Emergent grammars and the dynamics of language acquisition: an application of Dynamical Systems Theory to grammar construction

Núria Bosch
nb611@cam.ac.uk

St John's College, University of Cambridge

CRISSP Seminar (KU Leuven) - 14/12/2022

Roadmap

1. Introduction and theoretical background
    - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
    - General overview
    - Properties of Complex Adaptive Systems (CASs)
    - DST in cognition
3. A DST-based model of grammar construction
    - Emergence and the role of Universal Grammar
    - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
    - Control parameters: features and contrast
    - Dynamics *vs* representations
4. Novel predictions
    - Emergence in a dynamic and representationalist system
    - Shape and substance in learning paths
    - The flexible life of features
5. Conclusions

Roadmap

1. Introduction and theoretical background
    - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
    - General overview
    - Properties of Complex Adaptive Systems (CASs)
    - DST in cognition
3. A DST-based model of grammar construction
    - Emergence and the role of Universal Grammar
    - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
    - Control parameters: features and contrast
    - Dynamics *vs* representations
4. Novel predictions
    - Emergence in a dynamic and representationalist system
    - Shape and substance in learning paths
    - The flexible life of features
5. Conclusions

Introduction

- At least two fundamental aspects to any theory of language acquisition: (a) how do grammars emerge and develop? (b) what are the epistemological and ontological foundations of individual grammars?
    - ... But this has been perenially disputed.
    - Classical ”dichotomies”: nativism, emergentism, functionalism, usage-based, connectionism, generativism...
- I adopt a mathematical (and metatheoretical) perspective on these debates, presenting a Dynamical Systems theoretic view on language acquisition and generative grammar:
    - I suggest that DST may facilitate the understanding of facts about acquisition and extant generative approaches to learnability.

Neo-emergentism and biolinguistics

- **Biolinguistics** → language as a biological property of human beings, identical across humans, pathologies aside (Berwick and Chomsky, 2016; Hauser et al., 2002).
    - Approach language in the same way any other scientist would look at a biological system.
- Evolutionarily- and biologically-oriented perspective on language entails posing various questions:
    - What language is, which components it comprises, how it develops in children and adults and why it evolved in some way and not another.
    - The extent to which principles of linguistic systems are unique to this cognitive system or whether similar arrangements can be observed in different cognitive areas in humans or other species.
    - How much of language can be given a principled explanation based on well-known properties of natural and biological systems.

Neo-emergentism and biolinguistics

- **Biolinguistics** → language as a biological property of human beings, identical across humans, pathologies aside (Berwick and Chomsky, 2016; Hauser et al., 2002).
    - Approach language in the same way any other scientist would look at a biological system.
- Evolutionarily- and biologically-oriented perspective on language entails posing various questions:
    - What language is, which components it comprises, how it develops in children and adults and why it evolved in some way and not another.
    - The extent to which principles of linguistic systems are unique to this cognitive system or whether similar arrangements can be observed in different cognitive areas in humans or other species.
    - **How much of language can be given a principled explanation based on well-known properties of natural and biological systems**

I'll be focusing in particular on this last point

Neo-emergentism and biolinguistics

- In linking together biolinguistics, systems theory and acquisition together, I take two approaches as my main point of departure: Chomsky (2005)'s Three Factors approach and Biberauer (2019)'s Maximise Minimal Means (MMM).
- 3 factors contribute to the growth of language systems (adapted from Biberauer, 2019):

(1) (Poor) Universal Grammar + Primary Linguistic Data (PLD) + 3rd factors → Adult (steady-state) grammar

where 3rd factor = ”principles of data analysis” or ”principles of structural architecture and developmental constraints, [...] including principles of efficient computation”

Neo-emergentism and biolinguistics

- I'll also be assuming a *neo*-emergentist approach to language variation, in which UG is maximally impoverished (no innate parameters, no innate features, no innate cartographic spines, etc.).
    - E.g., UG may contain Merge, Agree and some notion of formal feature (Biberauer, 2019).
- Following on from Biberauer and Bosch (2021) → linguistic and cognitive systems show evidence for a third-factor principle - **Maximise Minimal Means** or MMM (Biberauer, 2019) - that attempts to make maximal use of minimal means. It unifies two independent linguistic manifestations:
    - **Feature Economy** (cf. Roberts and Roussou, 2003): minimise postulation of FFs and, in the MMM approach, only postulate them when contrastive (cf. Hall, 2007; Dresher, 2009).
    - **Feature Generalisation** (cf. Roberts, 2021, on Input Generalisation): maximise the use of a given FF ("recycle" it as much as possible).

