated by Table 27:

	CHI	MOT
PROTO	1148[22]	3329[23]
TODDL	2454 [24]	1319 [25]

Table 27: Distributions of occurences of marker for laughing in diverse subsets of CHILDES corpus.

and a lot of ink could be spilled by tentatives trying offer a serious, scientific, fully cartesian and p-value-endowed answer to question "*how is it possible that the CHA format's marker **laugh** is 2.5 times more frequent in transcripts of prototoddlerese when it contains 4013 less transcripts than the corpus of toddlerese?*".

But given the importance, intensity, diversity and perennial actuality of the topic (Aristotle, 5 BC), the role of laughing in development of mind can not to be addressed in the current pamphlet in extent it merits.

21 $0.128/0.088 = 1.45; 0.344/0.243 = 1.42$
22 $ grep laugh ./prototoddler/* | grep CHI | wc -l
23 $ grep laugh ./prototoddler/* | grep MOT | wc -l
24 $ grep laugh ./toddler/* | grep CHI | wc -l
25 $ grep laugh ./toddler/* | grep MOT | wc -l

Instead of doing so let's now fully admit that in case of analyses of corpus so strongly biased towards english language as the one hereby studied, it's maybe wiser to stop babbling about "universals" [26] and rather start assessing, evaluating and interpreting the central tenets of our theory in "english-specific" terms.

END UNIVERSALS 13.3

13.4 ENGLISH-SPECIFIC

In this section we shall present results of few data-mining experiments which concerned only those parts of CHILDES corpora which:

1. which transcribe interaction between english-speaking adults and english-speaking children
2. which also contain morpohological and grammatical annotations (i.e. every utterance line is also followed by %mor line and %gra line)

Table 28 contains overall statistics[27] of the datasets fulfilling these conditions and obtained by running the script langsort.pl[28].

	PROTO (<16 months)	TODDL (<16 months >31 months)
Investigators	10	35
Subjects	86	288
Transcripts	330	1335
CHI utterances	42229	293751
MOT utterances	196781	370972
CHI words	132927	1035341
MOT words	1076028	1921131

Table 28: Counts related to morphologically annotated english-language transcripts analyzed in this section.

As can be seen, the corpus still contains a non-negligeable amount of data describing interactions from almost hundred younger toddler subjects and almost three hundred older toddler subjects. Given that the data were collected and transcribed by dozens of diverse investigators, it can be expected that certain knowledge about generic ten-

26 Or do so elsewhere, c.f. (Hromada, 2016e).
27 All values were obtained by means of a standard UNIX utility wc (e.g. the amount of letters in CHI utterances was obtained by executing the shell command *$grep -P '^CHI' ../toddl_english/*.cha |wc -c*. Note that for wc's operational definition of what "word" (i.e. a continuous sequence of characters separated from other words by blank spaces) means strongly overlaps, but is nonetheless not completely equivalent, to what "word" means in linguistics.
28 http://wizzion.com/thesis/code/langsort.pl

dencies could be attained if ever the data was to be processed in a stringently quantitative manner.

Instructions and definitions of CHILDES Manual (MacWhinney and Snow, 1991) should be also taken into account more strictly than was the case in the preceding "universals" section. Other details of text preprocessing are mentioned in annex (**??**).

13.4.1 UTTERANCE-LEVEL CONSTRUCTIONS

Table 29 contains top most frequent utterance-level constructions obtained by launching one simple command[29].

That communication of younger children is dominated by non-linguistic behaviours (vocalizations, babbling, crying, laughing etc.) is hardly surprising. Nor is much surprising that younger children tend to produce shorter utterances. Nor the fact that vaste majority of multi-word motherese utterances are short fixed expressions (e.g. "come on", "that's right", "oh dear", "good girl").

Observation of certain similiarities between distributions of child-directed speech and speech produced by older children can one lead to a hypothesis that these distributions correlate. In order to verify the hypothesis, a simple script was programmed[30] which merged two complete distributions into one table. Subsequently, Pearson correlation coefficients were calculated and are presented on table Table 30.

One may thus observe the existence of statistically significant (i.e. $p<0.05$) correlations in all cases except the one: no statistically significant correlation was observed between MOT_{TODDL} and $CHILD_{PROTO}$. These seems reasonable, for how could the language of a young toddler a priori correlate with language which the mother shall use when the child will be older? In reverse direction, however, a weak (cor=0.022) but nonetheless statistically significant correlation is observed: thus, there exists certain relation between distributions of utterances in language produced by older children and distribution of utterances in language heard by younger children.

The strongest correlation, however, is to be observed between MOT_{TODDL} and CHI_{TODDL} which can be potentially explained in terms of convergence of toddlerese towards "the golden standard" actualized by language of the mother.

This being said, let's now conclude this brief overview of utterance-level distributions with Table 31 which presents following quantities:

- N_d : the number of distinct utterances present in the corpus
- P_d : probability that the utterance is distinct (N_d normalized by number of all utterances (c.f. Table 28)

29 grep -h -P "CHI: [^o]" ./toddl_english/* | sort |uniq -c |sort -g -r
30 http://wizzion.com/thesis/code/correlator.pl

- H : Shannon entropy of utterance distribution, calculated[31] as $H = -\sum_i P(x_i) \log_2 P(x_i)$ (Shannon, 1948) where $P(x_i)$ denotes the probability of occurence of i-th utterance (e.g. its relative frequency of occurence)

Given that the Shannon entropy can be understood as a measure of uncertainity and unpredictability, it may be stated that production of younger children yields most predictable transcripts. Production of older children is much less predictable and every new utterance seems to bring about twice as much information content (≈ 13.7 shannons instead of 6.8) . And utterances produced by mothers are even less predictable.

END UTTERANCE-LEVEL CONSTRUCTIONS 13.4.1

13.4.2 PIVOT SCHEMAS

In item 9.4.4, a pivot schema was defined as a two-word schema in which one word ("the pivot") recurres frequently in the same position and the other word varies. In order to detect potential pivot words, let's define a sort of "pivoteness" score as:

$$score_{pivoteness} = F_{N-gram} * length(N - gram) = F_{N-gram} * N$$

which is to be calculated for every continuous N-gram which occurs in the corpus and has more than X characters (i.e. $N > X$) . For example, if the corpus have contained only 4 utterances containing only the expression "dogs" and one utterance containing the expressions "dog", and if the parameter X was set to 3, the score-attributing script[32] would attribute score $4 * 4 = 16$ to tetragram "dogs", score $15 = 5 * 3$ to trigrammaton" "dog" and score $12 = 4 * 3$ to 3gram "ogs". However, bigrams "do", "og", "gs" as well as unigrams "d", "o", "g", "s" would be ignored since parameter $X = 3$.

Table 32 lists top thirty 8+grams (i.e. X=7[33]) extracted from all CHILDES transcripts of english-speaking children not older than 2 years and 7 months[34]

As may be seen, more than half of most salient pivots are onset expressions initiating the utterance (marked by the starting symbol ^) and the rest is divided between expressions which end the utterance ("in there", "'s that?", "on there", etc.) or are in midst of it (" in the ", " on the ", "another", "little").

31 http://wizzion.com/thesis/code/entropycalc.pl
32 http://wizzion.com/thesis/code/exh.pl
33 Note that the choice of the parameter X was in great extent arbitrary and only in much lesser extent motivated by "magical number seven, plus or minus two", postulated by (Miller, 1956).
34 The complete list of all 8+gram expressions and their associated $pivoteness_7$ scores is available at http://wizzion.com/thesis/results/pivots_english_7.

It is, however, quite probable that even among these pivot candidates there would be some which are not true pivots because they occur only in restricted amount of contexts. But in case like ours when all contexts are known, such "false pivots" can be potentially identified by an algorithm which, for every pivot candidate C:

1. assesses the distribution of contexts[35] D_C
2. calculates the Shannon entropy of D_C

and this is, indeed, the procedure actualized by the script pivotentropy.pl[36] whose outputs[37]introduced on the Table 33.

As may be seen, results presented on Table 33 are quite similar to results already presented on Table 32. There exists, indeed, a statistically significant correlation between $score_{pivoteness}$ and $H_{contextual}$ (i.e. Spearman's non-parametric rank correlation test yields p-value < 2.2e-16, $\rho = 0.474$). Since evaluation of $score_{pivoteness}$ is less costly than that of $H_{contextual}$, and since entropy values are, so to say, more precise than the $score_{pivoteness}$, the fact that these two measures tend to correlate may turn out to be quite useful in applied NLP practice.

Summa summarum, constructions extracted from CHILDES by means of above-mentioned methods strongly ressemble Bruner's "formats" (5).

END PIVOT SCHEMAS 13.4.2

13.4.3 PIVOT INSTANCES

Let's now focus on pivot instances, that is, on expressions which are matched by pivot schemas. We define: **an utterance U instantiates the pivot schema P if and only if U can be matched by the P-representing pattern**. In case we choose PERL regexes as a means of representation of pivot schemes this definition can be formalized as

$$\$U =\sim /\$P/$$

whereby =~denotes the regex-matching operator.

This notion is implemented by the script pivot_utterance_global.pl[38] which, when initialized with list of pivots as its input data, returns

35 In what shall follow, the term "context" means "two words to the right" if pivot initiates the utterance, "two words to the left" if it terminates the utterance and "one word to the left and one word to the right" if it is in the midst of it.
36 http://wizzion.com/code/thesis/pivotentropy.pl
37 The list of 1000 $pivot_7$ schemas with highest pivoteness and their CHILDES $H_c ontextual$ entropies is downloadable at http://wizzion.com/thesis/results/pivot7_entropies_english.1000
38 http://wizzion.com/thesis/code/pivot_utterance_global.pl

the frequencies of utterances which instantiate one among ten pivots with such high informational content. Most frequent among such pivot-instantiating utterancess are listed on Table 34 along their respective frequencies of occurrence.

The list makes evident certain usage-oriented, egocentric (e.g. "I want it", "that's mine"), attention-sharing ("look at this one") tendencies potentially inherent to human toddlers. But in order to be sure that it is, indeed the case and not just an artefact of the method with whic we treat the corpus, let's now slightly readjust the methodology: let's NOT throw all data coming from all children to one bag which is subsequently analyzed, but instead, let's keep all utterances well associated to their locutors in order to *identify* such utterances which are being spoken by biggest number of individual locutors.

Operationalization of such a methodology into the PERL script[39] makes it possible to pose the following question :

Which pivot-instantiating utterance was uttered by the biggest number of distinct children ?

45 top-ranking utterances are listed on Table 35 as an answer[40].

As before, this more horizontal an analysis indicates that toddlerese tends to be dominated by level-0 constructions:

1. encoding deictic focusing of attention to some object
2. expressing wanting or egocentric posession
3. asking for more information

or level-1 crossovers of such level-0 schemas like, for example,

$$\frac{\begin{array}{r}(another\ one)\\ \times\,(I\ want)\end{array}}{(I\ want\ another\ one)} \qquad \text{Q.E.D.}$$

END PIVOT INSTANCES 13.4.3

13.4.4 PIVOT GRAMMARS

Three tables which follow aim to ellucidate more closely the concrete substance of some pivots with big $H_{contextual}$ [41]:

Before leaving, we remind the reader of the fact that all "grammars" presented hereby are "intersubjective" in a sense that they were extracted from corpus of transcripts produced by distinct children.

39 http://wizzion.com/thesis/code/pivot_utterance_distinct.pl
40 http://wizzion.com/thesis/results/utterances_with_pivots_distinct_children_sorted
41 Quantity in square brackets denote utterance's "popularity", i.e. the number of distinct children which have uttered the construction.

Thus, it is more reasonable to label micro-grammars hereby introduced as "social" than purely individual (e.g. "cognitive"). But given the size of the CHILDES sample which was analyzed and given that the condition of "random sampling"[42] would hold - which is not granted - than it could be, more or less, expected, that "popularity" of utterances hereby unveiled characterizes not only English language as a mutually shared intersubjective entity, but could also characterize the intensity with which are certain structures encoded in the mind of an individual.

END PIVOT GRAMMARS 13.4.4

It seems as if a non-negligeable amount of salient phenomena were revealed during the analysis of English parts of CHILDES corpus.

Primo, distributions of utterance-level constructions indicated that

- communicative tentatives of prototoddlers are dominated by non-linguistic means (29)
- that distribution of utterances which mothers say to younger children significatively correlates with language produced by older children (30)
- productions of mothers and older toddlers are less predictable than productions of younger toddlers (31)

Secundo, analyses aimed at pivot schemas and their instances (35) suggest that

- most salient (32) and potent (33) pivot schemas coalesce around expressions used for location-related and deictic "pointing" and expressions of "wanting"
- most frequent instances of pivots tend to refocus interlocutor's attention to something else ("another one"), reinforcement of the current situation ("I want some more"), or express child's egocentricity ownership ("that's mine")
- certain utterances can be easily explained as crossovers between two frequent pivots (i.e. "I want" $\times$ "another one" $\rightarrow$ "I want another one")

Tertio, closer inspection of certain pivot schemas instantiated in utterances of biggest number of distinct children shows that

- non-abstract referents of child's linguistic pointing are mainly animates (Daddy, elephant, cow, horse) and color attributes (green, red, yellow, orange, blue) (36)

42 Id est, that CHILDES corpora in general represent a random sample of child's normal verbal interactions.

- children "want" a drink, to see, and to play (37)
- the pivotal affinity of the adjective "little" is in part caused by concrete referents (piggy, baby, ball), in part by fixed expressions ("a little bit") and in part by expressions belonging to both classes ("Mary had a little lamb", "twinkle twinkle little star")[43]

This being said, reader is cordially invited to explore the "results" files in order to identify other interesting (ir)?regularities potentially allowing us to increase the amount of knowledge we have about the *Weltanschauung* of a modal english-speaking toddler.

END ENGLISH-SPECIFIC 13.4

But many other, somewhat more universal "facts", were mined from the CHILDES corpus in the first half of this chapter. Asides the fact that mothers interacting with younger children laugh significatively more often than mothers interacting with older children (27), our initial attention was captivated by the relatively frequent[44] occurence of the consonant nasal labial M in productions of younger children (18).

When it comes to expressions composed of more than one signifier, one fact issued from analysis of child-directed motherese had struck us as particulary salient one. That is, the use of 2nd person singular pronoun "you" (13.3.2) significantly more common than in standard corpora.

Extraction of two or more replicas of 2+-gram sequences juxtaposed to each other within one utterance had lead us to conclusion that intralocutory duplications (13.3.3) are most frequently observed in motherese directed to younger children. Subsequently, the analyses of translocutory duplications - that is, repetitions spanning multiple utterances - has revealed a structural distinction between language of younger and older children: while prototoddlers use repetition of meaning-carrying "lexical" morphemes ("mama", "baby", "ball", daddy"), repetitions of older toddlers are populated by members of the closed set of "grammatical" morphemes ("the", "this", "yes", "here") (23) . The latter distribution being similiar to distributions of the adult grammar, it was hypothetized that during the process of development, child's language gradually adapts to language system of surrounding linguoracles, especially the mother.

A following analysis of "intersubjective replications" - i.e. of cases were a word uttered by the mother was immediately uttered after the child, or vice versa (25)- had indicated, that the hypothesis of a child unilaterally adapting to the mother is not sufficient. More concretely, the summary results presented on Table 26 caused us to state that "mothers tend to prefer reacting to topic changes than changing the topic themselves".

---

43 Child's growing exteroceptic, proprioceptic and/or spatial awareness of the fact that she's "little girl" also plays, of course, an important role.

44 I.e. in contrast with older children.

Thus, it seems, that in a long run - during weeks, months and years - it is the child who adapts to the mother, but in a short span - in concrete scenes lasting seconds and minutes - it is the mother who adapts her topic, her focus, her attention to that of the child.

Asides all these phenomena - and all other explicitely discussed on preceding chapters - we find it important to repeat once more the methodological objective behind this chapter. That is, to show that both relevant and interesting "knowledge" can be extracted from CHILDES corpora by means of a simple, fast and unambigously reproducible method of extraction of patterns attained by means of matching the corpus with PERL-compatible regular expressions.

END QUANTITATIVE 13

CHI				MOT			
PROTO		TODDL		PROTO		TODDL	
5206	&=vocalize .	10683	yeah .	2337	&=involuntary .	6609	oh .
1344	&=babble .	9126	no .	1969	yeah .	5219	no .
1180	&=nonspeech .	3856	oh .	1785	&=nonspeech .	3852	yeah .
1009	&=cry .	3003	mhm .	1458	okay .	3777	okay .
894	&=involuntary .	2769	yes .	1201	&hmm ?	2591	yes .
381	&uh .	2640	there .	1011	&=speechplay .	2563	mhm .
379	&=laugh .	1204	huh ?	1000	come on !	2002	right .
239	&ah .	1129	look .	985	here .	1900	&hmm ?
223	Mama .	816	that .	871	huh ?	1582	there .
206	&=cough .	773	here .	860	uhoh .	1576	what ?
106	Dada .	717	what's that ?	846	no .	1456	that's right .
104	&=labial .	601	Mummy .	764	&=laugh .	1377	what's that ?
87	&eh .	587	okay .	719	what ?	1156	well .
86	&=laughs .	514	uhhuh .	637	look !	1153	look .
79	ball .	489	yup [= yes] .	628	oh .	1144	come on .
67	ooh@b .	422	that one .	558	there you go .	1115	that's it .
55	&=raspberry .	412	&mm .	476	&=labial .	1020	oh dear .
54	&mm .	410	on there .	474	mhm .	1004	thank_you .
54	baby .	406	oh no .	472	thank_you .	820	pardon ?
52	&u:h .	404	in there .	472	hi .	764	no ?
52	Mommy .	396	this .	454	yes .	740	whoops .
49	byebye .	389	more .	433	come (h)ere !	734	what is it ?
48	guh@b .	382	what ?	414	that's right .	726	huh ?
47	up .	380	uhoh .	331	yay .	722	what's this ?
47	no .	375	oh dear .	329	ahhah .	690	what is that ?
45	oo@b .	354	baby .	317	whoa .	582	there you go .
42	&a:h .	337	car .	311	hello .	509	here .
38	Mom .	318	I don't know .	307	&mm .	464	oh no .
36	uguh@b .	317	me .	291	hey .	451	uhhuh .
35	heh@b .	310	what's this ?	281	what's that ?	409	good girl .

Table 29: Most frequent utterance-level constructions produced by english-speaking mothers and children in 2 phases of their development.

	MOT_{PROTO}	MOT_{TODDL}
CHI_{PROTO}	$t = 28.4625$ $d_f = 32679$ $p \leqslant 2.2e-16$ $cor = 0.155$	$t = 0.5$ $d_f = 74078$ $p = 0.6126$ $cor = 0.0019$
CHI_{TODDL}	$t = 5.9801$ $d_f = 70555$ $p = 2.24e-09$ $cor = 0.022$	$t = 317.27$ $d_f = 110006$ $p \leqslant 2.2e-16$ $cor = 0.692$

Table 30: Correlations between distributions of frequences of utterances.

	CHI	MOT
PROTO	$N_d = 3645$ $P_d = 0.0863$ $H = 6.824$	$N_d = 83267$ $P_d = 0.423$ $H = 14.2$
TODDL	$N_d = 120219$ $P_d = 0.41$ $H = 13.7$	$N_d = 199704$ $P_d = 0.538$ $H = 15.37$

Table 31: Number of distinct utterances in diverse datasets and entropies of their distributions.

Score	Pivot	Score	Pivot	Score	Pivot
18472	^that's	6507	another	4320	^I can't
16368	^ what's	5950	^that's a	4288	't know.
16360	^ I want	5841	want to	4239	^I wanna
13527	^ where's	5810	on there.	4235	^ there's a
10640	in there.	5808	little	3950	^ that one
9632	in the	4860	^this is	3790	going to
9513	^there's	4760	that one.	3740	^I want to
8328	's that?	4734	^ another	3600	^ it's a
7335	^I don't	4667	^ where's the	3591	, Mummy.
7320	on the	4608	don't know.	3552	^ here's

Table 32: Thirty 8+grams with highest $score_{pivoteness}$.

Pivot	$H_{contextual}$
^that's X Y	9.25876131528133
^I want X Y	8.96609935363205
^where' X Ys	8.95606540894548
^there's X Y	8.79381971491988
X in the Y	8.74578245657441
X on the Y	8.65695029616604
X Y in there.	8.5923250584143
^this is X Y	8.34192768618957
^that's a X Y	8.20433784100614
X little Y	8.01430314990973

Table 33: Ten CHILD-produced $\mathtt{pivot}_7$ schemas with highest contextual entropy (in shannons).

Utterance	Frequency
another one.	143
I want it.	84
where's it gone?	72
that's it.	69
what's in there?	68
that's mine.	67
where is it?	65
I want another one.	60
I can't do it.	49
look at this one.	48

Table 34: CHILDES utterances most frequently instantiating some $\mathtt{pivot}_7$ schema.

Utterance	Children	Utterance	Children	Utterance	Children
another one.	33	what’s in there?	13	little girl.	10
that’s mine.	27	that’s right.	13	I want another one	10
I want it.	27	look at this.	13	I can’t find it.	1
I want that.	26	it’s all_gone.	13	a little one.	10
that’s it.	25	in the car.	12	where go?	9
yes please.	23	I like that.	12	where are you?	9
I can’t do it.	22	here’s one.	12	there , look.	9
where is it?	19	what’s this one?	11	that’s red.	9
look at that.	17	there’s one.	11	I want this one.	9
go in there.	16	and there.	11	go in here.	9
I want that one.	15	what’s in here?	10	where’s this go?	8
where’s it gone?	14	there’s another one.	10	where’s other one?	8
that’s better.	14	that’s green.	10	that’s yellow.	8
little one.	14	that’s Daddy.	10	that’s a .	8
I want some more.	14	put it in there.	10	that one there.	8

Table 35: Pivot-instantiating CHILDES utterances pronounced by biggest number of distinct children.

that’s	mine [27]
	it [25]
	better [14]
	right [13]
	green [10]
	Daddy [10]
	red [9]
	yellow [8]
	orange [7]
	all [7]
	nice [6]
	blue [6]
	a elephant [6]
	a cow [6]
	you [5]
	my [5]
	horsie [5]
	good [5]
	a car [5]
	...

Table 36: Most popular instances of pivot "ˆthat’s X"

I want	it. [27]
	that. [26]
	that one.[15]
	some more.[14]
	another one.[10]
	this one.[9]
	this.[8]
	some.[7]
	a drink.[7]
	two.[6]
	to see.[6]
	one.[6]
	more.[6]
	to play.[5]
	down.[5]

Table 37: Most popular instances of pivot "^I want X"

little	one. [14]
	girl. [10]
	lamb. [5]
	boy. [5]
	bit. [5]
	man. [4]
	car. [4]
	piggy. [3]
	ball. [3]
	baby. [3]

a	little	one. [10]
		boy. [3]
		bit. [3]
		box. [3]

that	little	one [3].
twinkle twinkle	little	star. [4]
Mary had a	little	lamb. [3]

Table 38: Most popular instances of pivot "X little Y"

14 SUMMA III

Ideas are never static but develop across time and context, constantly cross-fertilizing with other currents of thought.

— Edwin F. Bryant

Hence ends the last part of the first volume of the Thesis aiming to offer certain *fragments of evidence* of the validity of the theory of intramental evolution. Two principal ways of acquisition of such fragments have been presented:

1. **qualitative**: holistic, naturalistic and multi-modal observations of development of one specific child, from prenatal period onwards
2. **quantitative** analyses of patterns recurrent in transcripts produced by hundred children immortalized in the CHILDES corpus

but their combination is, of course, also possible.

14.1 CROSSROADS OF THOUGHTS

As is often the case in science, a crossover between the methods can also lead to interesting results. Thus, it was by means of PERL regex pattern matching that the following schema was detected in behaviour of Kuczaj's (Kuczaj and Maratsos (1975) son named Abe (2;5.23):

Listing 7: Some lines from abe009.cha (PID 11312/c-00016245-1) transcript

```
301 *FAT:   okay (.) here's another one a cow ate the carrot
306 *CHI:   cow ate the carrot .
310 *FAT:   okay (.) now do this one the boy fell down the stairs
     .
315 *CHI:   boy fell down the stairs .
322 *FAT:   dinner was eaten by the boy .
325 *CHI:   dinner eaten by the boy .
...
365 *FAT:   the cow ate a carrot .
368 *CHI:   cow ate a carrot .
372 *FAT:   the little boy is happy .
375 *CHI:   happy .
...
398 *FAT:   a cow ate carrot the .
```

```
401 *CHI:   a cow ate carrot the .
...
507 *FAT:   the cow did not eat his dinner .
510 *CHI:   cow didn't eat his dinner he can't get snacks .
516 *FAT:   that's right no dinner (.) no snacks here's another
    game the elephant cannot go home
523 *CHI:   elephant can't go home .
526 *FAT:   nobody hit me .
529 *CHI:   nobody hit me .
533 *FAT:   the boy did not eat any cookies .
536 *CHI:   boy can't eat any cookies .
539 *FAT:   the cow cannot eat grass .
542 *CHI:   cow can't eat grass .
...
551 *FAT:   the boy did not sleep .
554 *CHI:   boy can't not sleep .
...
582 *FAT:   the goat eat did his dinner .
585 CHI:    goat didn't eat his dinner .
588 *FAT:   the boy not did eat any cookies .
591 *CHI:   boy can't eat any cookies .

642 *FAT:   we can play some more tomorrow too (.) okay .
647 *CHI:   tomorrow too (.) boy can't eat his carrots boy can't
    eat his carrots
654 *MOT:   do you want to go outside for awhile (.) Abe ?
659 *CHI:   play outside (.) boy can't eat his carrots .
```

Closer inspection of the above-listed father-son interaction unveils multiple interesting phenomena:

Primo, Kuczaj's son consistently used the construction "boy can't" in cases where he should repeat his father's "boy did not". Well beyond the objectives of this Thesis is the question whether this phemonenon is to be explained by mismatch on the level of passive, perceptive morphosyntactic C-structures (c.f. also 12.12.1) or whether it has more to do with mismatch of productive P-structures leaves a.

But mismatch there is, and for a reason unknown, Abe was consistently crossing-over the external schemata of a form "boy did not X" with private schema "boy can't X".

Secundo, other *crossovers between external stimuli (e.g. utterances produced by external linguistic oracles like parents, peers or teachers) and child's private world of needs, wants and protothoughts* are to be observed on lines 510 and 647 of the transcript. In the first case, father's utterance "cow did not eat his dinner" is augmented with Abe's private "he can't get snacks" which makes father to **react** to the "snack" topic[1]. without preliminary intention to do so.

1 Snacks are also mentioned in other Abe transcripts, on line 45 in abe004.cha mother urges the child to eat with a threat "okay (.) come eat or no snacks later on ." and on line 516 of abe017.cha where Abe offers his father an apple "as a snack"

Even more important - for the purpose of verification of theory hereby presented - is the crossover construction which emerges at the very end of the transcript, in the moment where father closes the session with words "we can play some more tommorrow", thus putting Abe in a position of a brief *vacuum* where anything can be said. The vacuum is immediately filled by Abe's production and replication (twice on lin 647, one on line 659) of the construction "boy can't eat his carrots".

Note that nowhere in the transcript had the father uttered a construction with "boy" as a subject and "carrot" as an object[2]. Thus, the expression with which Abe closes the language game seems to be his own invention, an invention which we consider to be the product of the crossover summarized on Table 39.

306	cow ate the carrot
361	cow ate a carrot
510	cow didn't eat his dinner he can't get snacks
542	cow can't eat grass
536	boy can't eat any cookies
554	boy can't not sleep
591	boy can't eat any cookies
647 647 659	boy can't eat his carrots

Table 39: Interphrastic crossover behind Abe's "boy can't eat his carrots".

Given all this, a question can be posed: "why carrots?". Why not "dinner", "cookies", "cheese" or "grass" which are also used as direct objects of "eat-ing" mentioned in the transcript? Why was it the substantive "carrot" which, as Tomasello (2009) would say, had "filled the slot"?

It may be the case that multiple cognitive processes and biases are to be taken into account in order to answer the question:

1. the primacy effect: term "carrot" is the first concrete object of eating mentioned in Kuczaj's "repeat after me" language game

2. the frequency effect: Abe was three times exposed to father's production of the term "carrot", i.e. more than in case of "cookies" (2 times), "grass" (2 times) or "cheese" (once)

2 The word "carrot", in fact, does not occur in any other Abe transcript, only in abe009.cha

3. the perturbation effect: term "carrot" was once heard (line 398) and once produced (line 401) in a syntactically anomalous construction "a cow ate carrot the" [3]

4. the semantic consistency effect: "boy can't eat his carrots" refers to more plausible a scenario than, for example, "boy can't eat his grass"

5. priming etc.

It seems to us evident that **all** these processes and biases are to be taken into account by anyone hoping to develop a reasonable *theory of crossover among linguistic structures* which does not contradict but rather naturally extends both cognitivist, connectionist, usage-based[4] paradigms which dominate contemporary developmental psycholinguistics.

But since the objectives addressed in the second volume of this work will be principially computational ones, let's now start concluding this theoretical volume with one sole principle which can be immediately deployed in a functional program.

14.1.1 THE LINGUISTIC CROSSOVER PRINCIPLE

Fitness of product of the crossover of (linguistic) schemes A and B is proportional to fitness of A and B as well as to amount of features which A and B share.