Neo-emergentism and biolinguistics

- MMM approach views language as an *emergent complex system* (a Humboldt system; Abler, 1989) that ”makes infinite use of finite means” (von Humboldt, 1836, p. 70) and ”whose synthesis creates something that is not present per se in any of the associated constituents” (von Humboldt, 1836, p. 67).

Roadmap

1. Introduction and theoretical background
    - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
    - General overview
    - Properties of Complex Adaptive Systems (CASs)
    - DST in cognition
3. A DST-based model of grammar construction
    - Emergence and the role of Universal Grammar
    - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
    - Control parameters: features and contrast
    - Dynamics *vs* representations
4. Novel predictions
    - Emergence in a dynamic and representationalist system
    - Shape and substance in learning paths
    - The flexible life of features
5. Conclusions

General overview of Dynamical Systems Theory

- Dynamical Systems Theory (DST) is a sub-branch of mathematics and physics that describes the long-term development of complex dynamical systems, usually via the use of differential equations or difference equations.
- Geometric approach to systems → understand states and the evolution of a system in terms of its position with respect to other states and features of the systems' landscape.
- What counts as a dynamical system?
  - Informally, any system that changes with time.
  - Formally, a DS is a triple $< T, X, \Phi >$, where
    T = set of times
    X = state space
    Φ = transition rule that specifies the evolution of X with time ($\Phi : X \times T \longrightarrow X$).

General overview of Dynamical Systems Theory

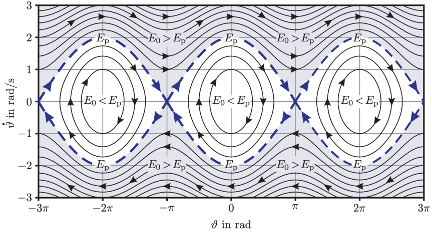


Figure 1: A phase portrait of the pendulum equation (Ochs, 2011)

General overview of Dynamical Systems Theory

- Some of the points of the system will be *attractive*, i.e., a set of states towards which a system tends to preferentially drift.
- New changes or reorganisations emerge due to (i) adaptation to the input and environment and (ii) due to changes in *control parameters*.
  - Control parameters as cues or parameters to which a system is sensitive → phase shifts.
- Many types of dynamical systems ((non)linear, (non)ergodic, chaotic, etc.). We'll focus mainly on one type: *complex adaptive systems* (CASs)
  - Type of *non-linear*, *open* and *complex* dynamical system.

Properties of complex adaptive systems

1. **Emergent, non-linear and self-organising**
    - New properties or forms come into existence due to interactions among more primitive components which do *not* have these properties themselves ("whole is more than sum of parts").
    - Therefore, they are *non-linear* and irreducible → process transformations cannot be reduced back to its original state and the change of the system output is not proportional to the change observed in the input.
    - Emergence is tightly linked to self-organisation: systems organise and structure themselves (order isn't pre-determined, rather "softly-assembled").

Properties of complex adaptive systems

2. **Sensitive to initial conditions**
   - Start at nearly identical state but can develop in opposite directions as the system amplifies initially small differences (cf. the butterfly effect).
   - Thus, system is recursive/iterative: it transforms x_t into x_{t+1}, x_{t+1} into x_{t+2}, x_{t+2} into x_{t+3}.
   - The output of some processes is therefore ”recycled” as new input for next step (”feedback sensitivity”).
   - Feedback occurs between levels of organisation, i.e., three dimensions of space-time:
     - ★ The **microscopic dimension** on the level of the individual **elements**.
     - ★ The **mesoscopic dimension** of the whole structure, limited by the **structural boundary**.
     - ★ The **macroscopic dimension** of the **field of interaction or relevant environment**, limited by the **system boundary**.

Properties of complex adaptive systems

2. **Sensitive to initial conditions**
   - DST advocates for a highly interactionalist view of systemic development → higher levels can "back-react" onto subunits, causing them to generate new patterns, which back-reacts again, etc.
   - Result of inter-dimension interactions is *structural homology*: later forms are built up of earlier forms (but crucially without losing their characteristic *emergent* property).

Properties of complex adaptive systems

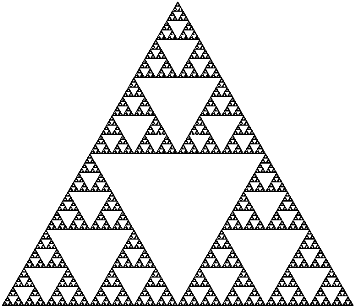
3. **Adaptive, critical and self-similar**
    - If the system has a macroscopic (i.e., environmental) dimension, then it'll be *adaptive* and *open* to the environment.
    - CASs are characterised by a high degree of adaptive capacity, but this adaptation does not proceed in just any conceivable way.
    - *Edge of chaos* → ability to avoid chaos by self-organising to a state roughly midway between globally static (unchanging, ordered) and chaotic (random, disordered) states, which ensures greater adaptability and resilience.
        - ★ Critical Brain Hypothesis or self-organised criticality: optimal information processing in non-linear systems is achieved close to phase transitions, a kind of "Goldilocks zone" of intermediate complexity that optimises the transfer and processing of information while still maintaining stability (for further information, see literature on criticality and "neuronal avalanches").

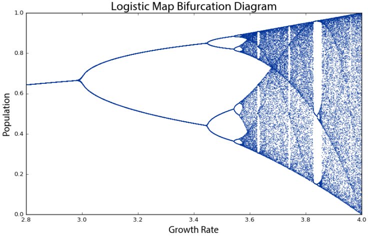
→ Allows systems to take advantage of sudden fluctuations in dynamics, but be stable enough so as to maintain order.

Properties of complex adaptive systems

3. **Adaptive, critical and self-similar**
   - What should be seen?
     - ★ Chialvo et al. (2008, p., 33) discuss the predictions of criticality for learning and cognitive development: ”increasing challenging levels that are easy enough to engage the learners but difficult enough not to bore them”.
     - ★ ”Raising the bar” effect → Life-long learning should be critical due to the effect of continuously ”raising the bar”.
   - One attractor that arises in edge of chaos dynamics, *strange attractors* → fractal (self-similar) structures appear naturally in CASs (complex patterns recur across different scales, by attracting the system to a recognisable structure). All strange attractors are geometrically fractals.

Properties of complex adaptive systems





DST in cognition

- New emphasis on the *how* of systemic change, instead of characterising *what* changes (Thelen and Smith, 1996).
- This gave rise to the Humean possibility that cognitive systems can be characterised in dynamic terms (Smolensky, 1988; Schöner, 2012).
- In this context, van Gelder (1998, p. 615) proposes his *Dynamical Hypothesis*:

The Dynamical Hypothesis

Cognitive agents are dynamical systems.

- My dissertation appends grammar construction to this list of dynamical processes and harnesses the tools and concepts supplied by DST to understand this cognitive phenomenon.

Roadmap

1. Introduction and theoretical background
   - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
   - General overview
   - Properties of Complex Adaptive Systems (CASs)
   - DST in cognition
3. A DST-based model of grammar construction
   - Emergence and the role of Universal Grammar
   - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
   - Control parameters: features and contrast
   - Dynamics *vs* representations
4. Novel predictions
   - Emergence in a dynamic and representationalist system
   - Shape and substance in learning paths
   - The flexible life of features
5. Conclusions

A DST-based model of grammar construction

- DST as a means of calling attention to the similarities among complex non-linear systems in nature and language acquisition.
- The starting point here is largely an adaptation of van Gelder's Dynamical Hypothesis formulated over acquisition:

The Dynamical Hypothesis of Acquisition

The cognitive process of grammar construction can be modelled as a complex adaptive system.

◎ Aim in this section: show how working under the DHA may provide a new characterisation of observations in acquisition and new tools with which to analyse learning paths.

Emergence, self-organisation and the role of Universal Grammar

- How do dynamical systems develop? → start in a relatively undifferentiated state. Sensitivity to initial conditions and a context against which to adapt cause them to diversify, self-organise and specialise.
- At initial period, the system's "learning potential" (its **degrees of freedom**) is much larger than in later developmental periods.
- Very **little pre-ordained instruction** needed: considerable complexity can emerge from a system with substantial degrees of freedom.
- Upon exposure to input/environment, the system complexifies and specialises to the task domain (read: target language), with reduction in degrees of freedom.
    - Applied to grammar construction, self-organisation and emergence seem to allow us to begin the acquisition path with little pre-specified structure.