END THE CROSSOVER PRINCIPLE 14.1

A more formal and geometric variant of this principle shall be furnished in the second volume of this work. For the time being, let's just elucidate that by the term of "features" we mean not only overlap between "semantic" features (c.f. "the semantic consistency effect" in enumeration above) of two "parental" schemes, but also overlap between prosodic, phonologic, morphologic, syntactic or even pragmatic characteristics of schemes which are to be fusioned.

As in case of any creative, poietic act, the form and content, the program and the data, fuse.

Thus, "AFE" and "OPICA" yield "API" (12.10.1), "BAJA" and "ANAN" yield "BANAN" (12.11.1) etc. not only because they denote the same meaning but also because they are phonetically similar."MAHLEN" and "BAUEN" yields "MAUEN" (12.11.2) not only because their signifiants can be matched by the pattern /Lab_{occ}A*EN/ but also because within certain subspace of the envelopping semantic space they

3 Exposure to such anomalous stimuli can be potentially asessed in terms of P600 event-related potential. It cannot be excluded that such a P600-related anomalies attain higher level of salience and activation than terms occurent in coherent contexts.

4 And with little bit of luck also mentalist

tend to be quite close (i.e. they both denote object-manipulating, constructive, creative activites etc.).

Hence, Abe joyfully utters "the boy can't eat his carrots" not only because "boy eats" is semantically closer to "carrot" than to "grass" but also because - on the morphosyntactic level - expressions "cow can't eat grass" and "boy can't eat any cookies" are similiar enough to induce the activation of the pathway like "X can't eat Y" $\rightarrow$ subsequently filled with most affine fillers ("boy" before "can't" and "carrot" after "eat").

Subtilities aside, the linguistic crossover principle can be further elucidated by the following aphorism:

14.1.2 OF CROSSOVERS AND ANALOGIES (APH)

If the reader has understood that events which we have labeled as "linguistic crossovers" could elucidate phenomena to which the traditional cognitive science refers by the term "analogy", then the reader has understood us well.

END OF CROSSOVER AND ANALOGIES 14.1.2

... and the precept

"whenever You notice an analogy or schematization, seek for existence the implicit structural crossover behind it"

can turn out to be useful methodological "rule of thumb" for any researcher potentially interested by our proposal.

END CROSSROADS OF THOUGHTS 14.1

14.2 AXES OF ANALYSIS

Diverse aspects of crossovers produced by IM, Abe or other toddlers, can be studied. Of non-negligeable importance is the analysis in terms of temporal interval between the last activation of crossover's input schemata and crossover's output product. Thus, in case of Abe's carrots, minutes had to pass between Abe's productions of all initial "carrot" and "boy can't" (inputs) expressions and his final "boy can't eat his carrots" (output).

Many crossovers uttered by IM also had a property of mixing together schemata separated from each other and their product by minutes of other content, c.f. the (MAMA + MIMI $\rightarrow$ MAMI, 12.11.2). But sometimes - as in case of PIJEN (12.10.1) - the timespan seemed to be even shorter and crossover seemed to be occuring in short term memory or even in a much more volatile phonologic buffer. And yet in other cases (12.11.1), a simple trick of letting child hear "AJAN" caused a 5-month old latent schema to get reactivated, fuse with much more recent BAJA and form the globally optimal form BANAN.

Another important aspect is the origin of input schemata. Analyzed from this perspective, one can state that nature of majority of crossovers noted down in the volume, was of following kind:

$$\frac{\begin{array}{c}\text{external}\\ \text{personal}\end{array}}{\text{CROSSOVER}}$$

whereby "external" denotes the schemata encoded in the stimuli to which the child is exposed (e.g. motherese utterances etc.) while "personal" denotes private and often unique idioglottic structures already encoded and productive within the mind of the given child.

Crossovers between two or more purely "personal" schemata also seem possible. Unfortunately for empiric science, they are either impossible to access (as is the case "dreaming"[5]) or difficult to recognize as what they truly are (e.g. certain babbling sequences etc.)

END AXES OF ANALYSIS 14.2.0

14.3 THE SOURCE OF VARIATION

Encoded in the material substrate of the brain, schemata are subjects to same physical laws of entropy and decay as the brain and body itself. Cognitive schemata are not engraved onto some kind of eternally lasting crystal. Humans forget [6].

Forgetting is a form of variation and as every form of variation, it can sometimes lead to disastrous loss of information. But less rarely, it can also cause one to discard previous "locally optimal" information, thus giving one the impetus to seek more globally optimal states.

Another source of variation inherent to the child is her tendency "to want another one" and "to play". While many phenomena related to craving and wanting more can be in large extent explicated in terms of standard behaviorist theories (reinforcement, reward etc.) or "3rd noble truth" already posited by Shakyamuni some 25 centuries ago ((Lama et al., 2005)), child's everactual readyness to play does not cease to struck us with such intensity that even after months of observations and empiric research, we still consider our initial definition of the "child" (Section 5.2) as a reasonable and a valid one.

5 For if there is a realm inaccessible to reason of an adult man than it is indeed the realm of toddler's dreams.

6 Sometimes the tendency to forget is so strong that some researchers (c.f. 9.4.2) have even forgotten that humans forget

What's more, our research had lead us to conviction that a modal toddler is much more a member of the species *Homo Ludens* (Huizinga, 1956) than of the species *Homo Sapiens*. And if there is one single thing which should be potentially reproached to otherwise most advanced and complete theory of linguistic development - i.e. the usage-based theory of Tomasello - then let it be this one:

14.3.1 EXTENDING USAGE-BASED PARADIGM (TXT)

That language development could not be possible without child's ability to share attention with other humans is true.

And it is also true that recurrence and distribution of patterns among and within diverse "usage scenarios" to which child is exposed and in which she is supposed to act, all that is an indispensable prerequisite to the success of the whole process.

But a similiarly indispensable prerequisite it's child's tendency to play with sounds, words, sentences and whole contexts. To laugh, to sing, to talk to herself, to say "no" when the child already *knows* that the only word which her interlocutor does NOT want to hear is..."no".

To playfully explore the limits of principles and rules and to do so in order to break them. To playfully explore limits of one's world.

END EXTENDING USAGE-BASED PARADIGM 14.3.1

And to feel Joy *during and because* all of that.

END THE SOURCE OF VARIATION 14.3

14.4 FROM SELECTION TO REPLICATION

Principial source of variation thus ellucidated, the theory of intramental evolution still lacks a component without which it could not be neither formalized nor translated into a functional computer program. That is, the description of the bridge between process of "selection" and process of "replication". What is still missing is such a fitness function which could be pertinent to the process of language development.

In other terms, what we still lack is a criterion by means of which one's language-processing system could evaluate which schemata (or their ordered sets) are "fit" for linguistic communication and which are not.

We posit the following principle in order to fill this gap.

14.4.1 THE PRINCIPLE OF EXOGENOUS SELECTION (DEF)

The more schema S encoded in cognitive system C matches the data produced by external oracle O, the more it is probable that S shall replicate into another region of C.

END THE PRINCIPLE OF EXOGENOUS SELECTION 14.4.1

Stated in more Piagetian and less probabilistic terms: whenever the schema succeeds to assimilate linguistic stimulus produced by the person endowed with implicit authority[7], then the schema gets copied in other region of child's mind.

Stated even more simply, the principle can be compressed in the following precept:

14.4.2 MPR PRECEPT (APH)

Matching Pattern Replicates.

END MPR PRECEPT 14.4.2

And that's it. Given that within the brain of a child replicated schemata are practically immediately subjected to forces of decay and (play | forget)ful variations, three words of MPR precept prepare the territory for great deal of adaptation which could potentially follow.

Under this view is the computational burden related to information-processing, noise-filtering and the selection of structures delegated to external oracles. By "uttering this and not that", by exposing child's schemata to this "data" and not that "data", indeed by such indirect mediated means do the model persons influence the development of structures in child's mind.

Asides few dozens of innate schemata is the mind of the nascent child filled mainly with unceasing swarming of images issued from the unknown realm of φαντασία. All the rest - including labels, rules and criteria - comes from outside, neatly packaged, preprocessed and preselected by caring oracles.

It is in this sense that the adjective **exogenous** is to be understood.

END THE PRINCIPLE OF EXOGENOUS SELECTION 14.4.2

7 Child experiences on a daily basis how persons like mother, father, grandparents, teachers, older siblings, older peers etc. succeed to solve problems which she is unable to solve on her own. In computational theory such problem-solvers able to yield immediate and correct answer are called oracle machines Turing (1939). C.f. Clark (2010) for discussion of how involvement of certain oracles, called Minimally Adequate Teachers, can reduce the computational complexity of the problem of grammatical inference of context free languages.

Nothing precludes that in a healthy symmetric relation between the parent and a child, parent can approach the child as if she was a computational oracle able to immediately solve certain types of problems. In such a case, adaptation and evolution leads to a sort of bilateral co-adaptative, co-evolutive interlock in which the child does for a parent what a parent does for a child.

That is, by selecting and exposing the parent with the data-to-be-matched, the child indirectly influences the population dynamics of schemata encoded in parent's mind. Et vice versa.

Willingness of many mothers to adapt to the topic proposed by their child (Table 25) as well as their readiness to perceive a fragile and powerless hominid not as an alien but as a "2nd person singular" (Table 23) lead us to belief that authentic, non-superficial comprehension of "the Other" (Buber, 1937) - i.e. "love" - is not a privilege but rather an essential prerequisite of successful co-adaptation of two minds and souls whose destinies are inexorably bound to each other.

Love (DEF)

« Strong positive emotional relation to persons, things, ideas or self. Conscious, effective, voluntary acceptation of value of the other in one's life. Readiness to be hostage for the other (Levinas). Platonic tradition accentuates that less perfect is attracted by more perfect (love as a longing for what one does not have, especially beauty). In christian tradition humans respond to the gift of existence (life, world, happiness, friendship, family) with love, id est by devotion to well-being of the other, which does not await anything in return (love as devotion). Love expresses itself on all levels of human being, physical, personal and spiritual. It is the only solid bound between humand and the ultimate source of everything in the world which has real value.» (Sokol, 1998)

END LOVE (DEF) 14.4.2

END FROM SELECTION TO REPLICATION 14.4

This being said, we end the first volume of this work with an expression of a simple hope. Of a hope that on preceeding pages we have already succeeded to furnish some fragmented, preliminary and undoubtably incomplete yet consistent evidence supporting the theory which was initiated by two words forming the Thesis

Mind evolves.

END SUMMA III 14

Part IV

SIMULATIONS

There is an appealing symmetry in the notion that the mechanisms of natural learning may resemble the processes that created the species possessing those learning processes.

— D.E. Goldberg and J. Holland

This part can be understood as a collection of four scientific articles. Each article describes a distinct simulation and can be read individually. Aspiration common to all articles is to provide different facets of cognitively plausible, *ex computatio et simulatio* proof-of-concept for theory of intramental evolution.

Zeroth simulation aspires to demonstrate that Evolutionary Computation (EC) can offer useful insights to an agent hoping to break the code of an unintelligible corpus (e.g. to help decode riddle as cryptic as the Voynich Manuscript). First simulation aspires to demonstrate that EC can be a useful means of multiclass classification of textual documents according to their semantic content (e.g. and in a Big Data scenario could potentially lead to results as good as those produced by connectionist "deep learning" methods). Second simulation aspires to demonstrate that EC can help to identify useful solutions to problem of multiclass part-of-speech classification. Third simulation aspires to demonstrate that EC can pave the way to induction of plausible micro-grammars from solely positive corpus of motherese utterances.

15 BREAKING INTO UNKNOWN CODE

15.1 GENERIC INTRODUCTION

A cryptologue posed with an unbroken cipher is, in certain sense, in a position similar to a child (P+19) which has just been born into our common world. Both cryptologue and a child are confronted with novel constellations of symbols and features. Both assume that the data with which they are confronted - a motherese (P+90-93) utterance perceived by a child or a cipher studied by a cryptologue - ultimately carry a certain meaningful message. Both combine their ingenuity with relentless perseverance: both accept that the path to success leads through ocean of trials and errors (P+22). Ultimately, they both transcend their initial state of limited knowledge and attain understanding: child shall understand the world and the scholar shall understand the cipher.

This analogy between a child and a cipher-breaker can be pushed even further in case we speak about the cipher stored in the enigmatic medieval Voynich Manuscript (VM). This is so because VM contains a non-negligible amount of *visual content* and it can be rightfully speculated that if VM contains a cipher to be decoded, than the deciphering process (and its subsequent evaluation) shall be founded on discovery of associations between VM's visual content and the adjacent "voynichese" script.

This is - we believe - similar to the position of a visually non-impaired human child who acquires a non-negligible amount of information about her world and her language by means of associating the components of surrounding visual scenes with simultaneously heard phonemic sequences (e.g. "red ball in mama's hand").

This being said, let's now present first implications of our "child as a cryptologue" analogy, as published in the article Hromada (2016a).

15.2 ABSTRACT

Voynich Manuscript is a corpus of unknown origin written down in unique graphemic system and potentially representing phonic values of unknown or potentially even extinct language. Departing from the postulate that the manuscript is not a hoax but rather encodes authentic contents, our article presents an evolutionary algorithm which aims to find the most optimal mapping between voynichian glyphs and candidate phonemic values.

Core component of the decoding algorithm is a process of maximization of a fitness function which aims to find most optimal set of substitution rules allowing to transcribe the part of the manuscript - which we call the Calendar - into lists of feminine names. This leads to microgrammars which allow us to consistently transcribe dozens among three hundred calendar tokens into feminine names: a result far surpassing both "popular" as well as "state of the art" tentatives to crack the manuscript. What's more, by using name lists stemming from different languages as potential cribs, our "adaptive" method can also be useful in identification of the language in which the manuscript is written.

As far as we can currently tell, results of our experiments indicate that the Calendar part of the manuscript contains names from baltoslavic, balkanic or Hebrew language strata. Two further indications are also given: primo, highest fitness values were obtained when the crib list contains names with specific in-fixes at token's penultimate position as is the case, for example, for Slavic feminine diminutives (i.e. names ending with -ka and not -a). In the most successful scenario, 240 characters contained in 35 distinct Voynichese tokens were successfully transcribed. Secundo, in case of crib stemming from Hebrew language, whole adaptation process converges to significantly better fitness values when transcribing voynichian tokens whose order of individual characters have been reversed, and when lists feminine and not masculine names are used as the crib.

15.3 INTRODUCTION

Voynich Manuscript (VM) undoubtedly counts among the most famous unresolved enigmas of the medieval period. On approximately 240 vellum pages currently stored as manuscript (MS) 408 in Yale University's Beinecke Rare Book and Manuscript Library, VM contains many images apparently related to botanics, astronomy (or astrology) and bathing. Written aside, above and below these images are bulks of sequences of glyphs. All this is certain.

Also certain seems to be the fact that in 1912, VM was re-discovered by a polish book-dealer Wilfrid Voynich in a large palace near Rome called Villa Mandragone. Alongside the VM itself, Voynich also found the correspondence - dating from 1666 - between Collegio Romano scholar Athanasius Kircher and the contemporary rector of Charles University in Prague, Johannes Marcus Marci. Other attested documents - e.g. a letter from 1639 sent to Kircher by a Prague alchemist Georg Baresch - also indicate that during the first half of 17th century, VM was to be found in Prague. The very same correspondence also indicates that VM was acquired by famous patron of arts, sciences and alchemy, Emperor Rudolf II. [1]

Asides this, one more fact can be stated with certainty: the vellum of VM was carbon-dated to the early 15h century (Hodgins, 2014).

15.3.1 PRE-DIGITAL TENTATIVES

Already during the pre-informatic era of first half of 20th century had dozens, if not hundreds, men of distinction invested non-negligible time of their life into tentatives to decipher the "voynichese" script.

Being highly popular in their time, many such tentatives - like that of Newbold who claimed to "prove" that VM was encoded by Roger Bacon by means of 6-step anagrammatic cipher (Newbold, 1928), or that of Strong (Strong, 1945) who claimed VM to be a 16th-century equivalent of the Kinsey Report" - may seem to be, when looked upon through the prism of computer science, somewhat irrational [2].

C.f. (d'Imperio, 1978) for a overview of other 20th-century "manual" tentatives which resulted in VM-deciphering claims. After description of these tentatives and and after presentation of informationally very rich introduction to both VM and its historical context, d'Imperio adopts a skeptical stance towards all scholars who associated VM's origin with the personage of Roger Bacon[3].

In spite of skeptic who she was, d'Imperio hadn't a priori disqualified a set of hypotheses that the language in which the VM was ultimately written was Latin or medieval English. And such, indeed, was the majority of hypotheses which gained prominence all along 20th century.[4].

15.3.2 POST-DIGITAL TENTATIVES

First tentatives to use machines to crack the VM date back to prehistory of informatic era. Thus, already during 2nd world war did the cryptologist William F. Friedman invited his colleagues to form

1 Savants which passed through Rudolf's court included Johannes Kepler, Tycho de-Brahe or Giordanno Bruno. The last one is known to have sold a certain book to the emperor for 600 ducats.

2 Note, for example, Strong's "translation" of one VM passage: "*When the contents of the veins rip, the child comes slyly from the mother issuing with leg-stance skewed and bent while the arms, bend at the elbow, are knotted like the legs of a craw-fish.*" Strong (1945) Note also that such translation was a product of man who was "a highly respected medical scientist in the field of cancer research at Yale University" (d'Imperio, 1978).

3 "I feel, in sum, that Bacon was not a man who would have produced a work such as the Voynich manuscript...I can far more easily imagine a small society perhaps in Germany or Eastern Europe (d'Imperio, 1978, 51)"

4 Note that such pro-English and pro-Latin bias can be easily explained not by the properties of VM itself, but by the simple fact that first batches of VM's copies were primarily distributed and popularized among Anglosaxon scholars of medieval philosophy, classical philology or occidental history

"extracurricular" VM study group - programming IBM computers for sorting and tabelation of VM data was one among the tasks. Two decades later - and already in position of a first chief cryptologist of the nascent National Security Agency - Friedman had formed the 2nd study group. Again without ultimate success.

One member of Friedman's 2nd Study Group After was Prescott Currier whose computer-driven analysis led him to conclusion that VM in fact encodes two "statistically distinct" (Currier, 1970) languages. What's more, Currier seems to have been the first scholar who facilitated the exchange and processing of Voynich manuscript by proposing a transliteration[5] of voynichese glyphs into standard ASCII characters. This had been the predecessor of the European Voynich Alphabet (EVA) (Landini and Zandbergen, 1998) which had become a de facto standard when it comes to mapping of VM glyphs upon the set of discrete symbols.

Canonization of EVA combined with dissemination of VM's copies through Internet have allowed more and more researchers to transcribe the sequence of glyhps on the manuscript into ASCII EVA sequences. Is is thanks to laborious transcription work of people like Rene Zandberger, Jorge Stolfi or Takeshi Takahashi that verification or falsification of VM-related hypotheses can be nowadays in great extent automatized.

For example, Stolfi's analyses of frequencies of occurrence of different characters in different contexts has indicated that majority of Voynichese words seems to implement a sort of tripartite crust-core-mantle (or prefix, infix, suffix) morphology. Later study has indicated that the presence of such morphological regularities could be explained as an output of a mechanical device called Cadran grill (Rugg, 2004). The "hoax hypothesis" is also supported by the study of Schinner (2007) who suggested that "the text has been generated by a stochastic process rather than by encoding or encryption of language". Pointing in the similar direction, the analysis also concludes that "glyph groups in the VM are not used as words".

On the other hand, a methodology based on "first-order statistics of word properties in a text, from the topology of complex networks representing texts, and from intermittency concepts where text is treated as a time series" presented in (Amancio et al., 2013) lead its authors to conclusion that VM "is mostly compatible with natural languages and incompatible with random texts". Simply stated, the way how diverse "words" are distributed among different sections of VM indicates that these words carry certain semantics. And this indicates that VM, or at least certain parts of it, are not a hoax.

5 In this article we distinguish transliteration and transcription. Transliteration is a bijective mapping from one graphemic system into another (e.g. VM glyphs is transliterated into ASCII's EVA subset). Transcription is a potentially non-bijective mapping between symbols one one side and sound- or meaning- carrying units on the other.

15.3.3 OUR POSITION

Results of (Amancio et al., 2013) had made us adopt the conjecture "VM is not a hoax" as a sort of a fundamental hypothesis accepted *a priori*. Surely, as far as we stand, it could not be excluded that VM is a work of an abnormal person, of somebody who suffered severe schizophrenia or was chronically obsessed by internal glossolalia (Kennedy and Churchill, 2005). Nor can it be excluded that the manuscript does not encode full-fledged utterances but rather lists of indices, sequences or proper names of spirits-which-are-to-be-summoned or sutra-like formulas compressed in a sort of private pidgin or a sociolect. But given VM's ingenuity and given the effort which the author had to invest into the conception of the manuscript and given a sort of "elegant simplicity" which seems to permeate the manuscript, we have felt, since our very first contact with the manuscript, a sort of obligation to interpret its contents as meaningful.

That is, as having the capability of denoting the objects outside of the manuscript itself. As being endowed with the faculty of reference to the world (Frege, 1994) which we, 21st century interpreters, still inhabit hundred years after VM's most plausible date of conception.

It is with such bias in mind that our attention was focused upon a certain regularity which we have later decided to call "the primary mapping".

15.3.4 PRIMARY MAPPING

Condition sine qua non of any act of deciphering is a discovery of rules which allow to transform initially meaningless cipher into meaningful information. In most trivial case, such deciphering is facilitated by a sort of Rosetta Stone (Champollion, 1822) which the decipherer already has at his disposition. Since both the cipher-text as well as the plain-text (also called "the crib") are explicitly given by the Rosetta Stone, discovery of the mapping between the two is usually quite straightforward.

The problem with VM is, of course, that it seems not to contain any explicit key which could help us to decipher its glyphs. Thus, the only source of information which could potentially help us to establish reference between VM's glyphs and the external world are VM's drawings. One such drawing present atop of folio f84r is shown on Figure 26.

Figure 26 displays twelve women bathing in eight compartments of a pool. Bathing women is a very common motive present in VM and there seems to be nothing peculiar about it. The fact that word-like sequences are written above heads of these women is also trivial.

Figure 26: Drawing from folio f84r containing the primary mapping.

One can, however, observe one regularity which seems to be interesting. That is, in case two women bath in the same compartment, the compartment contains two word-like sequences. If one woman bathes in the compartment, there is only one word-like sequence which is written above her head.

One figure - one word, two figures - two words. This principle is stringently followed and can be seen on other folios as well. What is more, the words themselves are sometimes similar but they are not the same. Such trivial observations lead to trivial conclusion: these word-like sequences are labels.

And since these names are juxtaposed to feminine figures, it seems reasonable to postulate that these labels are, in fact, feminine names. This is the primary mapping.

15.3.5 three conjectures

Method which shall be described in following sections can be considered as valid only under assumption that following conjectures are valid:

1. "the primary mapping conjecture" : voynichese words asides feminine figures are feminine names
2. "diachronic stability of proper names" : proper names are less prone to diachronic change than other language units
3. "Occam razor" : instead of containing a sophisticated esoteric cipher, VM simply transmits a text written in an unknown script

Further reasons why we consider "the primary mapping conjecture" as valid shall be given alongside our discussions of "the Calendar". When it comes to conjecture postulating the "diachronic stability of proper names", we could potentially refer to certain cognitive peculiarities or how human mind tends to treat proper names (Imai and Haryu, 2001). Or focus the attention of the reader to the fact that for practically every human speaker, one's own name undoubtedly belongs among the most frequent and most important tokens which one hears or utters during whole life. This results in a sort of stability against linguistic change and allow the name to cross the centuries with higher probability than words of lesser importance and frequency.

But instead of pursuing the debate in such a direction, let's just point out that successful decoding of Mycenaean Linear script B ((Ventris and Chadwick, 1953) would be much more difficult if certain toponyms like *Amnisos, Knossos or Pylos* haven't succeeded to carry their phonetic skeleton through aeons of time.

At last but not least, the "Occam razor conjecture" simply explicitates the belief that **a reasonable scientist should not opt to explain VM in terms of anagrams and opaque hermeneutic procedures if similar - or even more plausible - results can be attained when approaching VM as it was a simple substitution cipher.**

15.4 METHOD

The core of our method is an optimization algorithm which looks for such a candidate transcription alphabet A_x which, when applied upon the list of word types occurring in VM's Calendar section yields an output list whose members should be ideally present in another list, called the Crib. The optimization is done by an evolutionary strategy - an individual chromosome encode a candidate transcription alphabet and a fitness function is given as a sum of lengths of all tokens which were successfully transcribed from Calendar to a specified Crib.

15.4.1 CALENDAR

Six among twelve words present on Figure 26 occur only on folio f84r. Six others occur on other folios as well, and five of these six words occur also as labels near feminine figures displayed on 12 folios of the section commonly known as "Zodiac". It is like this that our attention was focused from the limited corpus of "primary mapping" towards more exhaustive corpus contained in the Zodiac.

Every page of Zodiac displays multiple concentric circles filled with feminine figures. Attributes of these figures differ - some hold torches, some do not, some are bathing, some are not - but one pattern is fairly regular. Asides every woman there is a star and asides every star, there is a word.

While some authors postulate that these words are names of stars or names of days, we postulate that these words are simply feminine names[6]. From Takahashi's transliterations of twelve folios of the Zodiac we extract 290 tokens which instantiate 264 distinct word types.

To evit possible terminological confusion, we shall denote this list of 264 labels[7] with the term Calendar. Hence, Zodiac is the term to refer to folios f70v2 - f73v, while Calendar is simply a list of 264 labels. Total length of this 264 labels is 2045 letters. These characters are chosen from 19-symbol ($|A_{cipher}| = 19$) subset of the EVA transliteration alphabet.

15.4.2 CRIBBING

Cribbing is a method by means of which a hypothesis, that the Calendar contains lists of feminine names, can potentially lead to deciphering of the manuscript. For if the Calendar is indeed such a list, then one could use lists of existing and attested feminine names as hypothetical target "cribs".

In crypt-analytic terms, an intuition that the Calendar contains feminine names makes it possible to perform a sort of known-plain-text attack (KPA). We say "a sort of", because in case of VM are the "cribs" upon which we shall aim to map the Calendar, not known with 100% certainty. Hence, it is maybe more reasonable to understand the cribbing procedure as the plausible-plain-text attack (PPA).

This beings said, we label as "cribbing" a symbol-substituting procedure $P_{cribbing}$ which replaces symbols contained in the cipher (i.e. in the Calendar) with symbols contained in the plain-text. Hence, not only cipher but also plain-text are inputs of the cribbing procedure.

Every act of execution of $P_{cribbing}$ can be followed an act of evaluation of usefulness $P_{cribbing}$ in regards to its inputs. The ideal procedure would result in a perfect match between the rewritten cipher and the plain-text, i.e.

$$P_{cribbing}(cipher) == plain - text$$

On the other hand, a completely failed $P_{cribbing}$ results in two corpora which do not have anything in common.

And between two extremes of the spectrum, between "the ideal" and "the completely failed", one can place multitudes other procedures, some closer to the ideal than the others.

This makes place for optimization.

6 It cannot be excluded, however, that they all this at once. Note, for example, that in many central European countries, it is still a fairly common practice to attribute specific names to specific days in a year, i.e. "meniny".

7 Available at http://wizzion.com/thesis/simulationo/calendar.uniq

Listing 8: Discrete cross-over

```
#discrete crossover
my $child_genome;
my $i=0;
for (@mother_genome) {
        if ($_ ne $father_genome[$i]) {
                rand > 0.5 ? ($child.=$mother_genome[$i]) : (
                    $child.=$father_genome[$i]);
        } else {
                $child_genome.=$mother_genome[$i];
        }
        $i++;
}
```

15.4.3 OPTIMIZATION

All experiments described in the next section of this article implement an evolutionary computation algorithm strongly inspired by the architecture of Canonical genetic algorithm (CGA, P+46) Holland (1992); Rudolph (1994). Hence, initial population is randomly generated and the fitness-proportionate (i.e. "roulette wheel", P+42) selection is used as the main selection operator. But contrary to CGAs, our optimization technique does not implement a classical single-point crossover but rather a sort of "discrete crossover" which takes place only in case that parent individuals have different alleles of a specific gene.

Another reason why our solution can be considered to be more similar to evolutionary strategies (Rechenberg, 1971) than to CGAs is related to the fact that it does not encode individuals as binary vector (P+48). Instead, *every individual represents a candidate mono-alphabetic substitution cipher* application of which could, ideally, transform the Calendar into a crib. More formally: given that cipher is written in symbols of the alphabet $A_{\mathtt{cipher}}$ and given that the crib is written in symbols of the alphabet $A_{\mathtt{crib}}$, then each individual chromosome will have length of $|A_{\mathtt{crib}}|$ genes and every individual gene could encode one among $|A_{\mathtt{cipher}}|$ values.

Size of the search space is therefore $|A_{\mathtt{cipher}}|^{|A_{\mathtt{crib}}|}$. Search for optima in this space is governed by a fitness function:

$$F_{P_{\mathtt{cribbing}}} = \sum_{w \in \mathtt{cipher} \wedge P_{\mathtt{cribbing}}(w) \in \mathtt{crib}} \mathtt{length}(w)$$

where *w* is a word type occurring in the cipher (i.e. in the Calendar) and which, after being rewritten by $P_{\mathtt{cribbing}}$ also matches a token in the input crib. Given that the expression $\mathtt{length}(w)$ simply denotes w's character length, the fitness function of the candidate transcription procedure $P_{\mathtt{cribbing}}$ is thus nothing else than the sum of char-

Listing 9: Cipher2Dictionary adaptation fitness function

```
#Fitness Function
my $text=$calendar;
my $old = "acdefghiklmnopqrsty";
my %translit;
@translit{split //, $old} = split //, $individual;
$text =~ s/(.)/defined($translit{$1}) ? $translit{$1} : $1/eg; # core transcription of calendar content
my %matched;
for (split/\n/,$text) {
        my $token=$_;
        if (exists $crib{$token}) {
                @antitranslit{split //, $individual} = split //, $old;
                $token =~ s/(.)/defined($antitranslit{$1}) ? $antitranslit{$1} : $1/eg;
                my $t=$token;
                $matched{$t}=1;
        }
}
for (keys %matched) {
        $Fitness[$i]+=length $_;
}
```

acter lengths of all distinct labels contained in the Calendar which $P_{cribbing}$ successfully maps onto the feminine names contained in the input crib.

15.5 EXPERIMENTS

Within the scope of this article, we present results of two sets of experiments which essentially differed in the choice of a name-containing cribs.

Other input values (e.g. Takahashi's transliteration of the Calendar used as the cipher) and evolutionary parameters (total population size = 5000, elite population size = 5, gene mutation probability <0.001) were kept constant between all experiments and subexperiments. Each experiment consisted of ten distinct runs. Each run was terminated after 200 generations.

15.5.1 SLAVIC CRIB

What we label as "Slavic crib" is a plain-text list of feminine names which we had compiled from multiple sources publicly available on the Internet. Principal sources of names were websites of western Slavic origin. This choice was motivated by following reasons:

1. The oldest more or less certain trace of VM's trajectory points to the city of Prague - the center of western Slavic culture.

2. Orthography of western Slavic languages relatively faithfully represent the pronunciation. That is, there are relatively few digraphs (e.g. a bi-gram "ch" which denotes a voiced velar fricative). Hence, the distance between the graphemic and the phonemic representations is not so huge as in case of English or french.

3. Slavic languages have rich but regular affective and diminutive morphology which is often used when addressing or denoting beloved persons by their first name.

The third reason is worth to be introduced somewhat further: in both Slavic and western Slavic languages, a simple in-fixing of the unvoiced velar occlusive "k" before the terminal vowel "a" of a feminine names leads to creation of a diminutive form of such a name (e.g. $alena \rightarrow alenka, helena \rightarrow helenka$ etc.) The fact that this morphological rule is used both by western as well as eastern Slavs indicates that the rule itself can be quite old, date to *common Slavic* or even *pre-Slavic* periods and hence, was quite probably in action already in the period when VM was written.

For the purpose of this article, let's just note that application of the substitution:

$$a\$ \rightarrow ka/$$

allowed us to significantly increase the extent of the "Slavic crib". Thus, we have obtained a list a of 13815 distinct word types which are in quite close relation to phonetic representation of feminine names used in Europe and beyond[8]. The alphabet of this crib comprises of 38 symbols, hence there exists 19^{39} possible ways how symbols of the Calendar could be replaced by symbols of this crib.

Figure 27 shows the process of convergence from populations of randomly generated chromosomes towards more optimal states. In case of runs averaged in the "SUBSTITUTON" curve, the procedure $P_{cribbing}$ consisted in simple mapping of the Calendar onto the crib by means of a substitution cipher specified in the chromosome. But in case of runs averaged in the "REVERSAL + SUBSTITUTION" curve, whole process was initiated by the reversal of order of characters present within individual tokens of the Calendar (e.g. $okedy \rightarrow ydeko, otedy \rightarrow ydeto$ etc.) Let's now look at contents of individuals which were "identified" by the optimization method.

More concrete illustrations can also turn out to be quite illuminating. Hence, if the most elite individual of run 1 (i.e. the one with fitness 197) is as a means of substitution of EVA characters contained in the Calendar, one will see appearance of names like ALENA, ALETHE,

8 Slavic crib is publicly available at http://wizzion.com/thesis/simulationo/slavic_extended.crib

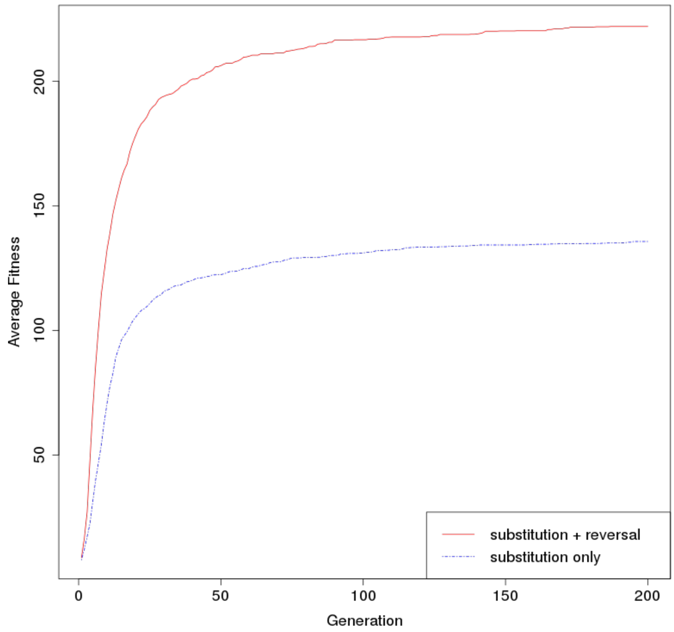


Figure 27: Evolution of individuals adapting label in the Calendar to names listed in the Slavic crib.

ANNA, ATENKA, HANKA, HELENA, LENA etc. And when the last one (i.e. the one with fitness 240 is used), the resulting list shall contain tokens like AELLA, ALANA, ALINA, ANKA, ANISSA, ARIANNKA, ELLINA, IANKA, ILIJA, INNA, LILIJA, LILIKA, LINA, MILANA, MILINA, RANKA, RINA, TINA etc.

This being said, the observation that all reversal-implementing runs have converged to genomes which:

1. transcribe e in EVA as nasal n
2. transcribe k in EVA as velar k
3. transcribe t in EVA as nasal n
4. transcribe y in EVA as vowel a
5. transcribe a in EVA as vowel (80% times as "i", 10% as "e", 10% as "o")
6. transcribe l in EVA as either a liquid consonant (80% "l", 10% "r") or "m" (10%)

...could also be of certain use and importance.

15.5.2 HEBREW CRIB

At this point, a skeptical mind could start to object that what our algorithm adapt to is in fact not the Calendar, but the statistical properties

Fitness																			
197	e	s	t	n	h	k	a	h	k	l	h	t	a	k	a	m	e	n	a
230	i	k	t	n	s	k	n	h	k	l	z	t	a	j	s	m	i	n	a
224	i	c	t	n	v	k	/	g	k	l	m	b	a	j	/	r	i	n	a
227	i		t	n	p	a	f	l	k	l	m	e	a	n	k	r	i	n	a
240	i	k	t	n	a	k	f	l	k	l	m	e	a	j	g	r	i	n	a
226	i		l	n	h	o		l	k	r	g	e	a	n	a	m	i	n	a
208	i	q	g	n	x	k	d	e	k	l	m	x	a	j	x	r	i	n	a
239	i	k	t	n	d	o	l	l	k	l	f	e	a	k	i	m	i	n	a
191	o	t	l	n	t	n	n	r	k	m	z	b	a	n	h	r	e	n	a
240	i	s	t	n	s	k	n	l	k	l	m	e	a	j	I	r	i	n	a
EVA	a	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	y

Table 40: Fittest chromosomes which map reversed tokens in the Calendar onto names of the Slavic crib

of the crib. And in case of such a long and sometimes somewhat artificial list like Crib_{Slavic}, such an objection would be in great extent justified. For the adaptive tendencies of our evolutionary strategy are indeed so strong that it would indeed find a way to partially adapt the calendar to a crib which is long enough[9]

For this reason, we have decided to target our second experiment not at the biggest possible crib but rather at the oldest possible crib. And given that our first experiment has indicated that it seems to be more plausible to interpret labels in the Calendar as if they were written in reverse, id est from right to left, our interest was gradually attracted by Hebrew language[10]. This lead us to two lists of names:

- $\text{Crib}_{Hebrew-men}$ contains 555 masculine names[11]
- $\text{Crib}_{Hebrew-women}$ contains 283 feminine names[12]

both lists were extracted from the website finejudaica.com/pages/hebrew_names.htm and were chosen because they did not contain any diacritics and

9 This has been, indeed, shown by multiple micro-experiments which we do not report here due to the lack of space. No matter whether we use cribs as absurd as list of modern American names or Enochian of John Dee and Edward Kelly, we could always observe a sort of adaptation marked by the increase of fitness. But it was never so salient as in case of Crib_{Slavic} or Crib_{Hebrew}.

10 Other reasons why we decided to focus on Hebrew include: important presence of Jewish diaspora in Prague of Rudolph the 2nd (c.f. the story of rabbi Loew and the Golem of Prague); ritual bathing of Jewish women known as mikveh; usage of VM-resembling triplicated forms (e.g. amen, amen, amen) in Talmudic texts; attested existence of so-called Knaanic language which seems to be principally a Czech language written in Hebrew script et caetera et caetera.

11 http://wizzion.com/thesis/simulation0/jewish_men

12 http://wizzion.com/thesis/simulation0/jewish_women

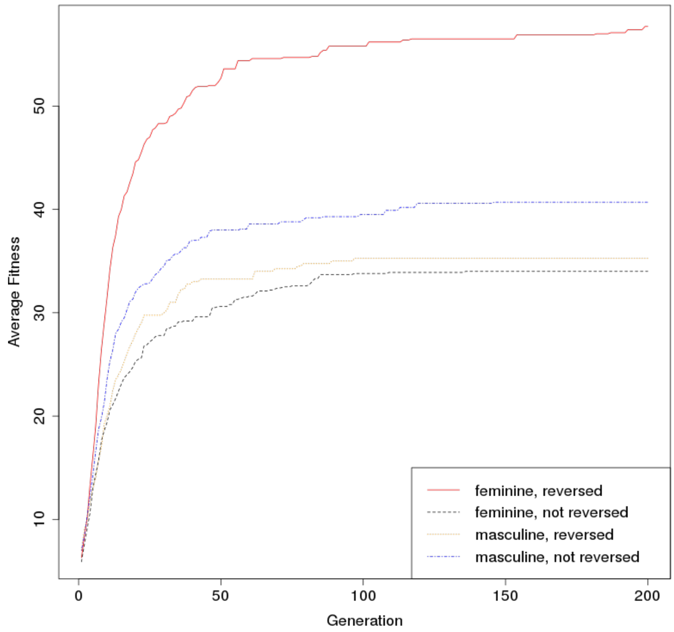


Figure 28: Evolution of individuals adapting label in the Calendar to names listed in the Hebrew cribs.

hence transcribing Hebrew names in a similar way as they had been transcribed millenia ago.

Figure 28 displays the summary of all runs which aimed to transcribe the Calendar with Hebrew names.

As may be seen, the whole system converged to highest fitness values when $Crib_{Hebrew-women}$ was used in concordance with reversal of order of characters. In such scenario, minimal attained fitness was attained by run converging to $F_{min(hebrew28,283,hfr)} = 52$, maximal attained fitness was $F_{max(hebrew28,283,hfr)} = 63$. Difference results of hebrew, reverse batch of runs and other results of other batches is statistically significant (Welch Two Sample t-test, p-value $<$ 7e-10).

Subsequently, a list of 283 was tokens randomly generated in a way that the distribution of lenghts of randomly generated sequences is identic to distribution of lenghts of names in the hebrew crib. Maximal attained fitness was $F_{max(random28,283,hfr)} = 26$ among 10 runs aiming to adapt the Calendar to such a random crib. Statistical difference between results of batch of runs adapting to valid character-reversed hebrew crib $hebrew28, 283, hfr$ and the equidistributed randomly generated crib $random28, 283, hfr$ turned out to be strongly significative (Welch two sample two sided non-paired t-test: t = 22.0261, df = 15.442, p-value = 4.384e-13).

The highest attained fitness value was was attained by the cribbing procedure which first reverses the order of characters whose EVA representations are subsequently substituted by a following chromosome:

A	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	Y
י	ה	ד	י	נ	ע	נ	ר	נ	ל	ב	ב	ה	צ	מ	ד	ת	ל	ג

This chromosome transcribes the voynichese Calendar labels *okam, otainy, otey, oty, otaly, okaly, oky, okyd, ched, otald, orara, otal, salal and opalg* to feminine Hebrew names

בינה גברילה גבורה גילה גלה גלילה גלינה גנה דגנה דינה דלילה ידידה לילה לילית עליצה

(i.e. Bina, Gabriela, Ghila, Gala, Galila, Galina, Gina, Degana, Diyna, Deliyla, Yedidya, Lila, Lilit and Alica).

Worth mentioning are also some other phenomena related to these transcriptions. One can observe, for example, that the label "otaly" - translated as Galina - is also present on folios f33v, f34r or f46v which all contain drawings of torch-like plants. This is encouraging because the word "galina" is not only a Hebrew name, but also a substantive meaning "torch". Similarly, the word "lilit" is not only a name but also means "of the night". This word supposedly translates the voynichese token "salal" which is very rare - asides the Calendar it occurs only on purely textual folio f58v and on a folio f67v2 which, surprise!, may well depict circadian rhythms of sunrise, sunset, day and night.

Or it could be pointed out kind that the huge majority of occurrences of voynichese trigram "oky" (potentially denoting the name "gina" which also means "garden") is to be observed on herbal folios. Or the distribution of instances of "okam" (transcribed as "bina" which means "intelligence and wisdom"[13] could, and potentially should, be taken into consideration. Or maybe not.

15.6 CONCLUSION

In 2013, BBC Online had announced "Breakthrough over 600-year-old mystery manuscript". The breakthrough was to be effectuated by Stephan Bax who, in his article, describes the process of deciphering as follows:

« **?**» (**?**)

What Bax does not add, unfortunately, is that the voynich crossword puzzle is so big that anyone who looks at it close enough can find in it small islands of order, local optima where few characters seem to fit the global pattern. Thus, even if Bax had succeeded, as he states, in "identification of a set of proper names in the Voynich text, giving a total of ten words made up of fourteen of the Voynich symbols and clusters", this would mean nothing else than that he had identified a locally optimal transcription alphabet.

13 Note that "bina" is one among highest sephirots located at north-western corner of kabbalistic tree of life. In this context it is worth noting that only partially readable EVA group "...kam" occurs as a third word near the north-western "rosette" of folio 85v2. Such considerations, however, bring us too far.

In this article, we have presented two experiments employing two different lists of feminine names. Both experiments have indicated that if labels in the Zodiac encode feminine names, then these have been originally written from right to left [14]. The first experiment led to identification of multiple substitution alphabets which allow to map 240 EVA letters, contained in 40 distinct words present in the Calendar, onto 35 feminine-name-resembling sequences enumerated among 13815 items of $\texttt{Crib}_{Slavic}$. Results of second experiment indicate that if ever the Calendar contains lists of Hebrew names, then these names would be more probably feminine rather than masculine.

This is, as far as we can currently say, all that could be potentially offered as an answer to the question « Can Evolutionary Computation Help us to Crib the Voynich Manuscript?» (Hromada, 2016a). Everything else is - without help coming from experts in other disciplines - just a speculation.

15.7 GENERIC CONCLUSION

Looked upon from a superficial point of view, an article presented in this "zeroth analysis" contains nothing else and nothing more than:

1. a very brief description of a particular enigma commonly known as "Voynich Manuscript"
2. introduction of a so-called "primary mapping" hypothesis potentially able to direct any future tentative to decipher the manuscript
3. discussion of inner workings of an "evolutionary algorithm" programmed whose source code is hereby transferred to the public domain[15]
4. presentation of *fairly reasonable* results obtained after confrontation of the manuscript with the algorithm which takes lists of Slavic and Hebrew names at its input

What is meant by the attribute *fairly reasonable* is, of course, a place for argument. And contrary to legions of other researchers, we do not pretend that we have succeeded to "crack" the manuscript. We simply state that after being executed on a single core of 1.8GHz CPU, a simple 160-line script written in pure PERL can yield, in just few hours, intelligible transcriptions of "*lattices of terms*" contained in a previously unknown corpus. Thanks to a fairly trivial derivative of

14 Note, however, that this does not necessarily imply that the scribe of VM (him|her)self had written the manuscript in right-to-left fashion. For example, in case (s)he was just reproducing an older source which (s)he didn't understand, his|her hand could trace movements from left to right while the very original had been written from right to left

15 http://wizzion.com/thesis/simulation0/voynich.PERL

a Canonical Genetic Algorithm, an average home PC can closely approximate a brute-force search which would otherwise run weeks (at least) even when executed at state-of-the-art computational clusters.

Simply stated, our 0th simulation indicates that, which has already been indicated many times before:

Evolution narrows-down the search to regions where most plausible hypotheses reside.

Non-negligible speed-up goes hand in hand with such narrowing-down. And it is evident that such speed-up can be useful for any system which can invest only limited amount of time and energy into its search of the most optimal hypothesis. It does not really matter whether the system in which we speak in this context is a PERL script, child's mind or the Nature herself: problem-solving system which implements evolutionary principles tends to converge (Rudolph, 1994) to "the answer" in less time, and with less resources wasted, than the system which does not implement such principles.

At least as fascinating as her ability to speed things up is evolution's propensity to produce adaptations. Zeroth simulation is particularly instructive in this regards: as noted in the footnote 9, the VM-to-crib transcribing EA produced certain results even in cases when cribs as "list of 20th century American names" have been used as target dictionaries. In spite of absurdity of such cribs - for it is indeed highly improbable that VM initially contained names like Butch or Mitch - the EA succeed to discover certain inherent similarities between two texts in order to exploit them in the future search.

Thus, the main conclusion of 0th simulation can be stated as follows:

Evolution is able to facilitate the search for optimal mapping between distinct corpora encoded in distinct forms of representation.

In this simulation, distinct forms of representation has been a so-called EVA alphabet (into which VM is transcribed) and phonemic alphabets common to Slavic or Hebrew languages. Mapping itself was nothing else than simple substitution of one symbol from one alphabet with one symbol from another alphabet. A mapping - a hypothesis - was considered "the fittest" if it succeeded to transcribe initial unintelligible EVA corpus to intelligible list of names. Both EVA corpus and the name list were EA's inputs and thus in certain sense "innate" to each individual run of the algorithm.

What was "acquired" during the process was the set of mono-alphabetic substitution rules. EA presented in 0th simulations is thus an example of evolution which processes strictly "symbolic" representations.

This will not be the case in simulations which are now to follow: let's now descend to the realm of sub-symbolic (vectorial) entities in order to propose an evolutionary solution to the problem of category induction.

16 EVOLUTIONARY LOCALIZATION OF SEMANTIC PROTOTYPES

16.1 GENERIC INTRODUCTION

How does a child create mappings between "signifiers" and "signifieds" (de Saussure, 1916), between words and their meanings? How do concepts emerge in the mind of a child?

These question are addressed on many places of Conceptual Foundations. Be it during our discussion of "ontogeny of lexicon and semantics" (P+72-78) or *classical* theories thereof (P+93-95), be it during the definition of "category prototype" (P+132) or in the Hebb/Harris analogy (P+133) suggesting *a sort of equivalence* between Hebb's law well-known to neuroscientists and so-called "distributional hypothesis" well-known to linguists, it has been indicated on multiple places that what contemporary linguists label as "vocabulary development" is, in its essence, nothing else than a usage-based, goal-oriented, associanist process. And that Chomsky's critic of Skinner (P+95), in regards to acquisition of meanings, quite inappropriate: in fact it does not even apply. This is so because first syntactic representations (P+173-179) are acquired, tuned and perfectioned later than first semantic constructions (P+179-184).

And how could such "vocabulary development" be simulated by an engineer willing to do so ?

In an ideal world, such an engineer would have to have, at least, two things at his disposition:

- a corpus C representing the world of a modal toddler: it should contain representations of objects with many attributes (some of them could and should mutually overlap)
- an algorithm A capable of clustering objects into categories in a "cognitively plausible" (P+13) way (i.e. similar to the way child's mind does it)

Unfortunately, as far as 2016, no such C is available, at least not in textual form which could be processed by means of methods commonly used in computational linguistics (P+112-164). The corpus CHILDES (P+207-209) is as close as one can get to C but, and this is a non-negligible "but", CHILDES contains transcripts representing interactions within certain worlds BUT does not contain descriptive representations of these worlds *selves*. And as we have noted elsewhere (Hromada and Gaudiello, 2014) construction of such corpus surpasses by far possibilities of any individual engineer and thus also possibilities of this dissertation.

Willing to develop A but without proper C, one is obliged to approximate. In regards to simulations of induction of meaning, a plausible approximation could be proposed as follows:

Let's suppose that text documents are "objects" and that groups of objects which have similar semantic content (i.e. refer to or speak about similar things) delimit a certain "semantic category".

Under such supposition - and under such supposition only - can one reduce the problem of vocabulary development to a problem of multi-class categorization of documents. Under such *ceteris paribus* - and under such *ceteris paribus* only - can one pretend that the model first published in the article « Genetic Optimization of Semantic Prototypes for Multi-class Document Categorization» (Hromada, 2015) could , in the long run, potentially lead to full-fledged computational models of vocabulary development.

16.2 INTRODUCTION

In computational theories and models learning, one generally works with two types of models: regression and classification. While in regression models one maps continuous input domain onto continuous output range, in models of classification, one aims to find mappings able to project input objects onto a finite set of discrete output categories.

This article introduces a novel means of construction of a particular type of the latter kind of learning models. Due to finite and discrete nature of its output range, classification - also called categorization by more cognition-oriented researchers - seems to be of utmost importance in any cognitively plausible (Hromada, 2014b) model of learning. But under these terms, two distinct meanings are confounded and the term categorization thus often represents both:

1. process of learning (e.g. inducing) of categories
2. process of retrieving information from already learned (induced) categories

which crudely correspond to training, resp. testing phases of supervised learning algorithms.

In the rest of this section we shall more closely introduce an approach combining notions of category prototype, dimensionality reduction and evolutionary computing in order to yield a potentially "cognitively plausible" means of supervised machine learning of a multi-class classifier. We shall subsequently present specificities of a Natural Language Processing (NLP) simulation which was executed in order to assess the feasibility of our approach. Results hence obtained shall be subsequently compared with comparable "deep learning" semantic hashing technique of (Salakhutdinov and Hinton, 2009). The article shall be concluded with few remarks integrating whole research into more generic theories of neural and universal Darwinism.

16.2.1 GEOMETRIZATION OF CATEGORIES

In contemporary cognitive science, categories are often understood as entities embedded in an Δ-dimensional feature space (Gärdenfors, 2004). The most fundamental advantage of such models, whose computer sciences counterparts are so-called "vector symbolic architectures" (VSAs) (Widdows and Cohen, 2014), is their ability to geometrize one's data, i.e. to represent one's data-set in a form which allows to measure distances (similarities) between individual items of the data-set.

Thus, even entities like "word meanings" or "concepts" can be geometrically represented, either as points, vectors or sub-spaces of the enveloping vector space S. One can subsequently measure distances between such representations, e.g. distance of the meaning of the word "dog" from the meaning of "wolf" or "cat" etc. Geometrization of one's data-set once effectuated, space S can be subsequently partitioned into a set R of $|C|$ regions $R = R_1, R_2, ..., R_{|C|}$.

In unsupervised scenario, such partitioning is often done by means of diverse clustering algorithms, the most canonical among which being the k-means algorithm (MacQueen et al., 1967). Such clustering mechanisms often characterize candidate cluster C_X in terms of a geometric centroid of the members of the cluster. Feasibility of a certain partition is subsequently assessed in terms of "internal clustering criteria" which often take into account distances among such centroids.

In the rest of this article, however, we shall aim to computationally implement a supervised learning scenario and instead of working with the notion of category's geometric centroid, our algorithm shall be based upon the notion of category's prototype. The notion of the prototype was introduced into science notably by theory of categorization of Eleanore Rosch which departed from the theoretical postulate that:

> "*the task of category systems is to provide maximum information with the least cognitive effort*" (Rosch, 1999)

In seminal psychological and anthropological studies which have followed, Rosch have realized that people often characterize categories in terms of one of their most salient members. Thus, a prototype of category C_X can be most trivially understood as such a member of C_X which is the most prominent, salient member of C_X. For example "apples" are prototypes of category "fruit" and "roses" are prototypes of category "flowers" in western cultural context.

But studies of Rosch had also suggested another, more mathematical, notion of how prototypes can be formalized and represented. A notion which is based upon the notion of closeness (e.g. "distance") in a certain metric space:

"*items rated more prototypical of the category were more closely related to other members of the category and less closely related to members of other categories than were items rated less prototypical of a category*" (Rosch and Mervis, 1975)

Given that this notion is essentially geometric, the problem of discovery of a set of prototypes can be potentially operationalized as a problem of minimization of a certain fitness function. The fitness function, as well as means how it can be optimized, shall be furnished in section 2. But before doing so, let's first introduce certain computational tricks which allow to reduce the computational cost of such search of the most optimal constellation of prototypes.

16.2.2 RADICAL DIMENSIONALITY REDUCTION

There is potentially an infinite number of ways how a data-set D consisting of $|D|$ documents can be geometrized into a $\Delta-$dimensional space S. In NLP, for example, one often looks for occurrences of diverse words in the documents of the data-set (e.g. corpus). Given that there are $|W|$ distinct words occurring in $|N|$ documents of the corpus, one used to geometrize the corpus by means of a N * M co-occurrence matrix M whose X-th row vector represents the X-th document N_X, Y-th column vector represents the Y-th word W_Y and the element on position $M_{X,Y}$ represents the number of times W_Y occurred in N_X.

Given the sparsity of such co-occurrence matrices as well as for other reasons, such bag-of-words models are more or less abandoned in contemporary NLP practice for sake of more dense representations, whereby the dimensionality of the resulting space, d, is much less than $|W|$, $d \ll |W|$. Renowned methods like Latent Semantic Analysis (LSA) (Landauer and Dumais, 1997) set aside because of their computational cost, we shall use the Light Stochastic Binarization (LSB) (Hromada, 2014c) algorithm to perform the most radical dimensionality-reducing geometrization possible.

LSB is an algorithm issued from the family of algorithms based on so-called random projection (RP). Validity and feasibility of all these algorithms, be it Random Indexing (RI, (Sahlgren, 2005)) or Reflective Random Indexing (RRI,(Cohen et al., 2010)) is theoretically founded on a so-called lemma of Johnson-Lindenstrauss, whose corollary states that "*if we project points in a vector space into a randomly selected subspace of sufficiently high dimensionality, the distances between the points are approximately preserved*" (Sahlgren, 2005).

Methods of application of this lemma in concrete NLP scenarios being described in references above, we precise that LSB can be labeled as "most radical" variant of RP-based algorithms because:

- it tends to construct spaces with as small dimensionality as possible (in LSB, $d < 300$; in RI or RRI models, $d > 300$)
- LSB tends to project the data onto binary and not real or complex spaces

It can be, of course, the case that such dimensionality-reduction and binarization can lead to certain decrease of discriminative accuracy of LSB-produced spaces. On the other hand, given that dimensionality reduction and binarization necessary bring about reduction of computational complexity of any subsequent algorithm which could be used to explore the resulting space S, such decrease of accuracy is to be more swiftly counteracted by subsequent optimization. The goal of this study is to explore whether such *post hoc* optimization of classifiers operating within dense, binary, LSB-produced spaces is possible, and whether the combination of the two can be used as a novel means of machine learning.

But before describing in more closer such evolutionary optimizations, let's precise that because of its low-dimensional and binary nature, LSB can also be understood as yielding a sort of "hashing function" aiming to attribute similar hashes to similar documents and different hashes to different documents. In this sense, LSB is similar to approaches like Locality Sensitive Hashing (LSH, Datar et al. (2004)) or Semantic Hashing (SH, Salakhutdinov and Hinton (2009)) often used, or at least presented, as *the* solution of multi-class classification of Big-Data corpora. It is with the results of the latter, "deep-learning" approach, that we shall compare our own results in section 16.5.

16.3 GENETIC LOCALIZATION OF SEMANTIC PROTOTYPES

Let $D = \{d_1, ..., d_{|D|}\}$ be a training data-set consisting of $|D|$ documents to which the training dataset attributes one among $|L|$ corresponding members of set of class labels $L = \{L_1, ..., L_{|L|}\}$.

Let Γ denote a tuple $\Gamma = C_1, ..., C_{|L|}$ whose individual elements are sets containing indices of members of D to which a same label L_l is attributed in the training corpus (e.g. $C_1 = \{3, 4, 5\}$ if training corpus attributes its 1st label only to documents d_3, d_4 and d_5).

Let $H = \{h_1, ..., h_{|D|}\}$ be a set of Δ-dimensional binary vectors attributed to members of D by a hashing function F_H, i.e. $h_X = F_H(d_X)$.

Let S be a Δ-dimensional binary (Hamming) space into which members of H were projected by application of mapping F_H.

Then a classificatory pertinence F_{CP} of the candidate prototype P_K of K-th class ($K \leqslant |C|$) can be calculated as follows:

$$F_{CP}(P_K) = \alpha \sum_{t \in C_K} F_{hd}(h_t, P_K) - \omega \sum_{f \not\subset C_K} F_{hd}(h_f, P_K) \quad (5)$$

whereby P denotes the position of the prototype in S, F_{hd} denotes the Hamming distance [1], h_t denotes the hash "true" document belonging to same class as the prototype, h_f is the vector of the "false" document belonging to some other class of the training corpus and α and ω are weighting parameters.

In simpler terms, an ideal prototype of category C is as close as possible to members of C and as far away as possible from members of other categories.