Emergence, self-organisation and the role of Universal Grammar

- UG is superfluous in this model, then? No.
  - A *maximally poor UG* is at home with systems theory.
- UG here counts as the very first set of initial conditions of the system - containing certain formal universal principles that constrain the shape of attested/attestable languages.
- **Navigation** or **steering** role. It restricts the possible phase spaces of dynamical systems towards organisations driven by Merge and structured in featural terms.
  - Retains some of the ”steering” role in the P&P era, but in a much more underspecified way.
  - It specifies the *shape* of grammars, but no substantive content.
  - Softly-assembled development (no hard-wired elements; parameters and representations are emergent, so is the substantive content of features).
- UG + iterativity + emergence/self-organisation = structurally homologous, (neo-)emergent systems.

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- Dynamical systems are **biased** in which subsets of the phase space they are drawn to (recall **attractors** earlier) ↔ acquisition systems are also biased in which facets of the input they pay particular attention to and which choices get priority.
- Thus, dynamical attractors = third-factor learning biases (at the very least).
- Attractors are of various kinds: point attractors, limit-cycle attractors or strange attractors (cf. also Lass, 1997; Roberts, 2021, on the role of attractors in diachrony).

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- Dynamical systems are **biased** in which subsets of the phase phase they are drawn to (recall **attractors** earlier) ↔ acquisition systems are also biased in which facets of the input they pay particular attention to and which choices get priority.
- Thus, dynamical attractors = third-factor leaning biases (at the very least).
- Attractors are of various kinds: point attractors, limit-cycle attractors or strange attractors (cf. also Lass, 1997; Roberts, 2021, on the role of attractors in diachrony).

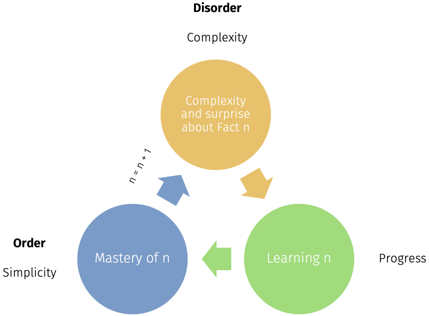
The focus here will be on the consequences of *strange attractors*

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- From DHA it follows that grammar construction systems go through edge of chaos stages and that they are biased by strange attractors.
    - *Intermediate* zone that is neither too ordered nor too disordered, but ”just right” for exhibiting complex development.
- Self-organised criticality is found in physical, biological and other systems. Two (possibly one) analogue(s) in linguistic systems:
    - The **Goldilocks Principle** (Kidd et al., 2012, 2014) → infants direct attention towards input sequences that are neither too simple (already learned) or nor too complex (unlearnable), but ”just right”.
    - **Maximise Minimal Means**
- Grammar construction system will operate in such a way that it tackles the input and areas of the hypothesis space that are intermediately complex, from which most learning advantage may be extracted and from which it avoids complete systemic disorder (see Biberauer, 2018; Biberauer and Bosch, 2021; Gerken et al., 2011; Poli et al., 2020, 2022; Brochhagen and Boleda, 2022, for potential examples).

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- In edge of chaos systems the *complexity-to-simplicity-to-complexity* pathway falls out for free (a ”progress niche” in the sense of Oudeyer et al., 2007; Oudeyer and Smith, 2016)



Disorder

Complexity

Complexity and surprise about Fact n

Learning n

Progress

Mastery of n

Order

Simplicity

n = n + 1

Figure 2: The Goldilocks Cycle

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- In reducing input to intake (Evers and van Kampen, 2008; Lidz and Gagliardi, 2015), levels of either very low or very high complexity are left out. Children preferentially attend to *partially integrable stimuli* (cf. Gagliardi, 2012, on ”partial” encoding as a catalyst driving the acquisition system forward).
- Overall: edge-of-chaos perspective on MMM and Abler (1989)’s Humboldt systems.

Attractors and life at the edge of chaos: Goldilocks, MMM and fractals

- Coming back to the link fractals ↔ edge of chaos ↔ strange attractors.
- Extending this Goldilocks-style approach to natural language fractals → link between MMM, Goldilocks, fractals and edge-of-chaos. Fractals are related to acquisitional dynamics that operate at an edge of chaos and Goldilocks zone.
  - E.g., parallelism between CP and DP (Abney, 1987).
- Why do fractals pervade?
  - Minimal Kolmogorov complexity.
  - Role of ”partial encoding”, again.
  - Therefore, not too unexpected if they arise at edge of chaos zones.
- Goldilocks and MMM likely aren’t separate biases, however (Biberauer, 2018; Biberauer and Bosch, 2021).
  - MMM leads us to expect “Goldilocks” effects: maximise and work on what is ”just right” at each point in acquisition.
  - Here, they would also have a unified source: Goldilocks, MMM and fractals as linguistic and developmental manifestations of a strange attractor in CASs.

Control parameters: features and contrast

- Recall control parameters earlier: cues or sets of cues that drive system reorganisations.
- Control parameters have been argued to be established near the edge of chaos zone (Baym and Hübler, 2006).
- My suggestion here → control parameters play a direct role in (MMM-driven) **feature postulation** on the part of the child.
  - i.e., examining which correlations control parameters are sensitive to may prove informative in deciding how children systematise the PLD they are confronted with.
- Main cue that has been linked to control parameters in the DST literature:
  - **Contrast and discrepancies**: discrepancies between what the system expects (i.e., what the system has learned) and what the context provides drive the system forward through successive reorganisations (Tucker and Hirsh-Pasek, 1993).

Control parameters: features and contrast

- Therefore, featural (or categorial) organisations piggyback on contrast here:
  - Dresher (2009, 2014)'s Successive Division Algorithm.
  - Cowper and Hall (2014)'s *Reductiō ad discrīmen* and Hall (2007)'s Contrastivity Hypothesis.
  - Biberauer (2019)'s "departures from Saussurean arbitrariness".
  - Cui (2020): phonological categories emerge by making lexically-contrastive divisions in the acoustic space.

Dynamics *vs* representations: reconciling continuity with discreteness

- Dynamical systems have been associated with anti-representationalism (Haselager et al., 2003; van Gelder, 1995)
  - Or if they do have some sort of representations (cf. connectionist architectures), they have been argued to be continuous, *not* symbolic (Smolensky, 1988).
- Contrast this with computational and symbolic (”information processing”) cognitive science (e.g, Church, Turing, Fodor, Pylyshyn…).
  - Cognition viewed as consisting of ”effectively computable” processes, that is, produced by means of a finite number of basic operations that act upon symbols, specified by some algorithm (e.g., Turing machines).

Dynamics *vs* representations: reconciling continuity with discreteness

- But there's little in the flow of a dynamical system that would preclude it from being representational, or from being linked to a representational system.
    - Dynamicism and representationalism can be seen as distinct but complementary (Dale et al., 2009), or even consistent with each other given the implicit computations performed by dynamical systems (Giunti, 2006; Siegelmann and Fishman, 1998; Crutchfield and Mitchell, 1995).
- I support *explanatory pluralism* here[1]:
    - I follow insights from frameworks of DST that argue both systems of measurement (continuous/dynamic and discrete/symbolic) are dissociable and have important qualities of equivalence.
        - ★ This has consequences for our resulting view of language acquisition through DST.

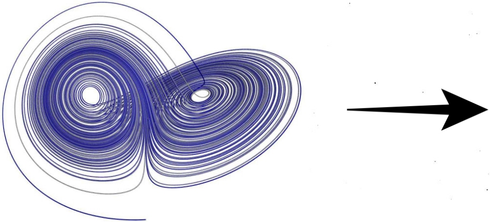
---

[1] "Practitioners of supposedly "hard" fields, such as physics, are very familiar with the fact that a total description of a given reality may require the use of several logically incompatible theories" (Mandelbrot, 1965, p. 553)

Dynamics *vs* representations: reconciling continuity with discreteness

- Main intuition: (acquisitional) dynamics and representations are *dissociable*, but tightly interconnected.
    1. Point of departure → dynamical systems can be viewed as geometric spaces.
    2. Assume that there exists a level of conceptual representation with contiguous domains that is also geometric (as famously suggested in Gärdenfors's, 2000; 2014 theory of Conceptual Spaces).
    3. Establish some equivalence and ontological relation between these two spaces.