Given such a definition of an ideal prototype, an ideal |C|-class classifier I can be trained by searching for such a set $P = \{P_1, ..., P_{|L|}\}$ of individual prototypes, which minimize their overall classification pertinence:

$$I = \min \sum_{K=0}^{K=|L|} F_{CP}(P_K) \quad (6)$$

In simpler terms, an ideal |C|-class classifier I is composed of |C| individual prototypes which are as close as possible to documents of their respective categories, and as far away as possible from all other documents.

Equations 1 and 2 taken together, one obtains a fitness function which can be optimized by evolutionary computing algorithms. And given that one explores the prototypical constellations embedded in a binary space, one can use canonical genetic algorithms (CGAs, Goldberg (1990)) for the optimization of the problem of discovery of ideal constellation of most pertinent prototypes. We choose CGAs for three principal reasons:

Primo, we choose CGAs mainly for their property, proven in Rudolph (1994), to converge to global optimum in finite time if ever they are endowed with the best-individual protecting, elitist strategy. Secundo, one can obtain practically useful and exploitable increase in speed simply due to the fact that CGAs are conceived to process binary vectors and do so on CPUs which are essentially built for processing such vectors. Tertio, CGAs offer a canonical, well-defined, "baseline" gateway to much more sophisticated evolutionary computing (EC) techniques and are well understood by both neophytes as well as the most experts of the EC community.

For this reason, we consider as superfluous to describe in closer detail the inner workings of a CGA: instead, references (Goldberg, 1990; Rudolph, 1994) are to be followed and read. Given that the particular values of mutation and cross-over parameters shall be specified

1 Hamming distance of two binary vectors h_1 and h_2 is the smallest number of bits of h_1 which one has to flip in order to obtain h_2. It is equivalent to a number of non-zero bits in a $XOR(h_1, h_2)$ binary vector.

in the following section, the only thing which in which the reader now needs to be reassured is her correct understanding of the nature of data structures which the algorithm hereby proposed shall implement, in order to encode an individual |C|-class classifier:

Given that equation 1 defines a prototype candidate as a position in Δ-dimensional Hamming space and given that equation 2 stipulates that an ideal |C|-class classifier is to be composed of representations of |C| ideal prototype candidates, the data structure representing an individual solution can be constructed by a simple concatenation of |C| Δ-dimensional vectors. Thus, the individual members of the populations which the CGA shall optimize are, *in essentia*, nothing else than binary strings of length |C|*Δ.

16.4 CORPUS AND TRAINING PARAMETERS

In order to be able to compare the performance of our algorithm with non-optimized LSB and SH, same corpus and dimensionality parameters were chosen as those, which are already reported in the previous studies (Salakhutdinov and Hinton, 2009; Hromada, 2014c). Thus, dimensionality of the resulting binary hashes was Δ=128. Every document of the corpus was hence attributed a 16-byte long hash.

A so-called "20newsgroups" corpus[2] has been used. The corpus contains 18,845 postings taken from the Usenet newsgroup collection divided into training set containing 11,314 postings, 7531 being the testing set ($|D_{training}| = 11313, |D_{testing}| = 7531$). Both training and testing subsets are divided into 20 different newsgroups which correspond each to a distinct topic. Given that every distinct topic represents a distinct category label, $|C| = 20$.

Documents of the corpus were subjected to a very trivial form of pre-processing: documents were split into word-tokens by means of [\^w] separator. Stop-words contained in PERL library Lingua::StopWords were subsequently discarded. 3000 word types with highest "inverse document frequency" value were used as initial terms to which the initial random indexing iteration attributed 4 non-zero values. Hashing function $F_H = LSB(\Delta = 128, Seed = 3, Iterations = 2)$ because there were 2 "reflective" iterations preceding the ultimate stage of "binarization".

Once hashes were attributed to all documents of the corpus, the Hamming space S was considered as constructed and stayed unmodified during all phases of subsequent optimizations and evaluations. As CGA-compliant algorithm, the optimization applied generated the new generation by crossing over two parent solutions chosen by the fitness proportionate (e.g. roulette wheel) selection operator. Each among 2560 (128*20) genes was subsequently mutated (i.e. a correspondent bit was flipped to its opposite value) with probability of

2 http://qwone.com/ jason/20Newsgroups/

0.1%. Population contain 200 individuals, zeroth generation was randomly generated. Elitist strategy was implemented so that all individuals with equally best fitness survived intact the transition to future generation. Parameters α and ω (e.g. equation 1) used in fitness estimation were both set to 1.

Information concerning the category labels guided the optimization during the training phase. During the testing phase, such information was used only for evaluation purposes. Multiple independent runs were executed and values of precision and recall were averaged among the runs in order to reduce the impact of stochastic factors upon the final results.

16.5 EVALUATION AND RESULTS

Every 250th generation, classificatory accuracy of an individual solution with minimal overall classification pertinence (c.f. equation 2) was evaluated in regards to 7531 documents contained in the testing part of the corpus. Following aspects of classifier's performance were evaluated in order to allow comparison with the results with Precision-Recall curves presented in (Salakhutdinov and Hinton, 2009; Hromada, 2014c):

$$\text{Precision} = \frac{\text{Number of retrieved relevant documents}}{\text{Total number of retrieved documents}}$$

$$\text{Recall} = \frac{\text{Number of retrieved relevant documents}}{|D_{testing}|}$$

The notion of relevancy is straightforward: an arbitrary document D_T contained in the testing corpus is considered to be relevant to query document D_Q if and only if they were both labeled with the same category label, $L_Q = L_T$.

On the other hand, the correct understanding of what is meant by "retrieved" is the key to correct understanding of the core idea behind the functionality of the algorithm hereby proposed. That is: **the prototypes induced by the CGA optimization are to be used as retrieval filters**.

We precise: given a hash h_Q of a query document d_Q, one can easily identify - among $|C|$ prototypes encoded as components of an quasi-ideal constellation I furnished by the CGA - such a prototype P_N which is nearest to h_Q. Subsequently, each among N documents whose hashes are N nearest neighbors of the prototype P_N, should be considered as retrieved by d_Q. Prototypes discovered during the training phase therefore primarily specify, during the testing phase, which documents are to be considered as retrieved, and which not. For all LSB curves present on Figure 29, the size of such retrieval neighborhood was set to N=2000.

Also, in order to obtain viable precision-recall curves, Radius R=(0, ..., $\Delta = 128$) of the Hamming ball was used as a trade-off parameter.

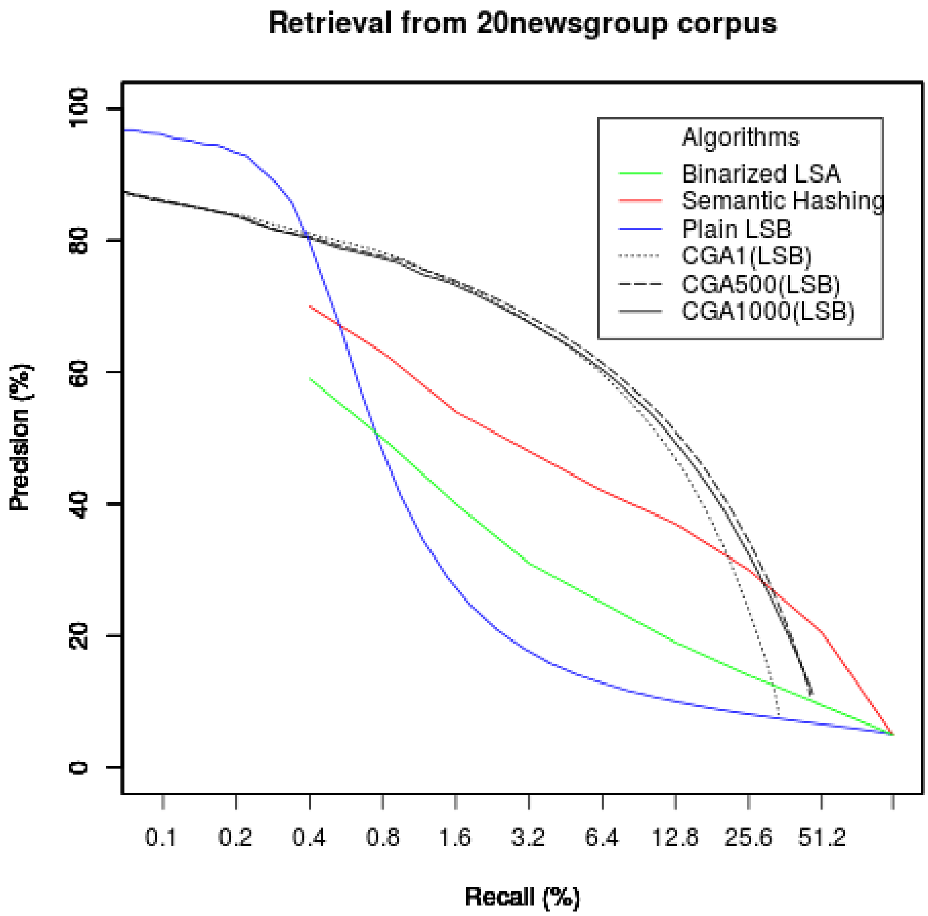


Figure 29: Retrieval and 20-class classification performance in 128-dimensional binary spaces. Non-LSB results are reproduced from Figure 6 of study (Salakhutdinov and Hinton, 2009), plain LSB from (Hromada, 2014c).

For every data-point of the plot on Fig. 1, h_N was considered as retrieved by query h_Q only if the hamming distance of query and the candidate document was smaller than R ($hd(h_Q, h_N) > R$). Points on the very left of the plot correspond thus correspond to R=0 (i.e. h_Q and h_N collide), while points on the right correspond to R=128 (i.e. h_Q does not have a single bit in common with h_N).

As comparison of curves on the figure indicates, *biggest increase in performance is attained by decision to use prototypes as retrieval filters*. Thus, when one uses the most fit among 200 randomly chosen prototype constellations as a retrieval filter (c.f. curve CGA1(LSB)), one obtains significantly better results than when does not use any prototypes at all (c.f. curve "Plain LSB"). If the process is followed by further genetic optimization (c.f. CGA500 for situation after 500 generations), one observes a non-negligible increase of precision in the high recall region of the spectrum. But it can also be seen that the optimization has its limits, hence there is a slight decrease between 500th and 1000th generation which potentially corresponds to situation whereby the induced prototype constellation tends to over-fit the training data-set. This leads to subsequent decrease in overall accuracy of classification of documents contained in the testing data-set.

Figure 29 also suggests that the genetic discovery of sets of prototypes - and their corresponding use as retrieval filters - seems to produce results which are better than those produced by both binarized Latent Semantic Analysis or SH. Exception to this is SH's 20% precision at recall level of 51.2%. Note, however, that since on page 6 of their article, Salakhutdinov and Hinton (2009) claim to have used their hashes as retrieval filters of neighborhood of size N=100, and given that the every size of the category in a 20newsgroup corpus $\approx$ 390 documents, such a result is not even theoretically possible. This is so because even in case the classifying system would retrieve only the relevant documents (i.e. precision would be 100%) the maximal attainable recall would still be just $100/390 \approx 25.6\%$. Both authors were contacted by mail with a request to rectify possible misunderstanding. Unfortunately, none of them replied.

16.6 CONCLUSION

Results hereby presented indicate that supervised localization of constellations of semantic prototypes can significantly increase accuracy of classifiers which use such constellations as retrieval filters.

Given that the localization of such constellations is governed by the training corpus but the increase is also significant in case when one confronts the system with previously unseen testing corpus, we are allowed to state that **our algorithm is capable of generalization**. This was principally attained by combination of following ideas:

1. projection of documents into low-dimensional binary space

2. definition of fitness of prototype in terms of distances to both documents of its category, as well as distance to document of other categories

3. search for fittest prototype constellations

4. use of the most fit prototype constellation as a sort of retrieval filter

In spite of its generalizing and thus "machine learning" capabilities, our algorithms is essentially a non-connectionist one. Thus, instead of introducing synapses between neurons, or speaking about edges between nodes of the graph, briefly, instead of speaking about *deep learning of multi-layer encoders of stacks of Restricted Boltzmann Machines fine-tuned by back-propagation* as (Salakhutdinov and Hinton, 2009) do - we have found as more preferable to reason in geometric and evolutionary terms. It is indeed due to this "geometric" perspective that the computational complexity of the algorithm is fairly low: $\Delta|D||C|$ for evaluation of fitness of one individual prototype constellation. In future study, we aim to explore the performance of slightly modified fitness function whose complexity $\Delta|D| + |C|^2$ could be of particular interest in cases of huge data-sets (i.e. big $|D|$) with fairly limited number of classes ($|C|$).

In practical terms, it is also advantageous that both fitness function evaluation as well as final retrieval assess distances in terms of binary hamming distance measure. In both cases, one can use basic logical operations like XOR + some basic assembler instructions which would furnish indices allowing to execute sort of "conceptual geometry" with particular swift and ease. Given these properties + the fact that hashes which are manipulated are fairly small (in one gigabyte of memory, one can store hashes for 8 million documents), one can easily predict existence of future application-specific integrated circuit (ASIC) potentially executing billions query2document comparisons per second.

Computational aspects aside, our primary motive in developing the algorithm hereby proposed was to furnish a sort of cognitively plausible (Hromada, 2014b) "experimental proof" for our doctoral Thesis which postulates that a sort of evolutionary process exists not only in the realm of biological species, but also in realms populated by "species" of a completely different kind.

Id est, in realms of linguistic structures and categories, in realms of word meanings, concepts and, who knows, maybe even in the realm of mind itself.

Being uncertain about whether our demonstrate, with sufficient clarity, that it is reasonable to postulate not only neural (Edelman, 1987), but also intramental evolutionary processes, we conclude by saying that the formula hereby introduced offers a simple yet reasonably accurate method of solving the problem of multi-class categorization of texts.

16.7 GENERIC CONCLUSION

Speaking less concretely, this article shows that model, implementing evolutionary search within a certain type of vector space, can bring practically applicable results. Given that results obtained with training data are in non-negligible extent transposable to testing data, one can consider such model to instantiate a particular case of machine learning (P+125-130). Training data-set is labeled and labels are exploited to direct the evolutionary search: hence, the algorithm can be understood as a supervised one.

Concretely speaking, this article shows how one can perform multi-class (N=20) classification of textual documents. Hence, newspaper articles were considered as entities which are to be classified and occurrence frequencies of words contained within the articles are used as features by means of which the articles are characterized.

And speaking less concretely again, this chapter indicates that evolutionary computation can provide the means to identify constellations of regions in a semantic space which roughly correspond to constellations of semantic categories [3]. Ideally, the process converges to state where correct category labels are attributed to correct regions with correct extension.

It is in this sense that the approach hereby introduced can be, mutatis mutandi, understood as a potential model of vocabulary development within individual child. This is so because the aim of vocabulary ontogeny is analogical: one aspires to attribute correct phonic representations ("words", "signifiers", "labels") to correct regions of the conceptual space. As has been observed by other researchers (P+173) or illustrated by the Borgesian *Ding-Dong Mystery* (P+177-179) such process of attribution *appropriate handel to appropriate vessels* is far from being a monotonic descent to most optimal state.

Rather, the process of acquisition of vocabulary is full of periods where the category is either too exhaustive or too specific, full of small adjustments, detours and returns. It is in this sense that the conjecture *learning of words is an evolutionary process* should be interpreted, and it is in this sense that the aspirations of the algorithm hereby introduced are to be understood.

3 Note that we use terms "semantic category", "semantic class" or "concept" as synonyms.

17

EVOLUTIONARY INDUCTION OF A LIGHTWEIGHT MORPHOSEMANTIC CLASSIFIER

17.1 GENERIC INTRODUCTION

The aim of previous chapter was to show that one can use evolutionary computation to induce sufficiently pertinent semantic categories from a corpus of text documents. Individual text documents were understood as "entities", words present within such documents were understood as their "features" and topics[1] to which diverse documents were attributed were understood as "semantic categories".

Analogies between such process of induction of semantic categories and the process of "vocabulary development", occurring in practically every human being since birth until death, have also been made.

In this chapter we shall explore evolutionary models of induction of yet another type of categories which also play a non-negligible role in human linguistic communication. Id est, induction of grammatical categories. And given that a commonly used definition of a grammatical category (GC) as a *grouping of language units sharing some common feature or function* is very general and vague, this chapter shall focus on particular type of GCs, that of "parts-of-speech" (e.g. "nouns", "verbs", "adjectives" etc.). There are three main "technical" reasons which motivate this choice:

- part-of-speech induction (POS-i, P+135-136) and POS-tagging are well-known NLP problems
- in spite of being well-known, relatively few researchers have proposed evolutionary means to solve these problems (P+137-139)
- certain transcripts within the CHILDES (P+196-222) corpus are tagged with POS-labels

and it is the 3rd reason which is to be understood as the most decisive one in regards to "psycho-linguistic" aims of this dissertation. But the ultimate reason for which we have opted to focus on part-of-speech categories is a theoretical one:

Part-of-speech categories tend to integrate word's semantic content with its grammatical function.

1 Note the congruence between the fact that the word "topic" is derived from the Greek τόπος which means "place" and the fact that in computational semantics, a topic is literally understood as a "place" within the semantic space

In other terms, the very information that "X belongs to category of nouns" informs the one, who already disposes of a certain notion of what a noun is, that X most probably denotes a thing or a state. And the very information that "Y is a member of category of verbs" suggests that Y most probably denotes a process or an activity. In this regards, the appartenance of the word W to the category C is an irreplaceable clue to not only of W's function and position in the enveloping utterance, but also to W's meaning. This is maybe not so important when the meaning of W is already known, but in case of a language-learning toddler, the ability to recognize that $W \in C$ could significantly reduce her difficulties in solving the problem "*to which components in a recently perceived scene should be a novel W associated?*".

Simply stated: POS-categories can help the child to bootstrap (Karmiloff and Karmiloff-Smith, 2009, pp.111-118) herself into the language.

But how does a child construct such categories in the first place? The aim of the article hereby introduced and recently submitted to journal Computational Linguistics (Hromada, 2016c), is to propose an evolutionary answer.

17.2 INTRODUCTION

What is the essence of linguistic categories, how are such categories represented in human mind and how do such representations develop? Questions which intrigue linguists and philosophers since time immemorial, questions of such elusive nature that any proposal aspiring to answer them have to be, per definitionem, only partial and incomplete.

Such epistemological problems notwithstanding, contemporary computer science tends to offer an instructive answer: categories are classes and classes can be operationalized as regions within an Δ-dimensional vector space S_Δ. Under such definition, training of a categorizing system (i.e a "classifier") can be simulated as a search for the most accurate partitioning of S_Δ. This holds for categories in general and hence it also holds for linguistic categories in particular.

One possible way how such partitioning can be performed is offered by so-called Support Vector Machines (SVM, Cortes and Vapnik (1995)). Basic idea behind SVMs is simple: the algorithm aims to find such a hyper-plane (also called a "decision boundary") which cuts the vector space into two sub-spaces each of which shall ideally contain only data-points attributed to one class. But not only that: given that many such decision boundaries are often possible and identifiable, an SVM tends to identify the one which maximizes the gap (i.e. margin) between data-points themselves and the boundary. Motivation behind such a choice is simple: the more margin is maximized in regards to objects extracted from the training data-set, the more it can be expected that object extracted from a previously unseen "testing data-set" shall be also projected onto the correct side of the boundary. And very often it indeed does: SVMs are able to generalize.

17.2.1 FROM PLANES TO PROTOTYPES

In spite of their theoretical elegance, SVMs - as well as their neural network "perceptron" counterparts - have one important drawback. That is: SVMs and perceptrons look for a "plane" which cuts the space into partitions. But as is illustrated by Figure 31, data-to-be-classified is very often not "linearly separable": a linear decision boundary is nowhere to be found (Minsky and Papert, 1969). In SVM practice, the problem is often solved by applying a certain "kernel function" (Hofmann et al., 2008) which projects the initial data-set onto the space of higher dimensionality where - if the kernel was well chosen - could be the data separated.

While kernel functions have other pleasing mathematical properties [2], they are highly abstract and of significant« mathematical slant» (Hofmann et al., 2008). This, we believe, makes it almost impossible that kernel-based models could ever be labeled as "cognitively plausible" (Hromada, 2014b). In other terms: it is highly improbable that human cognitive and neurolinguistic system would implement as mathematically precise, pure and fragile a machinery as kernels definitely are.

In this article we shall argue that it is in great extent possible to bypass the problem of "linear separability". This is to be attained by focusing one's attention on neighborhoods points $P_A, P_B, ..., P_X$ supposedly representing categories $A, B, ..., X$ instead of focusing it on linear boundaries B_{AB}, B_{AX}, B_{BX}... which supposedly represent the distinction between A and B; A and X etc.

Hence, categories are to be defined in terms of their prototypes (Rosch and Mervis, 1975; Hromada, 2015). Prototypes themselves are points in S tending to satisfy the following condition:

> *An point P_C can be understood as an optimal prototype of a category C if and only if all data-points attributed to C are closer to P than to any other prototype (P_X, P_Y) simultaneously represented within the system.*

In spite of its surface simplicity, the problem posed by this definition of "the optimal prototype" is not an easy one to tackle in a multiclass scenario: the constraint *closer than any other simultaneously represented prototype* substantially complicates the case. If this constraint wasn't present, the problem of identification of "optimal prototypes" would be trivial: prototype would be the centroid of all members of C. But the condition "closer than any other prototype" makes all components of the system mutually dependent on each other. In the end,

2 The most prominent of which is related to a so-called "kernel trick" which can significantly speed up the classifier-training process.

one is posed in front of the problem somewhat analogical to the famous three-body problem in physics. That is, a problem of which it is well known that it is insolvable by analytic means (Poincaré and Magini, 1899).

17.2.2 FROM PROTOTYPES TO CONSTELLATIONS

This article aims to demonstrate that the problem of discovery of constellations of optimal prototypes can be approximated by a nature-inspired non-connectionist method. In other terms, we shall use a relatively simple evolutionary algorithm in order to "induce" constellations of prototypes which are closer to training data-points to which they should be close and further from training data-points from which they should be far.

Thus, an individual solution contains a position of each component prototype. Every individual has a genome of length $|C|\Delta$ whereby $|C|$ denotes number of distinct classes and Δ is the dimensionality of the space within which the search is performed. As is common to evolutionary algorithms (EAs), these individual solutions are subjected to process replication, selection and variation across multiple generations. Notions of "far" and "close" are implemented directly in the fitness function so that the evolutionary search minimizes the number of incorrectly positioned "nearest prototypes".

Ideally - id est if EAs parameters have been correctly specified and iff the problem of prototype constellation is optimizable at all - the system should converge to such a constellation of prototypes which could accurately classify both testing and training data.

17.2.3 FROM CONSTELLATIONS TO LIGHTWEIGHT CLASSIFIERS

Note that if EAs could discover and optimize such constellations, then these constellations would yield truly "lightweight" classifiers: solution to the $C-$class classification problem of objects in $\Delta-$dimensional space has length $|C| * \Delta$. To be even more radical, let's precise that the search shall operate within binary $\Delta = 64$ spaces which means that position of every data-point as well as a candidate prototype could be defined by exactly 8 bytes. $5-$class classifiers presented in the next sections are described by no more and no less than $5 * 8 = 48$ bytes.

Another reason why these classifiers can be considered as "lightweight" is the nature of features used to project diverse textual tokens into such $64-$dimensional Hamming spaces. Being aware of results issued from our previous empiric simulations (Hromada, 2014a), we have decided to use three features only, i.e.

- suffix of the word *W* (i.e. last three characters of the word-to-be-categorized)

- suffix of the word $W_L eft$ (i.e. word immediately preceding W)
- suffix of the word $W_R ight$ (i.e. word immediately preceding W) are

in order to transform tokens into geometric entities. No other feature has been used during the geometrization phase of the algorithm.

All this in order to propose a nature-inspired model of induction of part-of-speech categories which is, we believe, at least as "minimalist" as Chomsky's "minimalist" program (Chomsky, 1995).

17.3 METHOD

Algorithm presented in this article is very similar to the one presented in (Hromada, 2015). Procedure starts with characterization of training-corpus entities (i.e. "words") in terms of their features (i.e. "suffixes" of W, W_L and W_R) . These features are subsequently used to project all entities into a 64-dimensional Euclidean space $S_{E(64)}$: this component is known as Random Indexing (Sahlgren, 2005). In following steps, whole "space" is reflected so that entities and features "implicitly connected" in the original corpus shall be more pushed to each other than entities and features which are not so connected: this component is known as Reflective Random Indexing (Cohen et al., 2010). At last but not least, all vectors are "binarized" by a simple binary thresholding procedure known as Lightly Stochastic Binarization (Hromada, 2014c). All this steps yield a binary Hamming space $S_{H(64)}$.

Once $S_{H(64)}$ is constructed, one can proceed to localization of most optimal constellations of category prototypes. This is being done by a fairly standard evolutionary algorithm (EA) which is more closely described in 17.3.2.

Most fit solutions obtained after certain number of generations are subsequently confronted with data extracted from the testing corpus in order to assess EA's capability beyond the training set.

17.3.1 CORPUS

This article is conceived as a part of dissertation addressing the possibility of developing evolutionary models of induction of linguistic categories in (and by) human children. This makes the choice of the corpus quite straightforward: the corpus from which we shall aim to extract first linguistic categories is to be contained in Child Language Data Exchange System (CHILDES, (MacWhinney and Snow, 1985)).

However, not all among 30 thousand transcripts contained in CHILDES (Hromada, 2016e) contain part-of-speech labels. Quality of labels also varies: this is no surprise given that some transcripts were manually labeled and/or corrected by multiple annotators while other transcripts have been labeled only by automatic NLP tools (Sagae et al., 2007).

For this reason we have ultimately focused our interest on one particular corpus: Brown's (Brown, 1973) transcriptions of verbal interactions of a girl named Eve. Primo because Brown's work is seminal for whole discipline of developmental psycho-linguistics. Secundo because it is indeed the Eve section of Brown's corpus whose POS-labels have been, according to (Sagae et al., 2007), manually corrected by human annotators.

Classes

According to (Sagae et al., 2007), each token of CHILDES corpus is labeled with one among 31 part-of-speech tags. However, majority of these tags are used only very rarely and/or denote such categories (e.g. AUX for auxiliaries, REL for relativizers or CONJ for conjunctions) of words which encode only little amount or semantic or deontic information.

It is certain that mastery of words belonging to categories like AUX, REL or CONJ play an important role in development of full/fledged adult-like competence. But given that an objective of our dissertation was to elucidate evolutionary computation can simulate bootstrapping of morphosyntactic categories from semantics (and vice versa), we have decided to focus on induction of five classes only. These are enumerated in table 17.3.1.

Class Tag	CHILDES POS tags	Example words
ACTION	v, part, cop	"think", "saying", "is"
SUBSTANCE	n	"cookies", "cow", "ball"
PROPERTY	adj, qn	"better", "blue", "three"
RELATION	prep	"on", "with", "to"
REFERENCE	pro, det, art	"I", "you", "this", "the"

Table 41: Five classes of interest, their corresponding CHILDES part-of-speech tags, some example word types which instantiate them.

What is common to these classes is, that their member words very often denote *visible* and tangible entities, states and processes. Id est, when a child *hears* these words it can be the case that she also perceives their referents by other senses.

Classification of words labeled with tags OTHER than "v", "part", "cop", "n", "adj", "prep", "pro", "art", "det", "qn" has been excluded from the following analysis. Primo,

- because such words do, more often than not, lack easily recognizable visual semantic contents and should not thus be mixed with words which encode such contents

secundo,

- because in ontogeny of a normal child, items belonging to such more abstract classes are mastered later (i.e. after the "toddlerese" (P+17) stage) than words denoting concepts subsumed under five classes listed in (Tomasello, 2009)

tertio,

- because problem of classification of words into 5 classes is, of course, less computationally complex and hence more tractable than problem of classification into 31 classes

and finally,

- it is far from certain whether categories like "auxiliaries" or "relativizers" are represented *per se* within minds of normal verbally communicating humans, or whether such categories are simply abstractions developed by linguists for their own purposes

All these arguments taken together had made us renounce to tentatives to train 31-class POS-classifier and made us focus on training of 5-class classifier only.

Pre-processing

10443 "motherese" utterances have been extracted from twenty transcripts of Brown's Eve corpus. These are very easy to detect because in CHILDES, every utterance is on a separate line and begins with the trigram denoting the locutor of the utterance (in case of mothers, the trigram is MOT). 10443 lines which follow these "motherese" utterances and begin with marker %mor have been also extracted: these are lines which contain manually annotated POS-labels.

Thus, 10443 line-couplets like this:

Listing 10: Motherese utterance from CHILDES corpus + associated morphological tier.

```
eve05.cha:*MOT: that s a duck .
eve05.cha-%mor: pro:dem|that cop|be\&3S art|a n|duck .
```

have been obtained by executing a simple shell command[3]. Lines beginning with MOT and %mor have been subsequently merged by a PERL script enrich_pos.pl[4] which yields output exemplified by the following listing:

Such is the primary data format of this simulation. In this format, each token is characterized on a separate line along with the utterance in which it occurred, as well as with its "gold standard" class-label which was attributed to it by manual annotators. Individual columns

3 cd Brown/Eve; grep -A3 -P '^MOT' * | grep -P '(MOT | %mor)'
4 Publicly available at URL http://wizzion.com/thesis/simulation2/enrich_pos.perl

Listing 11: Primary input format of this simulation.

```
that###REFERENCE###train###that s a duck .
s###ACTION###train###that s a duck .
a###QUANTIFIER###train###that s a duck .
duck###SUBSTANCE###train###that s a duck .
```

are separated by ### separator. The first column denotes the entity itself (the word token), second column contains its class, third column specifies whether the token occurred in a training or testing part of the corpus and the last column contains whole context within which the token entity occurred (i.e. the enveloping utterance).

Let's precise that the training corpus was extracted from first 12 Eve transcripts (i.e. files eve01.cha - eve12.cha) which describe verbal interactions which occurred before Eve attained 2 years of age. Testing corpus, on the other hand, was composed of 8 files (eve13.cha - eve20.cha) transcribed down as Eve was 2 - $2.\frac{1}{2}$ years old.

The script enrich_pos.pl thus outputs 12453 training corpus tokens and 8746 testing corpus tokens instantiating 972 (training) and 934 (testing) word types. Almost one half (449) of word types occurring in testing corpus does not occur in the training corpus.

17.3.2 ALGORITHM

This is the core of the model. It consists of two major components:

1. "vector space preparation" (VSP): a trivial suffix-extracting filter is used in order to project text from the primary input onto a 64−dimensional Hamming space

2. "evolutionary optimization": searches S_{H64} for most discriminative constellations of prototypes

Vector Space Preparation

Approach which was used to "geometrize" the primary textual input shares its essential features with that of Random Indexing ((Sahlgren, 2005)) as well as with other Vector Symbolic Architectures (Cohen et al., 2012) based on so-called Random Projection (Hromada, 2013). We describe it elsewhere as follows:

« Given the set of N objects which can be described in terms of F features, to which one initially associates a randomly generated d-dimensional vector, one can obtain d-dimensional vectorial representation of any object X by summing up the vectors associated to all features F_1, F_2 observable within X. **Initial feature vectors are generated in a way that out of** d **elements of vector, only S among them are set to either -1 or 1 value. Other values contain zero.** Since the "seed" parameter S is much smaller than the total number of elements in the vector (d), i.e. S «d, initial feature vectors are very sparse, containing mostly zeroes, with occasional value of -1 or 1.» (Hromada, 2014c).

Section 17.2.3 has already indicated the nature of features which we shall use to initiate the process of geometrization of textual input. We reiterate: we shall characterize every token T with three principal features only:

1. T's own suffix[5]

2. suffix of the token to T's right

3. suffix of the token to T's left

.

Asides this, only two other "lateral features" are used: token T has feature INIT if it is the initial (i.e. first) token of the utterance. Conversely, it is endowed with feature END if it is the last (i.e. terminal) token of the enveloping utterance.

These 3 principal and/or two lateral features are extracted - during the initial phase of VSP - by a following feature-extracting snippet.

Listing 12: PERL code of suffix-feature extractor

```
sub suffix3_featurefilter {
        my @f;
        my @wrdz=split / /,shift; #utterance in 1st parameter
        my $nam = shift; #token of focus in the 2nd
        my ($index)= grep { $wrdz[$_] eq $nam } 0..$#wrdz;
        $index+=1;
        my $pos = 1;
        for my $w (@wrdz) {
                my $w=lc $w;
                my $s=substr $w,-3;
                my $n=$index-$pos;
                $n=$n*-1; #features with minus to the left
                push @f, $n.$s if (abs($n)<2); #main 3 features
                $pos++;
        }
        push @f,"INIT" if $index==1; #lateral feature
        push @f,"END" if $index==scalar(@wrdz); #lateral feature
        return @f;
}
```

For example, when the Random Indexing procedure makes the following call:

suffix3_featurefilter("that s a duck","that")

5 What we label as suffix SFX_T of token T is, for the purpose of this text, equivalent to T's terminal character trigram (i.e. T's last three letters).

it returns three features characterizing this concrete occurrence (i.e. token) of the word "that":

INIT 0hat 1s

Accordingly, features −1hat, 0s, 1a would be used to characterize this instance of the token s and features −1a, 0uck, END would characterize this instance of duck.

This is the last level of representation which can still be understood as "symbolic". Subsequently, Random Indexing associates a random, sparsely non-zero init vector to each distinct feature (e.g. INIT, END, 0hat, −1hat, 1s, 0s, −1s, −1a, 1a, 0a, 1uck, 0uck etc.) present in any motherese utterance of the Brown/Eve corpus.

All in all, presence of 1321 distinct features has been assessed in the training corpus.

Once features are extracted, things go geometric. Vector representations for individual tokens are obtained as sums of vector representations of associated features. Subsequently, initial random feature vectors are discarded and features themselves are characterized as sums of vector representations of associated tokens. This steps marks the first "reflective" iteration of the process called Reflective Random Indexing (RRI). C.f. Cohen et al. (2010) for closer description of how and why RRI works.

For the purpose of this article, let's just precise that introduction of 2 max 3 "reflective iterations" practically always increases results of one's experiment. This is, in sense, quite expected: for what the reflective process does is not only enriching the representations of entities (e.g. tokens, documents) with information about their features (suffixes, resp. word occurrences) but also enriches representations of features with information about entities within which they occur.

For example, not only should be the word thinking characterized with the feature "ends with suffix −ing" but, conversely, the feature "ing is in part characterized by the fact of occurrence in the word thinking.

Note that all vectors produced by RI and RRI are euclidean. After every "reflection", vectors are normalized so that their unit length is 1. After last such reflection, each real number element of each vector is transformed into a Boolean value by a binary thresholding process known as Light Stochastic Binarization (Hromada, 2014c).

Such binarization is the last step of the vector space preparation. At its end, one obtains a binary vector "hash" tending to have a property common to other convergent[6] hashing methods (Datar et al., 2004; Salakhutdinov and Hinton, 2009):

6 A hashing function F_H is said to be convergent if similarity between its inputs implies similarity of its outputs. On the other hand, F_H is said to be "divergent" if similarity between inputs does not imply similarity between output hashes. Being of strongly divergent nature, functions like SHA2 or MD5 are not to be confounded with convergent hashing which we discuss here.

Similar inputs tend to have similar hashes.

The moment of attribution of binary hash to each token occurring in the corpus marks the end of the "vector space preparation" phase of the algorithm. In the current model, this VSP occurs only once - at the beginning of simulation and is not repeated.

Evolutionary Optimization

Ensemble of all binary hashes obtained from the corpus yields a hamming space S_H with fairly low dimensionality. This is technically very advantageous since measuring distances can be very swift in such spaces: calculating the hamming distance between two binary strings is definitely[7] less costly than calculating a distance between two real (or even complex) vectors.

The fact that we can measure distances swiftly is crucial for our evolutionary approach for measurement of distances constitutes the very core of the fitness function which is to evaluated for every individual member of every single generation of every single run of the simulation. This is exemplified by the following snippet of PERL pseudo-code.

Listing 13: PERL pseudocode of prototype-inducing fitness function

```
my $fitness=0;
for @individual (@population) {
        for $training_token (@training_tokens) {
                $training_token_hash=$hashes{$training_token};
                $training_token_class=$correct_classes{
                    $training_token};
                $true_prototype_distance=hamming_weight(
                    $training_token_hash XOR $individual[
                    $training_token_class]);
                for $incorrect_prototype ($incorrect_classes{
                    $training_token}) { #the innermost cycle
                        $fitness-- if (hamming_weight(
                            $training_token_hash XOR $individual[
                            $incorrect_prototype]) <=
                            $true_prototype_distance);
                }
        }
}
```

As may be seen that the innermost cycle of the fitness function evaluation contains three operations:

1. XOR between vector of the training object $\vec{o}$ and the vector of "false" prototype $\vec{p_F}$: this yields new vector with true values on those positions where elements of input vectors differ

7 Or at least on an ordinary transistor-based 21st century Turing machine

2. calculation of hemming weight (i.e. number of non-zero bits) of XOR's result[8]: this is equivalent to hamming distance $Hd(\vec{o}, \vec{p_F})$

3. penalization (decrementation of fitness value) for every incorrect prototype $\vec{p_F}$ which is not further from $\vec{o}$ than o's true prototype p_T, i.e.

$$Hd(\vec{o}, \vec{p_F}) <= Hd(\vec{o}, \vec{p_T}) \quad (7)$$

This concrete instance of prototype-inducing fitness function can be further elucidated by a formula

$$F_{object}(\vec{i}, \vec{o}) = \underset{p_x \neq p_T \wedge Hd(\vec{o}, \vec{p_x}) <= Hd(\vec{o}, \vec{p_T}) \implies p_x \hookrightarrow P_F}{|P_F|} \quad (8)$$

which defines the object-wise fitness $F_{object}(\vec{i}, \vec{o})$ of individual solution $\vec{i}$ in regards to vector representation of the training object $\vec{o}$ as a number (i.e. cardinality of a set) of "false" prototypes $|P_F|$ which are not further from $\vec{t}$ as $\vec{o}$'s corresponding (i.e. "true") prototype $\vec{p_T}$.

Subsequently, an overall fitness of the individual chromosome $\vec{i}$ in regards to each and every object occurring in a training corpus T, is a sum

$$F_{total}(\vec{i}) = -\sum_{o \in T} F_{object}(\vec{i}, \vec{o}) \quad (9)$$

The sum is inverted so that whole function is a maximization one. Under such definition the **maximum fitness value is 0 and corresponds to situation where all training corpus objects are closer to their true prototypes than to any other prototype**.

In theory, it may be the case that multiple global optima of such kind exist. In practice, and in case of many vector spaces, such global optima may not exist at all and fitness of any locally optimal states will have negative value.

Fitness function thus defined, the form of representation of individual solutions is quite straightforward:

An individual solution $\vec{i}$ encodes a **constellation** of all candidate prototypes of $|C|$ categories.

This means that, in regards to every single object $\vec{o}$ present in the training corpus T, $\vec{i}$ shall encode not only "true" prototype $\vec{p_T}$ associated to $\vec{o}$ by the training corpus. It shall also encode all prototypes which are not $\vec{o}$'s true prototypes and which - if ever located closer to

8 Assembler routine for hamming_weight calculation exploiting the POPCNT instruction implemented (on hardware level) of SSE4.2-compliant CPUs (Suciu et al., 2011) is accessible at URL http://wizzion.com/thesis/simulation2/popcount.asm

$\vec{o}$ than $\vec{p_T}$ - should be evaluated as members of a set of "false positive" P_F.

In practice, individual solution $\vec{i}$ is represented as a vector or an ordered tuple which **concatenates** all its components. Number of possible distinct individuals is

$$2^{\Delta*|C|}$$

where Δ is the dimensionality of the space and $|C|$ is the number of classes. Since in our simulations we have focused on partitioning of 64-dimensional space into five classes ($|C|$=5) there exist potentially $2^{64*5} = 2^{320}$ constellations.

Fitness landscape is thus finite but its complete traversal seems to be impossible to execute in a reasonable amount of time [9].

Two evolutionary heuristics has been deployed in order to explore the landscape:

1. CANONIC: a heuristic strongly reminiscent of Canonical Genetic Algorithms (Goldberg, 1990)

2. MERGE_1: an extension to CANONIC which merges independent runs of CANONIC into one big population and continues the evolution further

In both approaches, every generation starts with fitness evaluation for all individuals in the population. Subsequently, a so-called *2-way tournament selection* operator (Sekaj, 2005) selects members of the mating pool. Size of the mating pool equals the size of population. Members of new generation are obtained from the mating pool as follows: two parents (mother and father) individuals are randomly chosen from the mating pool in order to be subsequently "cut" at a randomly chosen point. Segment before the cut is taken from the mother, segment after the cut is taken from the father and new offspring is obtained. Any gene of offspring's genome can be mutated with 0.2% probability: mutation is equivalent to flipping of a bit. Elitism is not implemented and even the most fit individual can be subjected to decay.

There are thus only two aspects in which CANONIC and MERGE1 differ. One difference is the population size: in CANONIC, populations are fairly small (100 individuals) while MERGE1 implements somewhat bigger ones (1000 individuals).

Both heuristics also differ in the way how their initial population are generated. In CGAs one departs *ex nihilo* and CANONIC heuristics is no exception to this rule: genes present in the gene pool of generation 0 are randomly generated. Things are slightly different

9 At least on clusters of ordinary transistor-based 21st century Turing machines.

in case of MERGE1 heuristics: MERGE1 is initiated by populations yielded by different runs [10] of CANONIC after 200 generations.

CANONIC and MERGE1 taken together can be thus understood as a very primitive form of "parallel genetic algorithm" (PGA) (Sekaj, 2004).

Under this view, 100 independent runs of CANONIC can be understood as independent nodes on the lower level of the hierarchy and MERGE1 as the node of the higher level. A "migration" from all low-level nodes occurs after 200 generations. Follows a big tournament in which the initial MERGE1 population is constituted.

Subsequently, MERGE1 evolves further.

Parameters

VSP	Input corpus	Brown-Eve motherese [11]
	Feature Filter	suffix3
	Dimensionality	$\Delta = 64$
	Seed	S = 3
	Reflections	I = 3
CANONIC	Population size	N = 100
	Selection	Tournament
	Crossover	One-point
	Mutation rate	M = 0.2%
	Initial population	ex nihilo
	Generations	G = 200
	Elitism	E = 0
	Runs	R = 100
MERGE1	Population size	N = 1000
	Selection	Tournament
	Crossover	One-point
	Mutation rate	M = 0.2%
	Initial population	results of CANONIC
	Generations	G = 300
	Runs	R = 6
Machine Learning	Classes	$\lvert C \rvert = 5$

Table 42: Parameters of simulation 2.

10 Note that one common "vector space preparation" phase preceded all CANONIC runs. Hence, in spite of the fact that diverse runs of CANONIC followed different evolutionary trajectories, they always did so in the space S_{64} explored by other runs as well. This makes it possible to "merge" results of different runs.

11 Available at http://wizzion.com/thesis/simulation2/eve12-8-5classes.mot

17.3.3 EVALUATION

Accuracy of induced classifiers was primarily evaluated in terms of quantity of correctly predicted category labels (i.e. true positives). Hence, maximum score of 100% would correspond to situation when all objects have been successfully classified. On the contrary, a classifier attributing category membership by random would have precision of cca. 20% in case of classification into 5 equidistributed classes.

Overall classification accuracy of classifiers induced by CANONIC and MERGE1 heuristics has been evaluated after each 10 generations of the training process. Asides this, each class has been explored individually in order to yield class-specific precision and recall values.

Three other classification methods have been evaluated in order to compare the evolutionary method with non-evolutionary approaches:

- $\text{CENTROID}_{\text{HAMMING}}$ and $\text{CENTROID}_{\text{EUCLIDEAN}}$ baselines
- MSVM (i.e. a Multi-class Support Vector Machine)

Two baseline approaches characterize every class by their centroid. In $\text{CENTROID}_{\text{HAMMING}}$ approach is centroid C_X of a category X a hash obtained as an average of hashes of all objects belonging to X. Things are similar in case of $\text{CENTROID}_{\text{EUCLIDEAN}}$: the only difference being due to the fact that elements of objects and centroid vectors are now represented in their real-valued form. Id est, a representation issued from the last reflective iteration of the RRI component of the VSP phase of our algorithm.

At last but not least, binary vector space issued from the VSP phase has been partitioned by means of a MSVM implemented in the open-source package MSVMPack (Lauer and Guermeur, 2011). Default settings of the package have been used: linear kernel has been applied and training of MSVM2 (Guermeur and Monfrini, 2011) model has been stopped after converging to 98% accuracy level.

17.4 DISCUSSION OF RESULTS

Table 43 summarizes main results of five compared methods. Smallest amount of correctly classified tokens was attained by baseline CENTROID approaches: this was expected since these approaches do not include any optimization at all[12]. The observation that $\text{CENTROID}_{\text{HAMMING}}$ is less precise than $\text{CENTROID}_{\text{EUCLIDEAN}}$ is also trivial: transformation of real-valued vectors into binary ones brings about a non-negligible information loss. Worse performance of binary-based classifiers is a result of this information loss.

Optimization, however, can significantly reduce or even counteract impact of such loss. Hence, even a fairly simple CANONIC genetic

12 Note, however, that classification accuracy of these models is still significantly superior to a random classifier.

Method	Training corpus	Testing corpus
$CENTROID_{HAMMING}$	455 (42.12%)	412 (40.47%)
$CENTROID_{EUCLIDEAN}$	572 (52.96%)	533 (52.35%)
$MEAN(GA_{CANONIC})$	631 (58.44%)	589 (57.88%)
$MEAN(GA_{MERGE1})$	718 (66.51%)	657 (64.57%)
$FITTEST(GA_{MERGE1})$	772 (71.48%)	699 (68.66%)
MSVM2	781 (72.31%)	736 (72.30%)

Table 43: Overall results of five different approaches. GA results have been averaged across diverse runs (R = 6*100 for CANONIC, R=6 for MERGE1).

algorithm discovers, in just five sweeps through the hamming space, constellations of prototypes whose precision is higher than that of Euclidean centroids. This is exemplified by Figure 30 which plots evolution of precision across generations.

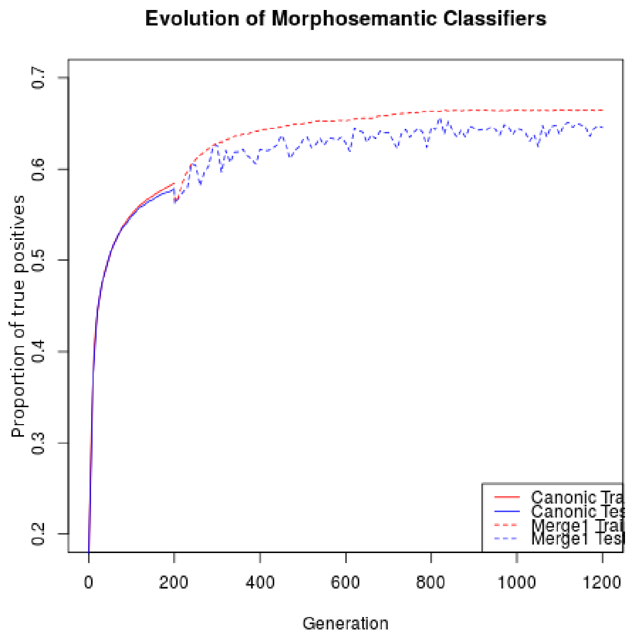


Figure 30: Evolutionary optimization increases the precision of a multi-class classifier. Curves represent results averaged across diverse runs (R = 6*100 for CANONIC, R=6 for MERGE1)).

It may be seen that introduction of PGA-like approach - as exemplified by MERGE1 - results in a significant increase in amount of precisely classified tokens. The score is still not so high as that of MSVM2 (compare 781 with 718 for training corpus, resp. 736 with 657 in testing corpus), but the jump between CANONIC and MERGE1 suggests that that another PGA architecture, introduction of elitism or a different choice of parameters or operators can potentially result in significant boost.

Table 44: MSVM2 training corpus confusion matrix.

	ACT	SUB	PROP	REL	REF
ACT	266	54	0	0	1
SUB	55	495	4	0	0
PROP	21	66	18	0	0
REL	20	12	1	2	0
REF	15	47	3	0	0

Table 45: MSVM2 testing corpus confusion matrix.

	ACT	SUB	PROP	REL	REF
ACT	271	38	4	0	1
SUB	55	450	8	1	0
PROP	21	62	15	0	0
REL	20	6	3	0	0
REF	20	38	5	0	0

Table 46: Training corpus confusion matrix produced by FITTEST(GA_{MERGE1}).

	ACT	SUB	PROP	REL	REF
ACT	278	28	4	4	7
SUB	56	427	34	18	19
PRO	19	39	43	3	1
REL	15	5	1	11	3
REF	9	35	7	1	13

Table 47: Testing corpus confusion matrix produced by FITTEST(GA_{MERGE1}).

	ACT	SUB	PROP	REL	REF
ACT	269	21	9	6	9
SUB	62	371	41	25	15
PRO	16	35	40	3	4
REL	15	3	4	5	2
REF	11	26	8	4	14

As may be seen on confusion matrices shown on tables 5 - 6, MSVM2 fails to correctly classify any testing corpus token attributed to minor REL and REF categories (i.e. recall = 0%) and the situation is not better in case of PROP class neither (testing recall 15.3%)[13]. On the other hand, this handicap is counteracted by MSVM's higher recall rates in regards to dominating SUB and ACT classes. This could potentially suggest that MSVM still tends to behave like a good old "dualist" Support Vector Machine rather than a truly multi-class classifier.

Confusion matrices on tables 7-8 indicate that FITTEST(GA_{MERGE1}) also performs quite well when it comes to classification of tokens into major categories ACTION (86.6% recall; 73.74% precision) and SUBSTANCE (73.74% recall; 79.96%precision). Asides this, it also attains 40% testing recall for PROPERTY class and 22% testing recall for the REFERENCE class. This suggests that even categories of minor importance play a certain role in models induced by evolutionary search for prototype constellations.

13 These low recall rates imply that the average F1 score of MSVM is, in fact, inferior that of FITTEST(GA_{MERGE1}). This is the case for both training ($F_{MSVM} = 0.481; F_{FITTEST(MERGE1)} = 0.518$) as well as testing ($F_{MSVM} = 0.426; F_{FITTEST(MERGE1)} = 0.474$) phases.

17.5 CONCLUSIONS

17.5.1 COMPUTATIONAL CONCLUSION

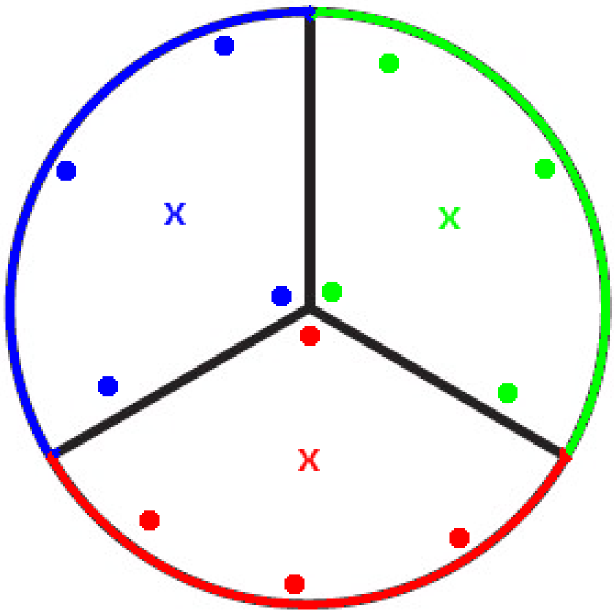


Figure 31: Centroidal tessellation of twelve data-points belonging to three distinct classes. Dots represent data-points, crosses are category prototypes and colors denote category membership. Black lines denote tesselation boundaries.

Figure 31 displays a potential training data-set composed of twelve data-points attributed to three distinct classes. One can observe that it is not possible to draw a single straight line which would separate all datapoints of one class from data-points of other classes. Hence, these data-points are plainly not separable by a linear boundary: many a researcher would be tempted to say that in order to classify such data-set, one would be obliged to apply a certain kernel and project it into space with higher dimensionality.

This is, however, not necessary, if one applies a machine learning strategy which looks for constellation of points instead of lines, planes or hyper-planes. Denoted on Figure 31 by crosses of different colors, such points - labeled as "category prototypes" - satisfy one simple condition:

Every data-point is closer to its prototype than to any other prototype.

Search for constellations of prototypes which satisfy such condition can be thus understood as a problem closely related to problems of Voronoi-Dirichlet tessellations (Aurenhammer, 1991). But contrary to such approaches where "seed" "generator" points are given in advance, positions of such points are, in our approach, induced by means of evolutionary computation.

Inductive process described in this article took place in a 64-dimensional binary space. Reasons behind this choice were of pragmatic nature:

1. optimization involving calculation of Hamming distances can be very fast, especially when implemented on arrays of dedicated Field Programmable Gate Arrays (Sklyarov and Skliarova, 2014) or Application Specific Integrated Circuits

2. binary hashes are very concise form of representation: our approach could be thus useful in Big Data scenarios[14]

These reasons aside, nothing forbids to bypass the "binarization procedure" and search for constellations of prototypes in an Euclidean space. It can be expected that precision of classifiers induced in Euclidean space would be higher than precision of classifiers induced in binary spaces. However, since there is no free lunch, such euclidean search would be undoubtedly more demanding when it comes to consumption of both memory and computational resources.

Shortcomings related to decision to execute the search in binary space notwithstanding, obtained results are quite encouraging. Hence, in a scenario aiming to classify tokens occurring in Brown-Eve section of the CHILDES corpus into 5 morphosemantic classes, classifiers induced by evolutionary optimization identified almost as many true positives as a multi-class SVMs (Lauer and Guermeur, 2011). In terms of F1-Score obtained as a harmonic mean of average recall and average precision, the performance of the most fit prototype constellation FITTEST(GA_{MERGE1}) turned out to be even higher than that of MSVM2. This, however, is more a residuum of a F1-score metrics than a result which would merit to be reported elsewhere than in a $fotnote_{13}$.

17.5.2 PSYCHOLINGUISTIC CONCLUSION

Table 48 lists tokens located in closest neighborhoods of three major prototypes which have been encoded in the constellation FITTEST(GA_{MERGE1}).

A subsequent inspection of false positives present in Table 48 turns out to be quite instructive. Hence, the token "building", present in the utterance "what are you building here?" on line 5417 of eve05.cha transcript is clearly not a noun, as CHILDES annotators supposed, but rather a participle - and hence an instance belonging to ACTION class, as correctly predicted by FITTEST(GA_{MERGE1}). Idem for "hit" present in the utterance "did you hit your head?" present on line 4145 of eve01.cha transcript: the token is clearly not a noun, as postulated

14 In case of 64-bit hashes one could potentially need as little as 800 Megabytes of storage volume in order to store hash representations of 100 million documents.

P_ACTION			P_SUBSTANCE			P_PROPERTY		
H	TOKEN	POS	H	TOKEN	POS	H	TOKEN	POS
10	pointing	part	10	penny	noun	18	**whistle**	noun
10	tripped	v	10	tummy	noun	20	**bent**	v
11	slipped	v	11	cracker	noun	21	**graham**	noun
11	squashing	part	11	graham+cracker	noun	21	**part**	noun
12	**building**	noun	11	key	noun	21	tough	adj
12	burped	v	11	**matter**	v	22	alright	adj
12	cutting	part	11	nap	noun	22	other	adj
12	dripping	part	11	paddle	noun	22	**pitcher**	noun
12	fixed	v	12	drinker	noun	22	**sweetheart**	noun
12	mix	v	12	letter	noun	23	**a**	art
13	dropped	v	12	**numbers**	v	23	**cough**	noun
13	hit	v	12	paper	noun	23	**fun**	noun
13	**hit**	n	12	snowman	noun	23	**grannie_hart**	noun
13	playing	part	12	**worse**	adj	23	**lemon**	noun
13	are	v	13	bx	noun	23	little	adj
13	saw	v	13	face	noun	23	**through**	prep
13	standing	part	13	maam	noun	24	all_gone	adj
13	swim	v	13	purple	noun	24	good	adj
13	want	v	13	soup	noun	24	bigger	adj
13	wiped	v	13	stove	noun	24	busy	adj

Table 48: Testing corpus tokens closest to prototypes of ACTION, SUBSTANCE and PROPERTY encoded in FITTEST(GA_{MERGE1}) constellation. Hamming distance H(token, prototype) and token's CHILDES part-of-speech annotations. False positives are marked by bold font.

by CHILDES annotators, but, as predicted, a verb and hence member of ACTION class. And one can continue: the token "matter" annotated on lines 2152 and 5688 of CHILDES corpus as a verb is clearly not a verb but a noun - and hence a member of a class SUBSTANCE - because it twice occurs in the utterance "what's the matter?. And in spite of the fact that CHILDES labels the token "numbers" as a verb, it is definitely not a verb when it occurs in the utterance "the numbers are going around too" (eve15.cha, line 6276). Et caetera et caetera.

Thus, in spite of the fact that POS tokens in Brown/Eve section of the CHILDES corpus are supposedly « annotated with high accuracy» (Sagae et al., 2007) it is, sometimes, not really the case. In this regards, one would be tempted to state that, as of 2016 AD, is the frontier between developmental and computational psycholinguistics still resembling a structure standing on clay feet. This is a first conclusion which could be potentially useful to any (comp | dev)psycholinguists willing to undertake the path initiated by the study hereby introduced.

The fact that our approach has allowed us to identify errors in the corpus which even humans didn't succeed to identify, is indeed encouraging. And it is moreso encouraging when one realizes how simple was a feature set which has been used to construct the vector space in which all subsequent classifications took place. We repeat:

Every token T was primarily characterized by:

1. *T's three last characters*
2. *three last characters of the token which precedes T*
3. *three last characters of the token which follows T*

asides this, only other information taken into account concerned T's potential position at the very beginning or end of the utterance.

Reason to depart from such a restricted feature set has been in part empiric (Hromada, 2014a). But there exist others, more profound reasons why we have initiated the *training of a verbally interacting computational agent* with focus on suffix-like features. Primo: the "less is more" hypothesis whose implication for neural-network-based processing of natural language has been so beautifully demonstrated by Elman (1993).

Secundo, note Slobin's operating principle A:

« Pay attention to the ends of words.» (Slobin, 1973)

which, according to its author, is a "general developmental universal".

In this regards does our analysis indeed demonstrate that "ends of words" offer features strong enough to initiate a supervised process of induction of categories which have been, for the purpose of this article, labeled as "morphosemantic".

And that the whole process can yield fruit even when representing a 5-class classifier with representation as concise, as 40-bytes long vector definitely is.