Dynamics *vs* representations: a proposed architecture



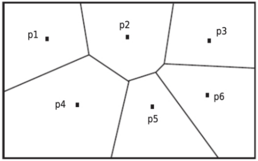


Figure 6: Voronoi tessellation of the plane into convex sets.

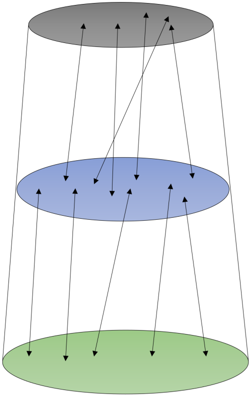
Dynamics *vs* representations: a proposed architecture

- Topological mapping between the dynamics of a system and the conceptual spaces derived from those dynamics.
- Fekete (2010) → activity spaces (= spaces of spatio-temporal events generated by a dynamical system) act as representational media.
  - Result: the representation of a system can be analysed in terms of the geometrical and topological properties of the systems' activity space.

Dynamics *vs* representations: a proposed architecture

- But Fekete's approach does not quite reach symbolic representations (computations performed over conceptual spaces ≠ computations over symbols).
- Extending the approach to symbolic computation → take Conceptual Spaces as a *mesolevel* between the macrolevel of symbolic computation and the microlevel of dynamics (Aisbett and Gibbon, 2001).

Dynamics *vs* representations: a proposed architecture



Macrolevel

Symbolic computation

Mesolevel

Conceptual spaces

Microlevel

Acquisitional (subsymbolic) dynamics

Figure 3: A topological mapping between acquisitional dynamics, conceptual spaces and representations

Dynamics *vs* representations: a proposed architecture

- This in the spirit of two main frameworks of analysis of dynamical systems:
    1. **Symbolic dynamics** (Dale and Spivey, 2005): analyse dynamical systems by *discretising* space. A space of representational symbols Σ and its shift map σ (the progression in time of symbols emitted by a system) have a certain geometrical equivalence to a dynamical system's continuous mapping and the set of states it visits.

Symbolic dynamics

- In symbolic dynamics, the systems (X, Φ) and (Σ, σ) are thus related to each other by:

  (2) $\pi \circ \Phi = \sigma \circ \pi$

  which can be represented as a commutative diagram:

$$\begin{tikzcd} x \arrow[r, "\Phi"] \arrow[d, "\pi"'] & \Phi(x) \arrow[d, "\pi"] \\ s \arrow[r, "\sigma"] & \sigma(s) \end{tikzcd}$$

  where $\pi : X \rightarrow \Sigma$ acts as an intertwiner and where s consists of the sequence of symbols emitted in the space Σ.

Dynamics *vs* representations: a proposed architecture

- This in the spirit of various frameworks of analysis of complex systems:
  1. **Symbolic dynamics** (Dale and Spivey, 2005): analyse dynamical systems by *discretising* space. A space of representational symbols Σ and its shift map σ (the progression in time of symbols emitted by a system) have a certain geometrical equivalence to a dynamical system's continuous mapping and the set of states it visits.
  2. **Evolutive Systems** (Ehresmann and Vanbremeersch, 2007, and also Bosch, in progress): hybrid model with DST and Category Theory, with a division of labour:
     - ★ DST: dynamical system plays role of 'regulator': regulates and helps build the representational system, specifying the changes to be made at each time.
     - ★ Category Theory: models the state of the representations at each time t and the relations between all its symbols. Also models the broader system as a whole, namely the connection between the dynamical and representational systems.

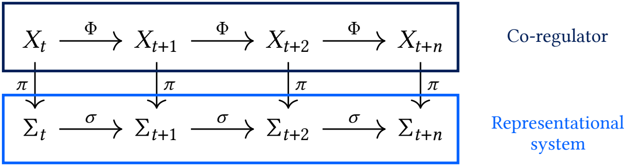
Symbolic dynamics and evolutive systems

- Abstracting even further, and concentrating on the systems as a whole and their evolution in time, they relate to each other as follows:

$$\begin{tikzcd}
X_t \arrow[r, "\Phi"] \arrow[d, "\pi"'] & X_{t+1} \arrow[r, "\Phi"] \arrow[d, "\pi"] & X_{t+2} \arrow[r, "\Phi"] \arrow[d, "\pi"] & X_{t+n} \arrow[d, "\pi"] \\
\Sigma_t \arrow[r, "\sigma"] & \Sigma_{t+1} \arrow[r, "\sigma"] & \Sigma_{t+2} \arrow[r, "\sigma"] & \Sigma_{t+n}
\end{tikzcd}$$