17.6 Generic discussion

This chapter has presented an algorithm which succeeds to correctly classify a significant amount of tokens into so-called "morphosemantic classes" (MS-classes). But why should one speak about such MS-classes instead of staying faithful to well-established term "parts of speech" ?

An answer is simple: because MS-classes are sometimes not equivalent to parts-of-speech categories. For example, an MS-category labeled as "ACTION" includes not only verbs, but also participles. Motivation behind this distinction is quite simple: it may potentially make sense for an expert linguist to state that "eating" functions is a participle but "to eat" a verb. However, a modal toddler of 20 months shall most probably turn out be ignorant of such a distinction (Tomasello, 2009). For what counts for such a toddler is the fact that he can associate both words "eat" and "eating" with the fact of simultaneously observing certain invariant structural property of her[15] surrounding environment (i.e. observes *activity* of putting something into one's mouth).

Table 17.3.1 introduced five initial MS-classes[16]. These MS-categories have been defined very loosely in the limited scope of this study: all substantives where defined as belonging to the class SUBSTANCE, diverse verbal, participial and infinitival forms as those instantiating ACTION, adjectives and numerals were collapsed into the MS-class PROPERTY, everything which had something to do with pointing, specification and deictics was subsumed under REFERENCE and prepositions were told to instantiate notion of RELATION.

Said in more practical terms: introduction of the notion of MS-class allowed us to enrich certain section of the CHILDES corpus (i.e. Brown's 20 transcripts of a girl named Eve) (Brown, 1973) with certain amount of *loosely semantic* information. *Loosely* because MS-classes, as used in this chapter, are loosely constructed themselves. For it is not always true that $POS_{substantives}$ always denote substances and POS_{verba} always denote actions: no serious linguist could defend such a general view in more than one article and still stay unostracized by the linguistic community.

Loosely, but in regards to "motherese" addressed to a modal toddler (P+17), also *semantic*. For what is more vital for a 18-month old child, to understand&express the difference between verb "eat" and participle/property "eating", or rather understand&express the difference between act of eating and the object being eaten ?

We summarize: act of making a notational turn from the concept of "parts-of-speech" to the notion of "morphosemantic class" led to enrichment of CHILDES corpus with few bits of semantic information. Few bits maybe, but still more bits than noise. Subsequent coupling of this information with morphological information contained in suffixes followed by optimization by means of an evolutionary algorithm allowed us to converge to very concise, 40-byte long multiclass classi-

15 To stay consistent with Conceptual Foundations as well as with other books of psycholinguistic tradition, we refer to toddlers and children with feminine pronouns "she", "her" etc.

16 We leave to reader's own ingenuity the exploration of an extent in which could these MS-classes correspond to Aristotle's categories, or Kant's and Piaget's "forms of pure reason".

fiers. These classifiers have subsequently resulted in identification of errors produced by much more complex and - so the authors pretend-also « highly accurate» (Sagae et al., 2007) POS-tagging systems supposedly corrected by multiple human annotators.

These considerations make us believe that a notion of morphosemantic classifier could be of certain use and applicability for any present or future researcher aiming to deploy, develop or fine-tune certain nature-inspired yet cognitively plausible (Hromada, 2014b) models of ontogeny of linguistic categories [17].

17.7 SECOND SIMULATION BIBLIOGRAPHY

17 Proof-of-concept source code of this simulation is freely available at URL http://wizzion.com/thesis/simulation2/ELLA.tgz under mrGPL licence.

18 EVOLUTIONARY INDUCTION OF 4-SCHEMA MICROGRAMMARS FROM CHILDES CORPORA

18.1 GENERAL INTRODUCTION

First simulation has indicated that one can use evolutionary computation in order to partition a semantic feature space into regions which roughly correspond to certain "topics". Second simulation has shown how an evolutionary search succeeds to increase the accuracy of so-called morphosemantic classifiers. Both simulations differed in regards to corpus-which-was analyzed (20 newsgroups corpus in simulation 1, CHILDES/Brown/Eve corpus in simulation 2) as well as in a feature set used to project initial text into binary vector space.

However, both simulations:

1. were optimized by means of an evolutionary algorithm
2. succeed to transpose knowledge present in the training set in order to correctly classify the elements of the testing set (i.e. generalization)
3. used labeled corpus as input of the learning process

Taken together, points two and three indicate that simulations 1 and 2 can be understood as particular instances of *supervised* machine learning. That is, a case of learning which demands more than exposition to the plain input corpus. In case of supervised learning, one needs to have another, parallel, source of information as well. Category labels which have been manually attributed by human annotators are most common cases of such "parallel" source of information.

It may be the case, however, that certain problems do not necessitate the exposure to such additional input at all. Such is, according to some linguists, also the problem of grammar induction (P+148-162) whereby one aims to infer a grammar of a language L solely from the corpus of utterances of L.

Because of this, computational models of GI are considered to be particular cases of *unsupervised* machine learning[1].

This chapter shall aim to present one particular model of GI. That is, an evolutionary model strongly resembling models presented in

1 Note, however, that the very act of choosing, in the moment T_0 (and not in T_1) and input corpus C_X and not C_Y can also be considered as an act of supervision. C.f. (Hromada, 2014b, 2016f) for further discussion of the "unsupervised" vs. "semi-supervised" dilemma.

previous chapters. But also a model aspiring to induce certain generic "micrograммars" from nothing else than the Brown/Eve section of the CHILDES corpus.

Article presented in this chapter has been submitted to journal Evolutionary Computation (Hromada, 2016b).

18.2 INTRODUCTION

Input of Grammar Induction (GI) process is a corpus of sentences written in language L, its output is, ideally a grammar (P+117-P+124) or a transparent language model able to generate sentences of L, including such sentences that were not present in the initial training corpus.

In spite of a seemingly simple nature of the problem, induction of grammars from natural language is quite a difficult nut to crack. Thus, symbolic models like the Syntagmatic-Paradigmatic GI (Wolff, 1988), graph-based ADIOS (Solan et al., 2005; Brodsky et al., 2007) do, indeed, attain interesting results in their efforts to extract English grammar from English corpora.

But given the deterministic nature of these models, they tend to converge to certain local optima from which there is no way out. To make things worse, such models often do not dispose of means which would allow them to purge themselves from unwanted over-regularizations (P+83).

In this chapter, we shall present a GI model aiming to harness evolution's ability to *discard the unwanted*. What's more, we shall exploit the genotype - phenotype distinction (Fogel, 1995) in order to perform sub-symbolic variation of sets of symbolic sequences. By doing so, we shall obtain a models which integrates entities represented at two levels of abstraction:

1. sub-symbolic feature vector spaces
2. symbolic PERL-compatible regular expressions

Ideally, such a model could be both robust as well as flexible enough to find its middle path between grammars which cover just one thing, and grammars which cover everything.

18.2.1 TWO EXTREMES

The nature of resulting grammar is closely associated to the content of the initial corpus as well as to the nature of the inductive (learning) process. According to their « expressive power », all grammars can be located somewhere on a « specificity – generality » spectrum. On one extreme lies the grammar having following production rules :

$$1 \rightarrow 2*$$

$$2 \rightarrow a|b|c \ldots Z$$

whereby $*$ means «repeat as many times as You Want» and | denotes disjunction.

This very compact grammar can potentially generate any text of any size and as such is very general. But exactly because it can accept any alphabetic sequence and thus does not have any « discriminatory power » whatsoever, is such a grammar completely useless as an explication of system of any natural language.

On the other extreme of the spectrum lies a completely specific grammar which has just one rule :

$$1 \rightarrow < Corpus >$$

This grammar contains exactly what $Corpus$ contains and is therefore not compact at all (in fact, it is even two symbols longer than $Corpus$). Such a grammar is not able to encode anything else than the sequence which was literally encoded by the training $Corpus$.

Such grammar is therefore completely useless for any scenario were novel sequences are to be generated (or accepted).

The objective of GI process is to discover, departing solely from $Corpus$ (written in language L), a grammar which is neither too specific, nor too general. If it is too general, it shall «over-regularize» (P+83). That is: such G shall be able to generate (or accept) sentences which the common speaker of L would never ever consider as grammatical.

On the other hand, if G is too specific, it shan't be able to represent all sentences contained in $Corpus$ or, if it shall, it shan't be able to generate (or accept) any sentence which is considered to be sentence of L but was not present in the initial training corpus $Corpus$.

18.2.2 DEFINITIONS

G-Category (DEF)

Let's have a set of N objects ($O_1, O_2, ..., O_N$) embedded within a Δ-dimensional space S (i.e. every object O_X can be described by a vector $\vec{o}_X = V_1, V_2, ..., V_\Delta$). Then geometrized category (G_Δ-Category) **C is defined as a content of S-embedded D-dimensional sphere** with

1. centroid whose coordinates are given by a vector $\vec{c} = C_1, C_2, ..., C_\Delta$
2. radius R

Under such definition, **all objects** $O_Y, O_Z, ...$ positioned **within volume of C are** to be understood as **members of C**.

END G-CATEGORY 18.2.2.0

We reinforce: under this view, a G_Δ-category is a convex region within S (Gärdenfors, 2004)[2]. Concrete geometric properties of such a ball (e.g. increase in volume in regards to increase of radius etc.) are, of course dependent on the nature of metric space in which the sphere is embedded (e.g. $V/r = 4/3\pi r^3$ for 3E-categories, i.e. categories embedded within 3-dimensional euclidean space).

In our simulations 2 and 3, we have used the Lightly Stochastic Binarization (Hromada, 2014c) algorithm to project initial objects onto positions within 128- or 64-dimensional binary Hamming spaces. We define categories within such spaces as follows:

H_Δ-*Category (DEF)*

H_Δ-Category is a Hamming ball within a Δ-dimensional Hamming space.

END H_Δ-CATEGORY 18.2.2.0

Given that

1. the radius of a H_Δ-Category cannot be higher than Δ (for such a sphere would envelop whole space S)
2. any integer Δ can be represented with $\log_2 \Delta$ bits
3. $\log_2 128 = 7$ and $\log_2 64 = 6$

it is evident that one needs exactly 135 bits of information[3] - in order to unambiguously specify a specific H_{128}-category embedded in a 128-dimensional hamming space.

And one needs 70 bits of information in order to unambiguously specify a H_{64}-category embedded in a 64-dimensional hamming space.

In this simulation, we shall juxtapose vectors representing diverse H_{64}-categories in order to obtain more complex schemata.

N_Δ-*Schema (DEF)*

An N_Δ-Schema is a result of concatenation of N vectors $\vec{g_1}, \vec{g_2}, \ldots, \vec{g_n}$ whereby each vector $\vec{g_1}, \vec{g_2}, \ldots, \vec{g_n}$ represents a G-category located within a Δ-dimensional space S_Δ.

END N_Δ-SCHEMA 18.2.2.0

Focus of the current simulation shall be on induction of schemata in case where $N = 4$. Given that basic units of such $4-$schemata will be H_{64}-categories, it can be easily seen that they such $4-$schemata could be encoded by no more and no less than $4 * 70 = 280$ bits.

END DEFINITIONS 18.2.2

2 Those endowed synesthesia could potentially visualize G-categories as Δ-dimensional pearls (Hesse, 1967) or *balls of certain material, state and color.*

3 128 bits to specify coordinates of the centroid and 7 bits to specify the radius

Under these definitions, the model and the simulation described in this text can be understood as a method which aims to infer - from plain-text Corpus written in language L_{Corpus} - a 4—schema or (a set of 4—schemata) able to *generate* utterances which were originally not in the Corpus but are nonetheless still syntactically correct utterances of language L_{Corpus}. END INTRODUCTION 18.2

18.3 MODEL

In its essence, model presented in this simulation is reminiscent of the model presented in (Chapter 17). Hence, during the phase of "vector space preparation", texts from English-language transcripts of CHILDES corpora are first projected into 64—dimensional Hamming space H_{64} . Subsequently, a search within H_{64} is realized by means of an evolutionary algorithm.

There exists, however, a certain difference which ultimately causes the algorithm hereby presented to be essentially a non-supervised one. Thus, in the present situation, a H_X-category increase the probability of its survival in time if and only if is H_X contained in the utterance-like N—schema which matches as many utterances as possible.

18.3.1 VECTOR SPACE PREPARATION

Listing 14: PERL code of neighbor-word feature extractor

```
sub word_juxtaposition_featurefilter {
        my @features;
        my @all_words = split / /, shift;
        my $word = shift;
        my ($word_position)= grep { $all_words[$_] eq $word }
            0..$#all_words;
        if ($word_position==0) { #word begins the utterance
                push @features,"INIT";
                push @features,"1".$all_words[$word_position+1];

        } elsif ($word_position==$#all_words) { #word ends the
            utterance
                push @features,"-1".$all_words[$word_position-1];
                push @features,"END";
        } else {
                push @features,"-1".$all_words[$word_position-1];
                push @features,"1".$all_words[$word_position+1];

        }
        return @features;
}
```

Method known as Light Stochastic Binarization (LSB) (Hromada, 2014c) is used to project the input text onto H_{64}. Note, however, that initial features slightly differ from both approach presented in Chapter 16 which used word frequency distributions to project documents onto a resulting *semantic* space, as well as from approach presented in Chapter 17 which used suffixal information to project words onto a resulting *morphosemantic* space.

In contrast to both these methods, the feature extractor presented on Listing 14 focuses on two sources of information only: the identity of the word W_L and the word W_R juxtaposed to the left (resp. to the right) side of the target word W_X.

For example, the function call:

word_juxtaposition_featurefilter("this is a dog","dog")

returns array @features characterizing this concrete token of the word "dog" in terms of two features:

$$-1a, \text{END}$$

In this case, the first feature encodes the fact that the token is preceded by an indeterminate article a while the second feature encodes the fact that "dog" is the last token of the utterance. Similarly, the token this would be characterized by features INIT, 1is; token is would be characterized by features −1this, 1a and the token a would be characterized by features −1is, 1dog.

Once each word of each utterance is characterized by its features, one follows a standard Random Indexing procedure (Sahlgren, 2005) in order to attribute each distinct feature a distinct randomly generated 64−dimensional sparsely non-zero "init" vector. Subsequently, euclidean representation of every word type W_X is obtained as a sum (i.e. unweighted linear combination) of features to which W_X is associated in the corpus.

These euclidean vectors are later normalized and enter the binarization procedure which leads to concise 8-byte hashes having the property:

The more words W_X and W_Y tend to occur in similar contexts, the less the Hamming distance between $\mathrm{LSB}(W_X)$ and $\mathrm{LSB}(W_Y)$ shall be.

It is, indeed, this property which shall potentially allow us to effectuate successful evolutionary searches within the H_{64} space which could be potentially labeled as "morpho-syntactic".

END VECTOR SPACE PREPARATION 18.3.1

18.3.2 BRIDGING THE SUB-SYMBOLIC AND SYMBOLIC REALMS

In order to better understand the model hereby presented, one needs to understand a certain distinction often implemented by proponents of evolutionary programming (Fogel, 1995) or evolutionary strategies (Rechenberg, 1971). Id est, the distinction between the genotype and the phenotype.

Genotype

Information-encoding substrate potentially modifiable by variation and replication operators. Unambiguously translatable into phenotype.

END GENOTYPE 18.3.2.0

Phenotype

Concrete manifestation of specific genotype against which fitness can be evaluated. A distinct phenotype P_X can potentially manifest multiple distinct genotypes.

END GENOTYPE 18.3.2.0

Listing 15: Transcription of vector representations (genotypes) into regular expression phenotypes

```
$regex = "";
$extension = "";
for $component (0..5) {
        $component_regex = "";
        $component_extension = 0;
        $radius=$genotype_radius[$component];
        for $word (@all_words_in_corpus) {
                $word_hash=$word_hashes[$word]};
                $word_hcategory_distance = hamming_weight(
                    $word_hash XOR $genotype[$component]);
                if ($word_hcategory_distance<$radius) {
                        !$cregex ? ($cregex = '('.$word) : (
                            $cregex .= ('|'.$word));
                        $cextension++;
                }
        }
        $cregex ? ($regex .= ($cregex.')')):($regex .= '');
        $extension *= $cextension if ($cextension);
}
$regex='^'.$regex.'$'; #utterance-based
```

In context of the current simulation, $N-$schemata (18.2.2) of length $N = 4$, i.e. $4-$schemata, are to be understood as individual genotype instances. As is always the case in evolutionary computation,

Word	Hash	Word	Hash
this	BABA	that	BABB
it	BAAB	is	0F23
are	0F11	a	C123
the	C125	not	5FF5
duck	7720	dog	7725

Table 49: Words of a $\text{Corpus}_{\text{Mini}}$ and hexadecimal representations of their potential hashes.

Syntagma[5]	H1		H2		H3		H4		H5	
	Center	Radius	Center	R	Center	R	Center	R	Center	R
	BABC	17	0F20	5	5FF0	7	C124	3	7723	7

Table 50: A candidate genotype which could be potentially induced from the hypothetic $\text{Corpus}_{\text{Mini}}$.

these schemata replicate, mutate, cross-over etc. But in order to get their fitness attributed, these genotypes have to be translated into phenotypes. Such translation is realized by means of the procedure displayed on Listing 15

The core idea of the genotype - phenotype translation is to be found on lines 9-11. On line 9, a hamming distance between hash of each among 5 components of the candidate genotype 4−schema is evaluated in regards to hash of each word W_X represented in the H_{64} vector space. On line 10, algorithm checks whether the obtained distance is smaller than the radius which is also included in the genotype. If yes, then the literal sequence of signifiant of the word W_X is injected into the resulting phenotype in a way, so that the resulting phenotype would be a syntactically correct Perl-Compatible Regular Expression (Wall et al., 1994; Hromada, 2011, 2016e) .

In other terms, the code displayed in Listing 15 can be understood as a method of translation of sub-symbolic (feature-based) binary vector representations into symbolic representations known as regular expressions.

For example, let's look at Table 49 which illustrates a small hypothetical $\text{Corpus}_{\text{Mini}}$ containing only words *that*, *this*, *it*, *is* ... and their corresponding binary hashes [4].

Then if ever a $5-schema$ like the one presented in Table 50 would be identified by the evolutionary search, it would be translated into a regular expression:

4 As usual, 64-bit hashes are presented in hexadecimal format as sequences of four characters from range 0-9A-F

5 In order to stay aligned with traditional linguistics, we shall sometimes use the term "syntagma" (resp. its abbreviated form "syn") as a synonym for the term "component".

^(this |that|it)(is)(not)(a |the)(dog |duck)$

which represents the microgrammar

$$\begin{aligned} Utterance &\rightarrow Syn_1 Syn_2 Syn_3 Syn_4 Syn_5 \\ Syn_1 &\rightarrow this \mid that \| it \\ Syn_2 &\rightarrow is \\ Syn_3 &\rightarrow not \\ Syn_4 &\rightarrow a \mid the \\ Syn_5 &\rightarrow dog \mid duck \end{aligned} \tag{10}$$

potentially covering 12 distinct utterances[6]. It would, however, not match utterances of a sort "this are not the dog" because the Hamming distance between the word *are* and the centroid of the 2nd component is bigger than the radius of the very same component (i.e. $HD(LSB("are"), Centroid_2) = HD(0F11, 0F20) = 9 > Radius_2$).

In such a way, one can determine the exact form of a Perl-Compatible regular expression (PCREs) by means of distance measurements in the underlying H_{64} space. And given that PCREs are

1. strings of symbols which describe sets of strings of symbols
2. a sort of *lingua franca* of many engineers active in the domain of Natural Language Processing, data-mining or information retrieval
3. well-tuned and optimized by almost three decades of development by not only PERL but also C++, Python, or R communities
4. transparent to inspections by human examinators[7]

one can potentially start to see a certain utility usefulness in developing an architecture which can unambiguously transform subsymbolic geometrized genotypes into comprehensible, symbolic, and manually modifiable PCRE-compatible phenotypes.

18.3.3 Fitness function

Fitness of N−schema N_X is principally determined by two characteristics:

6 We shall further denote the quantity of *maximal theoretical number of covered utterances* with the term **extension**.

7 Only 5 PCRE meta-characters are used in this article: **(** denotes beginning of a disjunctive group; **)** denotes end of a disjunctive group; | is a separator between two members of a disjunctive group; ^ denotes beginning of expression and **$** denotes the end of expression

1. **extension** E, or a maximal theoretically possible sensitivity, is a finite natural number representing the quantity (i.e. the cardinality of a set) of all utterances which could be matched by N_X

2. Corpus **sensitivity** Y is a number of utterances, present in the Corpus, which have been matched by N_X

More formally: Let's have a $N-$schema X composed of N H_{64}-categories $H_{X1}, H_{X2}...H_{XN}$. Then X is said to have an overall extension E defined as a multiplicative product of extensions of individual categories:

$$E_X = \prod_{k=1}^{N} I_{H_k} \tag{11}$$

whereby the individual extension I_{H_k} of a $k-$th category H_k is defined as number of members of H_k. I.e. $|I_{H_k} = |H_k|$ where $|H_k|$ denotes the cardinality of set of objects whose distance from centroid $\vec{h_k}$ is less than the radius of category H_k.

For example, extension E of the $5-$schema presented in Table 50 is 12 because $I_{H_1} * I_{H_2} * I_{H_3} * I_{H_4} * I_{H_5} = 3 * 1 * 1 * 2 * 2 == 12$.

In contrast to E which is more an information-theoretic quantity, is the sensitivity Y a value which is always relevant in regards to certain corpus.

$$Y_X = N_X matchesCorpus$$

This notion is further exemplified by first line of following listing.

Listing 16: PERL code behind fitness function $Fitness_1$

```
my $sensitivity = true { /$regex/ } @corpus;
#returns number of utterances in @corpus matchable by $regex
if ($sensitivity) {
        $f=($sensitivity**2)/$extension;
} else {
        $f=0;
}
```

Extension and sensitivity thus defined, the fitness value of the schema N_X has been, for the purpose of the current simulation, defined as:

$$Fitness_1(N_X) = \frac{Y_X * Y_X}{E_X} \tag{12}$$

Rationale behind our choice of this and not other [8] is simple: given that we shall tend to maximize the fitness function, we put extension

8 Many other fitness functions, of course, are possible and only very few of them have been tested. It cannot be excluded that more useful fitness functions shall be identified in the future. If not, then the fitness function $Fitness_1$ hereby defined could be potentially thought of as an expression of certain cognitive law. Such conjectures, however, would bring us too far.

in the denominator (i.e. divisor) while putting the sensitivity into numerator (i.e. dividend).

Thus aligned, it may be expected that implementation of such a fitness function shall direct the evolutionary search towards schemata with both low extension as well as with high sensitivity. For this reason, sensitivity is squared in order to somewhat *counteract* the impact of extension which itself is a multiplicative product of its components.

18.3.4 EVOLUTIONARY STRATEGY

The INDUCTOR$_1$ evolutionary algorithm implemented in this simulation is similar to the algorithm CANONIC presented in 17.3.2. Tournament operator is used as the main and only method of selection of fit individuals from the population to the mating pool. Size of the mating pool is equal to population size and mutations of centroid coordinates are equivalent to "bit flipping".

There exist, however, certain important differences which distinguish the algorithm hereby presented from the CANONIC:

1. implementation of phenotype-genotype distinction
2. evolution of both centroid coordinates as well as category radii
3. zeroth population is not generated pseudo-randomly
4. crossover occurs only at specific locations
5. re-focusing strategy is implemented

Taken together, this differences result in an algorithm endowed with certain characteristics of an evolutionary strategy (Rechenberg, 1971) or evolutionary programming (Fogel, 1995).

18.3.5 EVOLUTION OF BOTH CENTROIDS AND RADII

As had been already indicated, individual solutions identified by INDUCTOR$_1$ are essentially nothing else than $4-$schemata. That is, binary vectors which encode a syntagmatic sequence of four H_{64}-categories.

Given that a H_{64} category are defined in terms of both their center as well as radius, INDUCTOR$_1$ tries to identify not only the most optimal coordinates of category's centroid (as was the case in Chapter 17), but also the most optimal "extension" which is principally represented by H's radius.

Information about radius of each category is thus also part of the chromosome and is encoded as an integer value from range $< 0, \Delta >$. Probability of mutation of radius-encoding gene is 0.2%. If subjected to mutation, radius is either decremented or incremented with 1: this corresponds to category becoming less, resp. more exhaustive.

18.3.6 pseudo-random initialization of 0th population

Every single individual of the initial population of N−schemata is generated as follows:

1. choose a random word W_1 occurring in the corpus and retrieve its geometric coordinates $\vec{w_1}$
2. define $\vec{w_R}1$ as the center of first category H_1
3. choose a random word W_2 occurring in the corpus and retrieve its geometric coordinates $\vec{w_2}$
4. define $\vec{w_2}$ as the center of the second component H_2
5. ...
6. choose a random word W_N occurring in the corpus and retrieve its geometric coordinates $\vec{w_N}$
7. define $\vec{w_N}$ as the center of the last syntagmatic component H_N

Subsequently, a radius which is neither too big nor too small is attributed to each among N components. In case of $\mathsf{INDUCTOR_1}$, the radius was set-up to value 13[9] which, in context to 64−dimensional Hamming space, seems to denote a distance which is neither too small nor too big.

Thus, contrary to *ex nihilo* initialization of CANONIC which started the induction process from randomly generated positions of all centroids, is $\mathsf{INDUCTOR_1}$'s initial 0-th population only partially random.

This is so because at the end of initialization process, center of each component of every individual N−schema is the same as the position of a certain word present in the Corpus. [10]

18.3.7 locus-constrained cross-over

$\mathsf{INDUCTOR_1}$ cross-overs took place only at specific loci: namely at positions 64, 128 and 192 of the chromosome specifying centers of diverse G − categories. In more practical terms, such a design choice assured that information precising all coordinates of G − category of the parent individual X have been substituted by information precising all coordinates of another G − category encoded in another parent individual Y.

9 Big radius results in big extension of the corresponding category and hence to many false positives. Small radius causes the category to have small extension and hence to potentially miss many true positives.

10 Such an approach significantly boosts the inductive process which could have otherwise certain difficulties in *booting itself up*.

This distinction aside, the usage of cross-over in $\mathrm{INDUCTOR}_1$ strategy has been fairly standard: every individual of a new generation was obtained as a result of cross-over between two randomly chosen members of the mating pool.

18.3.8 RE-FOCUSING STRATEGY

Another particular aspect is related to $\mathrm{INDUCTOR}_1$'s ability to prioritize, with every new run, induction of new schemata. In practice, this is attained by starting every new run with execution of the code present in Listing 17.

Listing 17: PERL code behind re-focusing strategy

```
@corpus =grep {!/$previous_fittest_schema/ } @corpus;
```

Literally speaking, this line of code removes from the corpus all utterances matched by the most fittest $N - schema$ of the previous run. This results in gradual shrinking of size of the corpus against which the fitness of all future candidate schemata shall be evaluated.

In more general terms, the re-focusing strategy orients the process to *inference of schemata from such utterances, from which* ***no*** *schema has been yet induced*. [11].

And said in more "cognitive" terms, the algorithm invests more attention into exploration of structural regularities within data which have not yet been explored.

11 Inductive process lacking the re-focus strategy would often "lock" itself to most salient patterns present in the corpus which would result in distinct runs often converging to similar schemata.

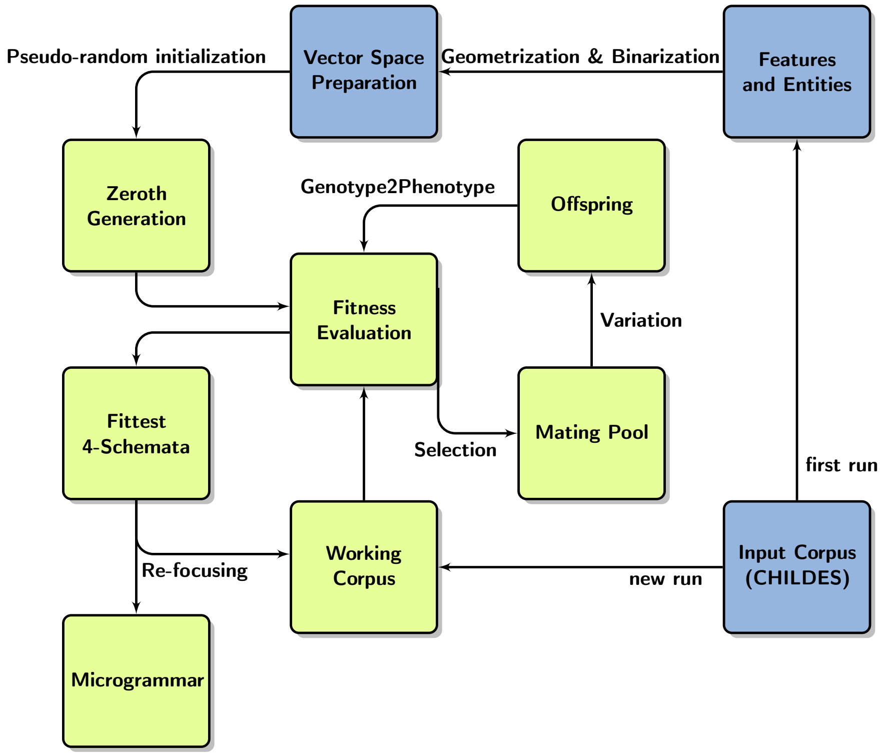


Figure 32: Data flow among main components of INDUCTOR. Lime color denotes components related to evolutionary optimization, royal blue color denotes components of the preliminary VSP phase.

18.4 SIMULATION

Simulation presented in this section has implemented the evolutionary strategy INDUCTOR in order to induce sets of regexp-like rules from four-word English utterances contained in CHILDES corpus. Diagram elucidating relations between main INDUCTOR components is visible on Figure 32. The simulation was invoked twice, once in 64–dimensional space ($INDUCTOR_{64}$) and once in 128–dimensional space ($INDUCTOR_{128}$).

The vector space preparation phase (c.f. Section 18.3.1) yielded a vector space in which all subsequent INDUCTOR runs took place. Each among 2 * 100 distinct runs of INDUCTOR was initialized by a pseudo-random generation of zeroth population.

18.4.1 CORPUS

This article is conceived as a part of dissertation addressing the possibility of developing evolutionary models of induction of linguistic rules in (and by) human children. This makes the choice of the corpus quite straightforward: the corpus from which we shall aim to extract first linguistic categories is to be contained in Child Language Data Exchange System (CHILDES, (MacWhinney and Snow, 1985)).

Inspired by the "less is more hypothesis" (Elman, 1993), input corpus used in simulation hereby presented consisted of 1047 four-word "motherese"[12] utterances extracted from English section of CHILDES. No other data has been used to guide the inductive process.

18.4.2 PARAMETERS

Table 51: Parameters of diverse components of the INDUCTOR algorithm.

VSP	Input corpus	$CHILDES_{English}$[13]
	Feature Filter	word_juxtaposition
	Dimensionality	$\Delta = 64$ or $\Delta = 128$
	Seed	S = 3
	Reflections	I = 0
INDUCTOR	Population size	N = 100
	Selection	Tournament
	Crossover	One-point
	Mutation rate	M = 0.2%
	Initial population	pseudo-random
	Generations	G = 100
	Elitism	E = 0
	Runs	R = 100
Machine Learning	Syntagms	N = 4

18.5 OBSERVATIONS

?? lists 100 regexp-like rules which have been evaluated as "fittest" at the end of distinct INDUCTOR runs which took place in a H_{64} space. These hundred rules match 176 from 1047 utterances present in the input corpus (16.8%).

?? lists 100 regexp-like rules which have been evaluated as "fittest" at the end of distinct $INDUCTOR_{128}$. These runs took place in a H_{128} space. These hundred rules match 176 from 1047 utterances present in the input corpus (15.8%).

As marked in both Appendices by the token GENERAL, INDUCTOR was also able to identify many completely grammatical 4-schemata which able to accept (or generate) even utterances which have not been present in the input corpus.

Such generalization faculty was observed in 82% resulting individuals in case of H_{64} and in 77% individual $4 - schemata$ induced in H_{128}.

Among these individuals induced in H_{64}, **32** have been manually evaluated as ALLGOOD, id est capable of accepting | generating only grammatically correct utterances of English language.

12 In CHILDES, lines containing motherese utterances begin with the marker *MOT.
13 Available at http://wizzion.com/thesis/simulation3/utterances.4

For example, the most fit schema of sixth run of INDUCTOR$_{64}$:

^(that)(is)(a)(bag | banana | basket | bridge | cherry | cow | gate | horse | kleenex | motorcycle | puzzle | rabbit | raccoon | shoe | spoon | story | timer | tractor)$

is able to accept|generate 18 grammatically correct English utterances in spite of the fact that only 5 among these 18 sentences have been explicitly present in the input corpus.

Excessive over-regularization was observed in case of 21 individuals willing to accept|generate at least one WRONG utterance.

Asides this, 4-schemata issued from 28 runs of INDUCTOR$_{64}$ have been marked as DISPUTABLE. That is, as capable of accepting|generating utterances which would be classified as "ungrammatical" by an orthodox grammarian, but could nonetheless occur in a real-life usage.

This border cases include utterances as:

where is the clever (individual 9)
what are we joey (individual 18)
there is what one (individual 34)
there does he go (individual 55)
what are you joey (individual 83)
oh what is i (individual 87)

as well as utterances which are syntactically correct, but semantically doubtful:

oh you are strawberries (individual 63)
oh you are fries (individual 63)
okay that is thumb (individual 91)

et caetera, et caetera.

In case of INDUCTOR$_{128}$ **37** induced 4—schemata have been manually evaluated as ALLGOOD and 17 as DISPUTABLE.

Listing 18: First exemplar of a non-monotonic ontogenetic trajectory

```
#ITERATION 30 FITNESS 1.333333
^(do )(you )(like )(candy|some|strawberries)$
#ITERATION 40 FITNESS 1.14285714285714
^(do )(you )(like )(bananas|box|candy|cover|fell|ketchup|
nana|not|papa|popsicles|some|sorry|strawberries|tired)$
#ITERATION 50 FITNESS 1.8
^(do )(you )(like )(box|candy|ketchup|some|strawberries)$
```

18.5.1 DIACHRONIC OBSERVATIONS

A deeper time-oriented inspection of processes taking place during individual runs can also be of certain interest.

On Listing 18 it may be seen that after 30 iterations, INDUCTOR$_1$ has identified a 4-schema able to accept | generate utterances "do you like candy", "do you like some" and "do you like strawberries". However, this schema was lost in following 10 generations and fitness fell from 1.33 to 1.14[14]. Hence, an over-regular schema gained in prominence which was able to accept even such constructs as "do you like sorry" or "do you like tired".

But in following ten generations, population dynamics of the whole system not only lead to correction of the previous errors, but even brought about the increase in fitness to 1.8 which went hand in hand with scheme's ability to match utterances like "do you like box" or "do you like ketchup".

Another run presented on Listing 19 also exemplified such non-monotonic, error-correcting aspects of INDUCTOR$_1$ algorithm:

As it may be seen that an incorrect utterance "what is he going" was acceptable by the fittest individual of 40th and 50th iteration. This was corrected in 60th generation but further development brought about yet another batch of mistakes: utterances like "what is he cute" and "what is he share" were thus acceptable by the most fit individual of 80th generation. This has been subsequently corrected and the run terminated, after 100 generations, with a GENERAL, ALLGOOD 4-schema.

Listing 19: Second exemplar of a non-monotonic ontogenetic trajectory

```
#ITERATION 30 FITNESS 1.33333333333333
 ^(what )(is )(he )(doing|playing|saying)$
#ITERATION 40 FITNESS 1.8
 ^(what )(is )(he )(doing|going|holding|playing|saying)$
#ITERATION 50 FITNESS 1.5
 ^(what )(is )(he )(doing|drinking|going|holding|playing|saying)$
#ITERATION 60 FITNESS 1.8
 ^(what )(is )(he )(doing|drinking|holding|playing|saying)$
#ITERATION 70 FITNESS 2.25
 ^(what )(is )(he )(doing|holding|playing|saying)$
#ITERATION 80 FITNESS 2.28571428571429
 ^(what )(is )(he )(called|cute|doing|holding|playing|saying|
     share)$
#ITERATION 90 FITNESS 1.5
 ^(what )(is )(he )(doing|drinking|going|holding|playing|saying)$
#ITERATION 100 FITNESS 2.25
 ^(what )(is )(he )(doing|holding|playing|saying)$
```

14 This is, of course, due to the fact that INDUCTOR$_1$ does not implement any form of elitism which would safeguard the fittest individuals from destructive variations.

18.6 CONCLUSION

Almost one third (32%) of $4-schemata$ - identified by INDUCTOR$_1$ sweeping a 64−dimensional Hamming space representing 1047 English "motherese" utterances - produce only correct generalizations.

Collection of all induced N-schemata yields what we call a "microgrammar". Such a microgrammar is more a as construction-based (Fillmore et al., 1988; Lakoff, 1990) or usage-based (Tomasello, 2009) grammar than a grammar in sense of the Formal Language Theory (P117+122) or in the sense commonly accepted by proponents of the generativist doctrine (Chomsky, 2002).

But given that such a microgrammar (c.f. **??**) is capable of generating more syntactically correct utterances than those which had been presented through the training corpus, one can still consider it to be, in certain regards, modestly generative.

We say "modestly" because the generative faculty is kept on the leash by *evolution's tendency to discard such schemata which would be too concrete (i.e. have low sensitivity* Y*), or too exhaustive (i.e. have high extension* E*)*. Hence, the thorny problem of over-generalization is - at least in case of algorithm implementing the INDUCTOR$_1$ Evolutionary Strategy - not resolved by any *a priori* knowledge embedded in a some kind of chomskyan "Universal Grammar".

Far from it: we propose to depart from the idea that the grammar-inducing agents are not "ideal learners" in sense of Gold's Theorem (Gold, 1967; Johnson, 2004). On the contrary: the process of grammar-induction can only fully succeed if some information-encoding representations are, sometimes, irreversibly forgotten or subsumed to variation.

In this article, variation was attained by operators which:

1. mutate coordinates of centers of syntagmatic $G-categories$
2. mutate radii of syntagmatic $G-categories$ (i.e. increases or decreases category's extension)
3. substitute a $G-categories$ from one $N-schema$ with $G-categories$ from another $N-schema$ (i.e. locus-constrained cross-over)

By causing these operators to perform their operations in a sub-symbolic vector space, and by evaluating results of their activities on a symbol-sequence level, one can obtain a system able to induce simple $4-schema$ micrograms from simplified corpus of English "motherese" utterances which are four words long.

This[15], however, is only the beginning.

15 Proof-of-concept source code of this simulation is available at URL http://wizzion.com/thesis/simulation3/EGI.tgz under mrGPL license.

18.7 GENERAL DISCUSSION

There is an appealing symmetry in the notion that the mechanisms of natural learning may resemble the processes that created the species possessing those learning processes.

— D.E. Goldberg and J. Holland

More generally and beyond syntax, operators implemented in the 3rd simulation can be associated to following psychological phenomena:

1. mutation of an N-schema - synaptic pruning (P+38), information decay, forgetting etc.
2. crossover between two N-schemata - related to creativity, dreaming (P+89-90) and phantasia

Other variation operators - corresponding to certain forms of

1. playing certain *language games* (Wittgenstein, 1953; Nowak et al., 1999), or "intrapsychic" (Brams, 2011) games
2. imitating certain phenomena observed in linguistic behavior of human children (P+184-204)

could also be deployed.

Another subsequent enhancement of the GI method hereby introduced could potentially result from introduction of additional feature sets. For example, one could take a fit N−schema X, decompose it into its component G−categories $G_1, G_2, ..., G_N$ and, if ever a certain component G−category G_α turns out to be disjunctive, enrich vectorial representations of all its members with information that they belong to G_α. For example, one could enrich vectorial representations of tokens "doing", "holding", "playing", "saying" with information that they turned out to be subsumed under $G - category$ present in one quite fit $4 - schema$ (c.f. Listing 18). And enrich vectorial representations of tokens "ketchup", "strawberries" etc. with information that these tokens turned out to subsumed by yet another $G - category$ present in another schema (c.f. Listing 19).

Note that introduction of such feature-sets could be interpreted as introduction of a feedback-loop in the system. Essence of such a system could thus be considered to be not only linguistic, but also cybernetic (Wiener, 1961; Lorenz, 1973). It could be postulated that introducing of such feed-back, bootstrapping (Hromada, 2014b; Karmiloff and Karmiloff-Smith, 2009, pp.111-118) loop into the system would not only result in identification of more complex micrograrmmars, but would also cause the system to follow similar ontogenetic trajectories than those of children which undergo a so-called syntagmatic-paradigmatic shift (Nelson, 1977).

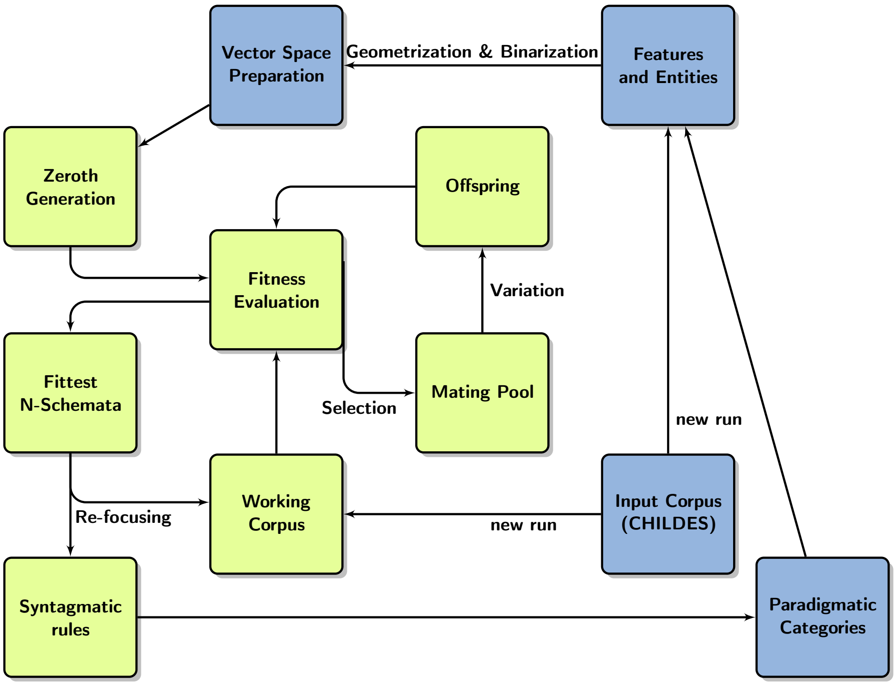


Figure 33: Data flow among main components of extended variant of INDUCTOR introducing a syntagmatic-paradigmatic feedback loop.

All such operators, features and feedback-loops taken together and coupled with

1. the fact that brain (P+5) is a finite material object with finite resources which is subjected to 2nd law of thermodynamics (P+7)
2. the fact that linguistic input which the child becomes is preprocessed by loving (P+241) and caring computational oracles (Turing, 1939; Clark, 2010) like mothers, fathers, care-takers etc.
3. the fact that acquisition of language takes place in informationally very rich, contextually grounded, usage-based scenarios (Tomasello, 2009)

one cannot exclude that a sort of evolutionary, ecological, equilibrium-seeking process indeed takes place in the mind of a modal healthy language-acquiring toddler.

And given that certain high-profile developmental linguists terminate their inquiry, concerning the informatic properties of the language input, with the conclusion

« internal mechanisms are necessary to account for the unlearning of ungrammatical utterances» (Marcus, 1993)

we allow us to conclude with a suggestion that the *internal mechanism* which Marcus mentions is, in reality, not a sort of universal grammar (P+98-101) black-box but instead a potentially "general cognitive process" (P+101, (Piaget, 1974)) whose very essence is to discard that, which is non-functional:

Evolution (P+3).

Part V

SUMMA

19 SUMMA SUMMARUM

The natural selection paradigm of such knowledge increments can be generalized to other epistemic activities, such as learning, thought and science.

— D.T. Campbell

The objective of this dissertation was to provide a computational evidence of the "operational thesis" (P+20):

«Learning of toddlerese can be successfully simulated by means of evolutionary algorithm processing textual representations of motherese.»

Given that

- the third simulation used no other input than the plain-text corpus of motherese utterances

and given that

- the third simulation resulted in identification of schemata able to generate grammatically correct utterances which have not been present in the initial corpus

A Popperian conclusion

one may consider the "operational thesis" as temporarily unfalsified.

In this sense, we consider any future effort to falsify or verify "the softest thesis" (P+17-19):

«Ontogeny of toddlerese can be successfully simulated by means of evolutionary computation.»

as effort worthy of interest.

It is worthy of noting in this regards that certain notions like that of a $4-\mathrm{schema}$ or *morphosemantic class* are not to be considered as some ultimate elements of some sort of *ewige Theorie* but rather as temporary, limited building blocks of an architecture which is to be surpassed.

Surpassed by what? Maybe surpassed by models which introduce not only $4-\mathrm{schemata}$ but also $2-\mathrm{schemata}$, $3-\mathrm{schemata}$, $5-\mathrm{schemata}$ ... $N-\mathrm{schemata}$. Or by procedures which integrate semantic, morphological and syntactic spaces within a single "linguistic" space S_L. Given what we have seen until now, it can not be a priori excluded that results of certain types of evolution-inspired simulations taking place within such S_L would turn out to be consistent with "the softer hypothesis" (P+14-16) which states that «learning of natural language can be successfully simulated by means of evolutionary computation»

But when speaking about optimization taking place within a linguistic space S_L, shouldn't it be also possible to speak also about optimization taking place within even more generic a space S_G ? For nothing prohibits that category-inducing methods hereby introduced could be used to induce classifiers of partially or even fully non-linguistic entities. For example, a research project stemming from this dissertation may potentially explore the extent in which the evolutionary search for prototypes could be useful in Computer Vision: the only thing which would be fundamentally different would be the essence of input entities (i.e. images and not texts) and features occurring in such entities (e.g. Haar features (**?**Hromada, 2010c; Hromada et al., 2010) or others).

Relation to Computer Vision

In fact, nothing forbids to use one among three CI models hereby introduced whenever one needs to perform:

1. multiclass classification of entities (exemplified by "supervised" simulations 1 and 2)

2. induction of rules from positive corpus only (exemplified by "unsupervised" simulation 3)

In other terms, the combination of "vector spaces" and "evolutionary computation" components can be understood as a "generic optimization toolbox" (GOT) which could potentially be applied upon any set of features. It is, however, primarily the nature of the input corpus and the nature of features which extracted from the corpus which should most closely determine the nature of categorization-performing agent thus induced.

Generic Optimization Toolbox

Hence, when applied upon data-sets describing "spatial" trajectories within a group of "labyrinths", one could aspire to induce rules allowing a certain robot, a certain automatized vehicle, or a certain sort of embedded artificial classifier system (Booker et al., 1989), to find its way out of the "labyrinth" it never saw before.

Induction of spatial trajectories

Or - if one would depart from so-called "morally relevant features" (Hromada and Gaudiello, 2014) - one could even hope to simulate ontogeny of categories and rules of a somewhat different kind. That is, of categories and rules which are commonly labeled as "aesthetic" (i.e. beautiful / ugly), "moral" (i.e. good / bad), "deontologic" (i.e. forbidden / allowed) (Hromada, 2016f).

Moral Induction

Asides "linguistic", "visual", "spatial" or "moral", implementation of EML GOTs in induction of other types of intelligence (Gardner, 2011) or their combinations (Karpathy and Fei-Fei, 2015) in artificial agents and robots is also a task to be explored. If successful, it cannot be excluded that such explorations would potentially bring scientific and engineering communities one step closer us to deployment meta-modular (Hromada, 2012a) artificial agents able to:

EML and theory of multiple intelligences

1. integrate (Tononi, 2004) multi-modal (i.e. linguistic, visual, proprioceptive etc.) information

2. use nature-inspired, evolutionary computational core to identify most fit groupings of such information

By doing so, an ultimate *ex computatio atque simulatio* proof of the "soft thesis" (P+11-13):

« learning can be successfully simulated
by means of evolutionary computation »

could be, potentially, given.

To offer such a proof, however, is a task which by far surpasses limits of any individual researcher. What is more, alternative machine learning paradigms (e.g. deep learning) currently predominate and it may be the case that popularity of such approaches decreases the amount of attention which could - and should - be focused on exploration of common grounds between computational models of learning and computational models of evolution.

Let's now enumerate certain advantageous properties of evolutionary machine learning (EML) models which have been presented in simulations one, two and three. These EML models are :

Evolutionary Machine Learning and its advantages

1. **functional**: function of the model is principally determined by choice of fitness function and selection/variation operators

2. **alternative**: in any moment T_X, the learning system contains multiple alternative solutions of the problem (P+8-10)

3. **population-based**: behavior of the learning system can be interpreted in terms of population dynamics (P+116)

Contrary to these, connectionist models are more "structural" than "functional", they do not explicitly encode representations of diverse solutions and their convergence towards optimal states is more easily interpretable in terms of differential "gradient descent" of "backpropagation" than in terms of population dynamics.

Comparison with connectionist models

What's more, by coupling the notion of evolution with that of a vector space, and by implementing a fairly trivial phenotype - genotype transcription (Section 18.3.2), one can obtain unsupervised EML models

1. bridging sub-symbolic (vectorial) and symbolic (regexps and grammars) realms

2. transparent to investigation and modulation by a human investigator (i.e. easy to interpret and *teach*)

Note that the property of being transparent to investigation and modulation is not a property which should be taken *à la légère*. For it could result in a creation of the inter-subjective bound between the artificial system which is being (investig | modul)ated and the human who (investig | modul)ates.

In other terms, it could, potentially, result in emergence of entities of non-organic origin *who* could, and should, be considered as not only *objects of machine-learning* but also as *subjects of machine-teaching*.

Of learners and teachers

Such considerations, however, bring us further than paradigms like machine learning or even computer science could ever bring us. Such considerations bring us towards meta-paradigm[1] of paedagogy and didactics (Komenskỳ et al., 1991) which solely can demonstrate the validity and usefulness of the Theory of Intramental Evolution (Hromada, 2015).

Such considerations bring us towards such regions of S_G whereby the very "hard thesis" (P+2-10)

«Learning is a form of evolution»

could be evaluated as valid.

Valid or not, nothing forbids the sign-manipulating[2] mind (P+1) to realize a transposition (P+190-192) which savants like Bateson (Bateson, 2006) once realized.

That is, a transposition between two terms each of which denote one big stochastic system, a transposition between "Mind" and "Nature", a transposition which obliges one to state:

Ultimate Chiasm

«Evolution is a form of learning[3]»

Such is, indeed, the ultimate result of the dissertation with which we aspire for attribution of the title *Philosophiae Doctor* in both cybernetics as well as cognitive psychology.

Such is, indeed, the result of work commenced by two words forming the "initial thesis" (P+1):

«Mind Evolves»

*
* *

1 A scientific paradigm (Kuhn, 2012) transfers knowledge about certain field of study. A scientific meta-paradigm transfers knowledge concerning the transfer of knowledge.

2 « Thinking is essentially the activity of operating with signs.» (Wittgenstein, 1934)

3 Lorenz (1973) states that the principal difference between learning and evolution is the ability of a learning system to "learn from one's own errors". System which learns is supposed to have such ability while system which "only" evolves does not. But is it really always the case?

BIBLIOGRAPHY

Adler, A. (1976). *Connaissance de l'homme*. Payot.

Amancio, D. R., Altmann, E. G., Rybski, D., Oliveira Jr, O. N., and Costa, L. d. F. (2013). Probing the statistical properties of unknown texts: application to the voynich manuscript. *PloS one*, 8(7):e67310.

Ambridge, B., Theakston, A. L., Lieven, E. V., and Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development*, 21(2):174–193.

Araujo, L. (2002). Part-of-speech tagging with evolutionary algorithms. In *Computational Linguistics and Intelligent Text Processing*, pages 230–239. Springer.

Aristotle (-335 BC). *Poetics: On Comedy*. Unknown.

Aristotle (342BC). *On Coming-to-be & Passing-way*. At the Clarendon Press.

Atkinson, Q. D. and Gray, R. D. (2005). Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic biology*, 54(4):513–526.

Augustine, S. (1838). *Confessions. Book I.*

Aurenhammer, F. (1991). Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405.

Aycinena, M., Kochenderfer, M. J., and Mulford, D. C. (2003). An evolutionary approach to natural language grammar induction. *Final Paper Stanford CS224N June*.

Baixeries, J., Elvevåg, B., and Ferrer-i Cancho, R. (2013). The evolution of the exponent of zipf's law in language ontogeny. *PloS one*, 8(3):e53227.

Bandura, A. and McClelland, D. C. (1977). Social learning theory.

Barrett, D. (2007). *Waistland: A (R) evolutionary View of Our Weight and Fitness Crisis*. WW Norton & Company.

Barrett, M. D. (1978). Lexical development and overextension in child language. *Journal of child language*, 5(02):205–219.

Bateson, G. (2006). Mind and nature: A necessary unity (advances in systems theory, complexity, and the human sciences).

Bee, H. L. and Boyd, D. R. (2000). *The developing child*. Allyn and Bacon Boston.

Bellegarda, J. R. (2005). Unsupervised, language-independent grapheme-to-phoneme conversion by latent analogy. *Speech Communication*, 46(2):140–152.

Bentley, P. (1999). *Evolutionary design by computers*. Morgan Kaufmann.

Best, K.-H. (2006). Quantitative linguistik: Eine annäherung. 3., stark überarbeitete und ergänzte auflage.

Blackmore, S. (2000). *The meme machine*. Oxford University Press.

Booker, L. B., Goldberg, D. E., and Holland, J. H. (1989). Classifier systems and genetic algorithms. *Artificial intelligence*, 40(1):235–282.

Borges, J. L. (1952). El idioma analítico de john wilkins. *Otras inquisiciones*, pages 158–159.

Braine, M. D. (1971). On two types of models of the internalization of grammars. *The ontogenesis of grammar*, pages 153–186.

Braine, M. D. and Bowerman, M. (1976). Children's first word combinations. *Monographs of the society for research in child development*, pages 1–104.

Brams, S. J. (2011). *Game theory and the humanities: bridging two worlds*. MIT Press.

Brighton, H., Kirby, S., and Smith, K. (2003). Situated cognition and the role of multi-agent models in explaining language structure. In *Adaptive agents and multi-agent systems*, pages 88–109. Springer.

Broca, P. (1861). {Remarque sur le siege de la facult\\'{e} du language articul\\'{e}, suivie d\'une observation d\'aph\\'{e} mie (perte de la parole)}. *{Bulletin de la soci\\'{e} t\\'{e} anatomique de Paris}*, 36:330–356.

Brodsky, P., Waterfall, H., and Edelman, S. (2007). Characterizing motherese: On the computational structure of child-directed language. In *Proceedings of the 29th Cognitive Science Society Conference, ed. DS McNamara & JG Trafton*, pages 833–38.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., and Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational linguistics*, 18(4):467–479.

Brown, R. (1958). Words and things.

Brown, R. (1973). *A first language: The early stages.* Harvard U. Press.

Bruner, J. S. and Watson, R. (1983). *Child's talk: Learning to use language*. Oxford University Press Oxford.

Bryant, E. F. (2009). The yoga sutras of patanjali.

Buber, M. (1937). I and thou. *Clark, Edinburgh*.

Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological review*, 67(6):380.

Campbell, D. T. (1974). An essay on evolutionary epistemology. *The philosophy of Karl Popper*, pages 413–463.

Champollion, J. F. (1822). *Observations sur l'obelisque Egyptien de l'Ile de Philae*.

Chomsky, N. (1957). *Syntactic structures*. Mouton.

Chomsky, N. (1959). A review of bf skinner's verbal behavior. *Language*, 35(1):26–58.

Chomsky, N. (1995). *The minimalist program*, volume 28. Cambridge Univ Press.

Chomsky, N. (2002). *Syntactic structures*. Walter de Gruyter.

Christodoulopoulos, C., Goldwater, S., and Steedman, M. (2010). Two decades of unsupervised pos induction: How far have we come? In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 575–584. Association for Computational Linguistics.

Clark, A. (2010). Distributional learning of some context-free languages with a minimally adequate teacher. In *Grammatical Inference: Theoretical Results and Applications*, pages 24–37. Springer.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, pages 1–33.

Clark, E. V. (2003). *First Language Acquisition*. Cambridge University Press.

Cohen, T., Schvaneveldt, R., and Widdows, D. (2010). Reflective random indexing and indirect inference: A scalable method for discovery of implicit connections. *Journal of Biomedical Informatics*, 43(2):240–256.

Cohen, T., Widdows, D., Schvaneveldt, R. W., Davies, P., and Rindflesch, T. C. (2012). Discovering discovery patterns with predication-based semantic indexing. *Journal of biomedical informatics*, 45(6):1049–1065.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cosmides, L. and Tooby, J. (1997). Evolutionary psychology: A primer. *Evolutionary Psychology: a primer*.

Currier, P. (1970). 1976." voynich ms. transcription alphabet; plans for computer studies; transcribed text of herbal a and b material; notes and observations.". *Unpublished communications to John H. Tiltman and M. D'Imperio, Damariscotta, Maine*.

Darwin, C. (1859). *The Origin of Species*. J. Murray.

Darwin, C. and Bettany, G. T. (1890). *Journal of researches into the natural history and geology of the countries visited during the voyage of HMS" Beagle" round the world: under the command of Capt. Fitz Roy, RN*. Ward, Lock.

Datar, M., Immorlica, N., Indyk, P., and Mirrokni, V. S. (2004). Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 253–262. ACM.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press Oxford.

De Chardin, P. T., Wall, B., et al. (1965). *The phenomenon of man*, volume 383. Harper & Row New York, NY, USA:.

de Saussure, F. (1916). *Cours de la linguistique générale*.

DeCasper, A. J. and Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant behavior and Development*, 9(2):133–150.

Dehaene, S. and Changeux, J.-P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience*, 1(3):244–261.

Dennett, D. C. (1995). Darwin's dangerous idea. *The Sciences*, 35(3):34–40.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., and Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, 32(04):759–786.

d'Imperio, M. E. (1978). The voynich manuscript: an elegant enigma. Technical report, DTIC Document.

Dubremetz, M. (2013). Vers une identification automatique du chiasme de mots. *TALN-RÉCITAL 2013*, page 150.

Dupont, P. (1994). Regular grammatical inference from positive and negative samples by genetic search: the gig method. In *Grammatical Inference and Applications*, pages 236–245. Springer.

Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection.* Basic Books.

Elbers, L. and Ton, J. (1985). Play pen monologues: the interplay of words and babbles in the first words period. *Journal of Child Language*, 12(03):551–565.

Ellis, R. and Wells, G. (1980). Enabling factors in adult-child discourse. *First Language*, 1(1):46–62.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1):71–99.