Symbolic dynamics and evolutive systems

- Abstracting even further, and concentrating on the systems as a whole and their evolution in time, they relate to each other as follows:



$$\begin{tikzcd}
X_t \arrow[r, "\Phi"] \arrow[d, "\pi"'] & X_{t+1} \arrow[r, "\Phi"] \arrow[d, "\pi"] & X_{t+2} \arrow[r, "\Phi"] \arrow[d, "\pi"] & X_{t+n} \arrow[d, "\pi"] \\
\Sigma_t \arrow[r, "\sigma"] & \Sigma_{t+1} \arrow[r, "\sigma"] & \Sigma_{t+2} \arrow[r, "\sigma"] & \Sigma_{t+n}
\end{tikzcd}$$

Co-regulator

Representational system

Dynamics *vs* representations: a proposed architecture

- Recapitulating: acquisitional dynamics ↔ conceptual spaces ↔ symbolic computation.
  - Some bidirectionality and top-down influence to be expected: e.g., acquirers do not seem to consider non-convex categories in artificial learning experiments (Heffner et al., 2019) and L1 acquisition is structure-dependent.
- Stepping back: how new is this in linguistics?
  - Representations reflecting the learning path is fairly widespread:
    - ★ Dresher (2009)'s Successive Division Algorithm
    - ★ Biberauer and Roberts (2015)'s NO>ALL>SOME acquisition path.
  - Maximising Minimal Means → recycle already-existent (parts of) a topological space.
  - New consequences, too, however (to be explored later).

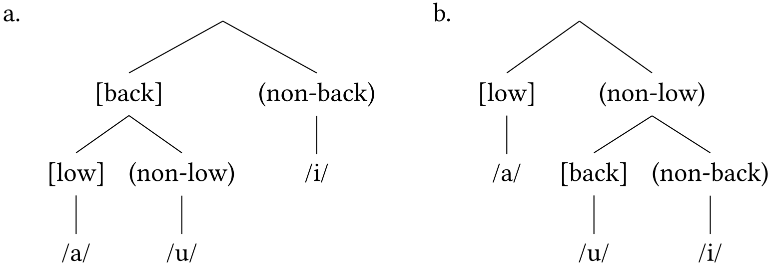
Roadmap

1. Introduction and theoretical background
    - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
    - General overview
    - Properties of Complex Adaptive Systems (CASs)
    - DST in cognition
3. A DST-based model of grammar construction
    - Emergence and the role of Universal Grammar
    - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
    - Control parameters: features and contrast
    - Dynamics *vs* representations
4. Novel predictions
    - Emergence in a dynamic and representationalist system
    - Shape and substance in learning paths
    - The flexible life of features
5. Conclusions

Emergence in a dynamic and representationalist system

- Symbolic systems hinge on learnability and conceptual spaces to a considerable degree here, meaning:
    - Most (all?) representations in steady-states reflect the learning path (cf. Dresher, 2009, 2021; Biberauer and Roberts, 2015).

(3)



a.

[back]

(non-back)

[low]

(non-low)

/i/

/a/

/u/

b.

[low]

(non-low)

/a/

[back]

(non-back)

/u/

/i/

Emergence in a dynamic and representationalist system

- Symbolic systems hinge on learnability and conceptual spaces to a considerable degree here, meaning:
  - Symbolic representations should obey the same structural principles as Conceptual Spaces (esp. its convexity requirement). This seems to be born out for:
    - Lexical acquisition (Dautriche and Chemla, 2016; Xu and Tenenbaum, 2007; Plunkett et al., 2008)
    - Perceptual categories (Biberauer and Bosch, 2021)
    - Linguistic categories (Gärdenfors, 2000, 2014; Heffner et al., 2019; Biberauer and Bosch, 2021)

Emergence in a dynamic and representationalist system

- Lexical acquisition (Dautriche and Chemla, 2016; Xu and Tenenbaum, 2007; Plunkett et al., 2008)
    - Both adults and children avoid discontinuity in word-learning experiments.
    - They avoid postulating a word for a discontinuous, disjunctive meaning (e.g., DOG OR BONE), and resort to homophony.