Erjavec, T. (2004). Multext-east version 3: Multilingual morphosyntactic specifications, lexicons and corpora. In *LREC*.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., and Stiles, J. (1994). Variability in early communicative development. *Monographs of the society for research in child development*, pages i–185.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language*, pages 419–439.

Fernando, C., Szathmáry, E., and Husbands, P. (2012). Selectionist and evolutionary approaches to brain function: a critical appraisal. *Frontiers in computational neuroscience*, 6.

Ferrer-i Cancho, R. and Elvevåg, B. (2010). Random texts do not exhibit the real zipf's law-like rank distribution. *PLoS One*, 5(3):e9411.

Fillmore, C. J., Kay, P., and O'connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, pages 501–538.

Fisher, R. A. (1925). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Flake, G. W. (1998). *The computational beauty of nature: Computer explorations of fractals, chaos, complex systems, and adaptation*. MIT press.

Floridi, L. (2011). *The Philosophy of Information*. Oxford University Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Fogel, D. B. (1995). Phenotypes, genotypes, and operators in evolutionary computation. In *Evolutionary Computation, 1995., IEEE International Conference on*, volume 1, page 193. IEEE.

Fogel, L. J., Owens, A. J., and Walsh, M. J. (1966). Artificial intelligence through simulated evolution.

Foiter, M. L. (2002). Symbolism: The foundation of culture. *Companion Encyclopedia of Anthropology*, page 366.

Fraisse, P. (1974). *Psychologie du rythme*. Presses universitaires de France Paris.

Frege, G. (1994). Über sinn und bedeutung. *Wittgenstein Studien*, 1(1).

Furrow, D., Nelson, K., and Benedict, H. (1979). Mothers' speech to children and syntactic development: Some simple relationships. *Journal of child language*, 6(03):423–442.

Galton, F. (1875). *English men of science: Their nature and nurture*. D. Appleton.

Gärdenfors, P. (2004). *Conceptual spaces: The geometry of thought*. MIT press.

Gardner, H. (1985a). *Frames of mind: The theory of multiple intelligences*. Basic books.

Gardner, H. (1985b). *The mind's new science*. Basic Books.

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. Basic books.

Gertner, S., Greenbaum, C. W., Sadeh, A., Dolfin, Z., Sirota, L., and Ben-Nun, Y. (2002). Sleep–wake patterns in preterm infants and 6 month's home environment: implications for early cognitive development. *Early Human Development*, 68(2):93–102.

Gödel, K. (1931). Über formal unentscheidbare sätze der principia mathematica und verwandter systeme i. *Monatshefte für mathematik und physik*, 38(1):173–198.

Gold, E. M. (1967). Language identification in the limit. *Information and control*, 10(5):447–474.

Goldberg, D. E. (1990). Genetic algorithms in search, optimization & machine learning. *Addison-Wesley*.

Goldberg, D. E. and Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3:95–99.

Gómez, R. L. (2011). Memory, sleep and generalization in language acquisition. *Experience, Variation and Generalization: Learning a First Language*, 7:261.

Grice, H. (1975). Logic and conversation'in p. cole and j. morgan (eds.) syntax and semantics volume 3: Speech acts.

Guermeur, Y. and Monfrini, E. (2011). A quadratic loss multi-class svm for which a radius–margin bound applies. *Informatica*, 22(1):73–96.

Haeckel, E. (1879). The evolution of man. *London: Kegan Paul*.

Hamilton, W. D. (1963). The evolution of altruistic behavior. *American naturalist*, pages 354–356.

Harris, M. (2013). *Language experience and early language development: From input to uptake*. Psychology Press.

Harris, Z. S. (1954). Distributional structure. *Word*.

Hebb, D. O. (1964). *The Organization of Behaviour: A Neuropsychological Theory*. John Wiley and Sons.

Hesse, H. (1967). *Das Glasperlenspiel: Versuch e. Lebensbeschreibung d. Magisters Ludi Josef Knecht samt Knechts hinterlassenen Schriften*, volume 842. Suhrkamp.

Hodgins, G. (2014). Forensic investigations of the voynich ms. In *Voynich 100 Conference www. voynich. nu/mon2012/index. html. Accessed*, volume 4.

Hofmann, T., Schölkopf, B., and Smola, A. J. (2008). Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220.

Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* U Michigan Press.

Holland, J. H. (1992). Genetic algorithms. *Scientific american*, 267(1):66–72.

Holly Smith, B., Crummett, T. L., and Brandt, K. L. (1994). Ages of eruption of primate teeth: a compendium for aging individuals and comparing life histories. *American Journal of Physical Anthropology*, 37(S19):177–231.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558.

Householder, F. W. (1981). *Apollonius Dyscolus: The Syntax of Apollonius Dyscolus*, volume 23. John Benjamins Publishing.

Hromada, D. (2008). 23 comments to the chomskian doctrine. personal communication with D.Sportiche.

Hromada, D. (2010a). Quantitative intercultural comparison by means of parallel pageranking of diverse national wikipedias.

Hromada, D. D. (2009). Basen o jablku. Master's thesis, Faculty of Humanities, Charles University, Prague, Czech Republic.

Hromada, D. D. (2010b). Concepts of «invasivity» and «reversibility» and their relation to past, present and future techniques of neural imagery. Course work written for Bordeaux team of Neural Imagery affiliated to Ecole Pratique des Hautes Etudes.

Hromada, D. D. (2010c). smiled : Sourire naturel et sourire artificiel. de l'utilisation d'opencv pour le tracking, la reconnaissaince des expressions faciales et la détection du sourire. Master's thesis, Ecole Pratique des Hautes Etudes, Paris, France.

Hromada, D. D. (2011). Initial experiments with multilingual extraction of rhetoric figures by means of perl-compatible regular expressions. In *RANLP Student Research Workshop*, pages 85–90.

Hromada, D. D. (2012a). From age&gender-based taxonomy of turing test scenarios towards attribution of legal status to meta-modular artificial autonomous agents. page 7.

Hromada, D. D. (2012b). Variations upon the theme of evolutionary language game. Written for prof. Vladimir Kvasnicka, downloadable at http://wizzion.com/papers/2012/.

Hromada, D. D. (2013). Random projection and geometrization of string distance metrics. In *RANLP*, pages 79–85.

Hromada, D. D. (2014a). Comparative study concerning the role of surface morphological features in the induction of part-of-speech categories. In *Text, Speech and Dialogue*, pages 46–52. Springer.

Hromada, D. D. (2014b). Conditions for cognitive plausibility of computational models of category induction. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 93–105. Springer.

Hromada, D. D. (2014c). Empiric introduction to light stochastic binarization. In *Text, Speech and Dialogue*, pages 37–45. Springer.

Hromada, D. D. (2014d). Geometrizacia ontologii - pripadova studia snomed. Written for doc. Mikulas Popper.

Hromada, D. D. (2015). Genetic optimization of semantic prototypes for multiclass document categorization. submitted to Elitech 2015 conference.

Hromada, D. D. (2016a). Can evolutionary computation help us to crib the voynich manuscript ? Submitted to JADT 2016 conference.

Hromada, D. D. (2016b). Evolutionary induction of 4-schema micro-grammars from childes corpora. submitted to journal Evolutionary Computation.

Hromada, D. D. (2016c). Evolutionary induction of a lightweight morphosemantic classifier. submitted to Computational Linguistics.

Hromada, D. D. (2016d). *Evolutionary Models of Ontogeny of Linguistic Categories: Four Simulations*. PhD thesis, Slovak Technical University and University Paris Lumieres.

Hromada, D. D. (2016e). Fast and frugal retrieval of linguistic universalia from childes transcripts. Submitted to JADT 2016 conference.

Hromada, D. D. (2016f). Narrative fostering of morality in artificial agents: Constructivism, machine learning and story-telling. In *L'esprit au-delà du droit: Pour un dialogue entre les sciences cognitives et le droit*. Mare et Martin.

Hromada, D. D. and Gaudiello, I. (2014). Introduction to moral induction model and its deployment in artificial agents. In *Sociable Robots and the Future of Social Relations*, pages 209–216. IOS Press.

Hromada, D. D., Tijus, C., Poitrenaud, S., and Nadel, J. (2010). Zygomatic smile detection: The semi-supervised haar training of a fast and frugal system: A gift to opencv community. In *Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), 2010 IEEE RIVF International Conference on*, pages 1–5. IEEE.

Huizinga, J. (1956). Homo ludens vom ursprung der kultur im spiel.

Imai, M. and Haryu, E. (2001). Learning proper nouns and common nouns without clues from syntax. *Child development*, 72(3):787–802.

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

Jakobson, R. (1960). Why "mama" and "papa". *Essays in honor of Heinz Werner*.

Jiménez López, M. D. et al. (2000). Gramar systems: a-formal-language-theoretic framework for linguistics and cultural evolution.

Johnson, K. (2004). Gold's theorem and cognitive science*. *Philosophy of Science*, 71(4):571–592.

Jones, W. (1788). The third anniversary discourse, delivered 2 february, 1786. *Asiatick Researches*, 1:415–431.

Jusczyk, P. W. and Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive psychology*, 29(1):1–23.

Jusczyk, P. W., Cutler, A., and Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of english words. *Child development*, 64(3):675–687.

Karmiloff, K. and Karmiloff-Smith, A. (2009). *Pathways to language: From fetus to adolescent*. Harvard University Press.

Karpathy, A. and Fei-Fei, L. (2014). Deep visual-semantic alignments for generating image descriptions. *arXiv preprint arXiv:1412.2306*.

Karpathy, A. and Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3128–3137.

Karypis, G. (2002). Cluto-a clustering toolkit. Technical report, DTIC Document.

Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. Oxford University Press.

Kelemen, J. (2004). Miracles, colonies, and emergence. In *Formal Languages and Applications*, pages 323–333. Springer.

Kelemenová, A. and Csuhaj-Varjú, E. (1994). Languages of colonies. *Theoretical Computer Science*, 134(1):119–130.

Keller, B. and Lutz, R. (1997). Evolving stochastic context-free grammars from examples using a minimum description length principle. In *1997 Workshop on Automata Induction Grammatical Inference and Language Acquisition*. Citeseer.

Kennedy, G. and Churchill, R. (2005). *The Voynich manuscript: the unsolved riddle of an extraordinary book which has defied interpretation for centuries*. Orion Publishing Company.

Kennedy, J., Kennedy, J. F., and Eberhart, R. C. (2001). *Swarm intelligence*. Morgan Kaufmann.

Keysers, C. and Perrett, D. I. (2004). Demystifying social cognition: a hebbian perspective. *Trends in cognitive sciences*, 8(11):501–507.

Komenský, J. A., Okál, M., and Pšenák, J. (1991). *Vel'ká didaktika: Didactica magna*. Slovenské pedagogické nakladatel'stvo.

Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection*, volume 1. MIT press.

Kuczaj, S. A. and Maratsos, M. P. (1975). What children can say before they will. *Merrill-Palmer Quarterly of Behavior and Development*, pages 89–111.

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

Küntay, A. and Slobin, D. I. (1996). Listening to a turkish mother: Some puzzles for acquisition. *Social interaction, social context, and language: Essays in honor of Susan Ervin-Tripp*, pages 265–286.

Küntay, A. and Slobin, D. I. (2002). Putting interaction back into child language: Examples from turkish. *Psychology of Language and Communication*, 6(1).

Kvasnicka, V. and Pospichal, J. (1999). An emergence of coordinated communication in populations of agents. *Artificial Life*, 5(4):319–342.

Kvasnicka, V. and Pospichal, J. (2007). Evolúcia jazyka a univerzální darwinizmus. *Mysel, inteligencia a zivot*.

Labov, W. and Labov, T. (1978). The phonetics of cat and mama. *Language*, pages 816–852.

Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*. Cambridge Univ Press.

Lama, D. et al. (2005). *In the Buddha's words: An anthology of discourses from the Pali Canon*. Simon and Schuster.

Landauer, T. K. and Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

Landini, G. and Zandbergen, R. (1998). A well-kept secret of mediaeval science: The voynich manuscript. *Aesculapius*, 18:77–82.

Lashley, K. (1950). In search of the engram. *Symposia of the Society for Experimental Biology*.

Lauer, F. and Guermeur, Y. (2011). Msvmpack: a multi-class support vector machine package. *The Journal of Machine Learning Research*, 12:2293–2296.

Li, W. (1992). Random texts exhibit zipf's-law-like word frequency distribution. *Information Theory, IEEE Transactions on*, 38(6):1842–1845.

Lieven, E. V., Pine, J. M., and Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of child language*, 24(01):187–219.

Lorenz, K. (1973). *Die Rückseite des Spiegels*. R. Piper.

Lotka, A. J. (1925). Elements of physical biology.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. California, USA.

MacWhinney, B. (1987). The competition model. *Mechanisms of language acquisition*, pages 249–308.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume I: Transcription format and programs*. Psychology Press.

MacWhinney, B. and Snow, C. (1985). The child language data exchange system. *Journal of child language*, 12(02):271–295.

MacWhinney, B. and Snow, C. (1991). Childes manual.

Maratsos, M. (1988). The acquisition of formal word classes. *Categories and processes in language acquisition*, pages 31–44.

Marchman, V. A. and Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of child language*, 21(02):339–366.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, 46(1):53–85.

Markman, E. M. and Hutchinson, J. E. (1984). Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations. *Cognitive psychology*, 16(1):1–27.

Maynard Smith, J. (1986). *The problems of biology*, volume 144. Oxford: Oxford University Press.

McAuley, J. D., Jones, M. R., Holub, S., Johnston, H. M., and Miller, N. S. (2006). The time of our lives: life span development of timing and event tracking. *Journal of Experimental Psychology: General*, 135(3):348.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., and Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition*, 29(2):143–178.

Menyuk, P., Liebergott, J., Schultz, M., Chesnick, M., and Ferrier, L. (1991). Patterns of early lexical and cognitive development in premature and full-term infants. *Journal of Speech, Language, and Hearing Research*, 34(1):88–94.

Miller, G. (1956). The magic number seven plus or minus two: Some limits on our automatization of cognitive skills. *Psychological Review*, 63:81–97.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

Mink, J. and Blumenschine, R. (1981). Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. *Am J Physiol*, 241(3):R203–R212.

Minsky, M. and Papert, S. (1969). Perceptrons.

Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2012). *Foundations of machine learning*. MIT press.

Morgan, J. L. and Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child development*, 66(4):911–936.

Morgan, T. H. (1916). *A Critique of the Theory of Evolution*. Princeton University Press.

Mouillot, D. and Lepretre, A. (2000). Introduction of relative abundance distribution (rad) indices, estimated from the rank-frequency diagrams (rfd), to assess changes in community diversity. *Environmental monitoring and assessment*, 63(2):279–295.

Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological bulletin*, 84(1):93.

Nelson, K. (2006). *Narratives from the crib*. Harvard University Press.

Newbold, W. R. (1928). *Cipher of Roger Bacon*.

Nowak, M. A., Plotkin, J. B., and Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, 200(2):147–162.

Ofria, C. and Wilke, C. O. (2004). Avida: A software platform for research in computational evolutionary biology. *Artificial life*, 10(2):191–229.

O'Neil, M. and Ryan, C. (2003). *Grammatical evolution*. Springer.

Pagel, M., Atkinson, Q. D., Calude, A. S., and Meade, A. (2013). Ultraconserved words point to deep language ancestry across eurasia. *Proceedings of the National Academy of Sciences*, 110(21):8471–8476.

Páleš, E. (1994). Sapfo–parafrázovač slovenčiny. *Veda vydavateľstvo SAV*.

Piaget, J. (1947). La psychologie de l'intelligence.

Piaget, J. (1965). *The Moral Judgment of the Child*. The free press.

Piaget, J. (1974). Introduction à l'épistémologie génétique. *Paris, PUF*.

Piatelli-Palmarini, M. (1980). Language and learning: The debate between jean piaget and noam chomsky.

Pine, J. M. and Lieven, E. V. (1997). Slot and frame patterns and the development of the determiner category. *Applied psycholinguistics*, 18(02):123–138.

Pinker, S. (1994). *The language instinct: The new science of language and mind*, volume 7529. Penguin UK.

Pinker, S. (2000). Survival of the clearest. *Nature*, 404(6777):441–442.

Planck, M. (1926). Über die begründung des zweiten hauptsatzes der thermodynamik. *Sitzungsberichte der Preussischen Akademie der Wissenschaarticle*.

Plato (380BC). *Republic*.

Pohlheim, H. (1996). Geatbx: Genetic and evolutionary algorithm toolbox for use with matlab documentation. *Online]. http://www. geatbx. com/docu/algindex. html.(Accessed May, 2004)*.

Poincaré, H. (1908). *L'invention mathématique*.

Poincaré, H. and Magini, R. (1899). Les méthodes nouvelles de la mécanique céleste. *Il Nuovo Cimento (1895-1900)*, 10(1):128–130.

Popper, K. R. (1972). *Objective knowledge: An evolutionary approach*. Clarendon Press Oxford.

Price, G. R. et al. (1970). Selection and covariance. *Nature*, 227:520–21.

Provasi, J., Anderson, D. I., and Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in psychology*, 5.

Ray, T. S. (1992). Evolution, ecology and optimization of digital organisms. *Santa Fe*.

Rechenberg, I. (1971). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Dr.-Ing*. PhD thesis, Thesis, Technical University of Berlin, Department of Process Engineering.

Rizzolatti, G., Sinigaglia, C., and Anderson, F. T. (2008). *Mirrors in the brain: How our minds share actions and emotions.* Oxford University Press.

Roffwarg, H. P., Muzio, J. N., and Dement, W. C. (1966). Ontogenetic development of the human sleep-dream cycle. *Science*.

Rosch, E. (1999). Principles of categorization. *Concepts: core readings*, pages 189–206.

Rosch, E. and Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4):573–605.

Rosenberg, A. and Hirschberg, J. (2007). V-measure: A conditional entropy-based external cluster evaluation measure. In *EMNLP-CoNLL*, volume 7, pages 410–420.

Rudolph, G. (1994). Convergence analysis of canonical genetic algorithms. *Neural Networks, IEEE Transactions on*, 5(1):96–101.

Rugg, G. (2004). An elegant hoax? a possible solution to the voynich manuscript. *Cryptologia*, 28(1):31–46.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., and Wintner, S. (2007). High-accuracy annotation and parsing of childes transcripts. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*, pages 25–32. Association for Computational Linguistics.

Sahlgren, M. (2005). An introduction to random indexing. In *Methods and applications of semantic indexing workshop at the 7th international conference on terminology and knowledge engineering, TKE*, volume 5.

Salakhutdinov, R. and Hinton, G. (2009). Semantic hashing. *International Journal of Approximate Reasoning*, 50(7):969–978.

Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210.

Schinner, A. (2007). The voynich manuscript: evidence of the hoax hypothesis. *Cryptologia*, 31(2):95–107.

Schleicher, A. (1873). *Die Darwinsche theorie und die sprachwissenschaft: Offenes sendschreiben an herrn Ernst Häckel*, volume 2. Böhlau.

Schmidt, J. (1872). *Die verwantschaftsverhältnisse der indogermanischen sprachen*. Böhlau.

Schwartz, R. G. and Leonard, L. B. (1982). Do children pick and choose? an examination of phonological selection and avoidance in early lexical acquisition. *Journal of child language*, 9(02):319–336.

Schwartz, R. G. and Terrell, B. Y. (1983). The role of input frequency in lexical acquisition. *Journal of child language*, 10(01):57–64.

Sekaj, I. (2004). Robust parallel genetic algorithms with reinitialisation. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 411–419. Springer.

Sekaj, I. (2005). *Evolučné vỳpočty a ich využitie v praxi*. Iris.

Shannon, C. E. (1948). A mathematical theory of communication.

Simonton, D. K. (1999). Creativity as blind variation and selective retention: Is the creative process darwinian? *Psychological Inquiry*, 10(4):309–328.

Skinner, B. F. (1957). *Verbal Behavior*.

Sklyarov, V. and Skliarova, I. (2014). Hamming weight counters and comparators based on embedded dsp blocks for implementation in fpga. *Advances in Electrical and Computer Engineering*, 14(2):63–68.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. *Studies of child language development*, 1:75–208.

Smith, T. C. and Witten, I. H. (1995). A genetic algorithm for the induction of natural language grammars. In *Proc. of IJCAI-95 Workshop on New Approaches to Learning for Natural Language Processing*, pages 17–24.

Sokol, J. (1998). *Malá filosofie člověka: Slovník filosofickỳch pojmu*. Vyšehrad.

Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005). Unsupervised learning of natural languages. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11629–11634.

Sosík, P. and Štỳbnar, L. (1997). Grammatical inference of colonies. In *New Trends in Formal Languages*, pages 236–246. Springer.

Spears, W. M., De Jong, K. A., Bäck, T., Fogel, D. B., and De Garis, H. (1993). An overview of evolutionary computation. In *Machine Learning: ECML-93*, pages 442–459. Springer.

Spencer, H. (1894). *Education: Intellectual, moral, and physical*. CW Bardeen.

Strong, L. C. (1945). Anthony askham, the author of the voynich manuscript. *Science*, 101(2633):608–609.

Suciu, A., Cobarzan, P., and Marton, K. (2011). The never ending problem of counting bits efficiently. In *Roedunet International Conference (RoEduNet), 2011 10th*, pages 1–4. IEEE.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts: with special reference to north american indians and eskimos. *Proceedings of the American philosophical society*, pages 452–463.

Tomasello, M. (2009). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M., Akhtar, N., Dodson, K., and Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language*, 24(02):373–387.

Tomita, M. (1982). Dynamic construction of finite-state automata from examples using hill-climbing. In *Proceedings of the fourth annual cognitive science conference*, pages 105–108.

Tononi, G. (2004). An information integration theory of consciousness. *BMC neuroscience*, 5(1):42.

Trevarthen, C. (1993). The self born in intersubjectivity: The psychology of an infant communicating.

Trivers, R. (1972). Parental investment and sexual selection.

Turing, A. M. (1939). Systems of logic based on ordinals. *Proceedings of the London Mathematical Society*, 2(1):161–228.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, pages 433–460.

Ventris, M. and Chadwick, J. (1953). Evidence for greek dialect in the mycenaean archives. *The Journal of Hellenic Studies*, 73:84–103.

Vygotsky, L. S. (1978). Mind and society: The development of higher mental processes.

Vygotsky, L. S. (1987). Thinking and speech. *The collected works of LS Vygotsky*, 1:39–285.

Wall, L. et al. (1994). The perl programming language.

Watson, J. D., Crick, F. H., et al. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356):737–738.

Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1):49–63.

Wernicke, C. (1874). {Der aphasische Symptomencomplex}.

Widdows, D. and Cohen, T. (2014). Reasoning with vectors: a continuous model for fast robust inference. *Logic Journal of IGPL*, page jzu028.

Wiener, N. (1961). *Cybernetics or Control and Communication in the Animal and the Machine*, volume 25. MIT press.

Wilson, E. O. (2000). *Sociobiology: The new synthesis*. Harvard University Press.

Wittgenstein, L. (1922). *Tractatus logico-philosophicus*. Kegan Paul.

Wittgenstein, L. (1934). The blue book.

Wittgenstein, L. (1953). *Philosophical Investigations*. Blackwell.

Wolff, J. G. (1988). Learning syntax and meanings through optimization and distributional analysis. *Categories and processes in language acquisition*, 1(1).

Wright, S. (1932). *The roles of mutation, inbreeding, crossbreeding, and selection in evolution*, volume 1. na.

Zipf, G. K. (1949). Human behavior and the principle of least effort.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CL Computational Linguistics

DP Developmental Psycholinguistics

EC Evolutionary Computing

EL Evolutionary Linguistics

ES Evolutionary Strategy

ET Evolutionary Theory

FLT Formal Language Theory

GA Genetic Algorithm

GE Genetic Epistemology

GS Grammar System

GI Grammar Induction | Grammar Inference

HT Hard Thesis

LA Language Acquisition

LD Language Development

MDL Minimal Description Length

MLU Mean Length of Utterance

ND Neural Darwinism

NLP Natural Language Processing

OT Operational Thesis

POS-i Part-of-Speech Induction

POS-t Part-of-Speech Tagging

ST Soft Thesis

S^2T Softer Thesis

S^3T Softest Thesis

UD Universal Darwinism

VD Vocabulary Development

ACKNOWLEDGEMENTS

Ideas presented in this dissertation are result of crossover of multiple sources of intellectual and cultural influence. Some of them are enumerated in the bibliography and some of them are mentioned in acknowledgments section of my bachelor and master dissertations. Their individual names being listed elsewhere, the influence of the teachers - affiliated with Charles University in Prague, National University of Mongolia, University of Nice Sophia-Antipolis as well as of researches affiliated to 3rd Section of Ecole Pratique des Hautes Etudes (be it in Paris, Dijon, Caen, Aix en Provence or Bordeaux) - is not to be underestimated.

Also not to be underestimated is the influence of EPHE's 4th section. Although the unpredictability of life has't allowed me to pursue the "philologic track" longer than one year, I stay in the state of *reconnaissance profonde* towards prof. G-J. Pinault, J.E.M. Houben, D. Petit, A. LeMarechal, Fanny Meunier and others for showing me how "muddy" are formalisms of any theory in comparison with gems of poetry and language which have been, were, had been or simply *are* still alive.

University Paris 8 St. Dennis - renamed to Paris Lumières amidst the work on this dissertation - is also to be praised as an entity without whose support of which this dissertation would not see the light of a day. On one side its teachers, such as S. Peperkamp who introduced me into both beauties and intricacies of psycholinguistic disputes, or M.B. Jover whose courses of epistemology reignited in me a long-forgotten interest in philosophy. On another side dozens of other *fonctionnaires* who allowed me - sometimes slowly, sometimes fast, but always with success - to pursue my investigations of unknown epistemic territories.

However, EPHE and Paris 8 are just drops on the surface of structure much greater and ancient: Lutetia Parisorium, the city of Paris. Verily, if there is a three-dimensional mineral structure which is to be attributed the acknowledgment as a primary source of inspiration for this work, then it is the web which surrounds the Notre-Dame. Deeply grounded and firmly represented in my hippocampus hippocampus, Paris is indeed the city where first drafts of this dissertation had been brought to life and discussed in wines-lasting debates with Simon Carrignon, Adil elGhali, Fabien Ruggieri, Ilaria Gaudiello, Jean-Marc Thiebaut, Mary Rougeux, Yann Leger, Christophe Chavatte, Kechadi Lagha, Anne Ronsheim, Barbarka Jarkovska, Maurice & Florence Benayoun, Ivan Bigorgne, Geoffrey Tissier, Jeremy Gardent, my first student Pauline Vallies and second student Geoffrey Vantalon, Jarmila Mendelova, Louise Hearsum, Jitka Pelechova, Ophelie Monsoreau, Julienne Michele and Julie Rocton. Help coming from Kristina Poliakova, Julian Bonnyaud, Mikaela Barankova & Monique Girodroux, François Jodelet, Anh Nguyen, stream of good books coming from antiquaire Andre from bookshop on the border of rue d'Ulm and rue Claude Bernard as well as support of bouqinists Julien & Lue from Quai Voltaire were quite vital during periods when dissertation existed only *in posse* and not yet *in esse*.

But it would be unjust to praise Paris but not address some praise also *à la République toute-entière*. For without help of her institutions and her establishements - as diverse as CROUS, CAF, ANPE, Cité des Sciences, Société d'exploitation de la Tour Eiffel (!), Mairie de Paris or Campus France - it would be highly unplausible that an ordinary Bratislava boy could ever dedicate years of his life to pure science. In this regards, the role of French Ministry of Foreign Affairs, Embassy of France in Slovakia and people like Mme. Monika Saganova are of particular importance because of their assistance which ultimately allowed me to cover significant part of material needs with the scholarship of french government for doctoral studes under double supervision.

It is also thanks to Michal Oravec and Zuzana Dideková that such a double supervision got realized. By a strange coincidence of events and independently from each other they have both attracted my attention to the fact that in my own country of origin, Slovakia, there already exists a well-established, firm and intellectually rich tradition of cybernetics in general and evolutionary computation in particular. Hence I met Mr. Ivan Sekaj who was not only willing to take me under his wings, re-introduced me to education system of my own homeland, made me program my first genetic algorithm and always somehow succeeded to adopt his agenda to my needs. It is thanks to him that I had opportunity to get in contact with other "wizards from Mlynska Dolina", including prof. V. Kvasnicka or M. Popper.

None of these meetings and encounters would take place, however, if it hadn't been for one man: professor Charles Tijus. This is so because it was mainly Charles - assisted by Francois Jouen and Joelle Provasi - who kept alive the curriculum "Cognition Humain et Artificielle" at EPHE/Paris8, it was Charles who guided the direction of my Master Thesis and asides this also managed the complexities of laboratory ChART and the research platform Lutin where the germs of this dissertation have been conceived. But asides all this, it was Charles who convinced me that pursuing the path of science is worth the effort in order to subsequently give me practically absolute *liberté* in finding my own method of such pursue.

At last but not least, my ultimate "thank you" is dedicated to a woman which has transformed herself, during 6 years of doctoral studies, from a completely unknown *féerie* into a virtual acquitance into my guest into my host into my tourist guide into my friend into my love into my muse into mother of our daughter into my fiancee into my wife. It is thanks to You, Lucia, and thanks to thousands of numinous adjustments You do that our Iolanda Maitreya sleeps her green ideas peacefully and not furiously, that our house fragrances with myriads essences and this dissertation could be hereby considered as finished.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of November 17, 2016 (`classicthesis` version 1).

DECLARATION

I declare that this Thesis is a fruit of my own work and that all citations and references to external sources are explicitly marked.

Daniel Devatman Hromada,
November 17, 2016