Emergence in a dynamic and representationalist system

- Perceptual categories (Biberauer and Bosch, 2021)
    - Object recognition appears to rely crucially on, firstly, contiguously-organised visual categories and, secondly, on edge detection (see Structural Information Theory and Recognition-by-Components Theory).
    - Visual cortex is specifically well-adapted to detecting edges of continuous visual domains (Elder and Sachs, 2004).

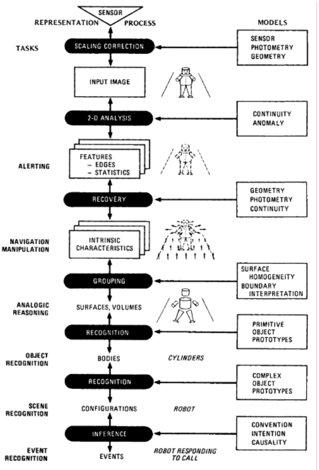
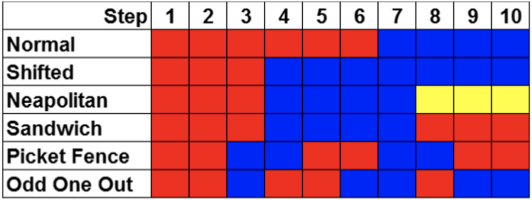


Figure 1. Computational architecture for a general-purpose vision system.

Emergence in a dynamic and representationalist system

- Linguistic categories (Gärdenfors, 2000, 2014; Heffner et al., 2019; Biberauer and Bosch, 2021)
  - Heffner et al. (2019): auditory categories varying along a unidimensional scale are more learnable for adults when they are contiguous, and get progressively harder as the discontinuity is amplified.



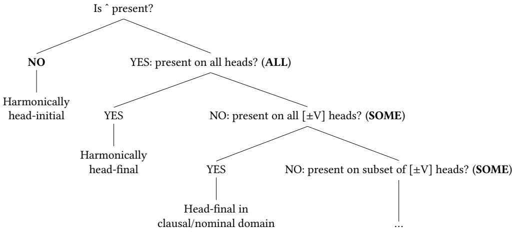
Step	1	2	3	4	5	6	7	8	9	10
Normal	🟥	🟥	🟥	🟥	🟥	🟥	🟦	🟦	🟦	🟦
Shifted	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟦	🟦	🟦
Neapolitan	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟨	🟨	🟨
Sandwich	🟥	🟥	🟥	🟦	🟦	🟦	🟦	🟥	🟥	🟥
Picket Fence	🟥	🟥	🟦	🟦	🟥	🟥	🟦	🟦	🟥	🟥
Odd One Out	🟥	🟥	🟦	🟥	🟥	🟦	🟦	🟥	🟦	🟦

Shape and substance in learning paths

- More fundamentally, a radically impoverished UG + hallmarks of CASs + topological mapping suggests that crosslinguistic universality in learning paths may relate to *shape* or the geometry of the flow, rather than *substance*.
- Common shape, topology and attractors leads to crosslinguistic uniformity, but the specific *substance* attached to this geometry may be subject to variation across L1s.
  - Learning paths are *dynamically* similar, but *substantively* dissimilar.

Shape and substance in learning paths

- What shape?
  - Broadly, general-to-specific or coarse-to-fine-grained (*pace* Westergaard, 2009).
- NO>ALL>SOME and emergent categorial hierarchy (Biberauer and Roberts, 2015)



Is ˆ present?

NO

Harmonically head-initial

YES: present on all heads? (**ALL**)

YES

Harmonically head-final

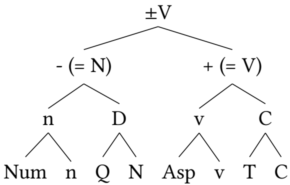
NO: present on all [±V] heads? (**SOME**)

YES

Head-final in clausal/nominal domain

NO: present on subset of [±V] heads? (**SOME**)

...



±V

- (= N)

+ (= V)

n D v C

Num n Q N Asp v T C

Shape and substance in learning paths

- If there is no substance in UG and development is largely ”softly assembled”, there is possibly very little (or no) room for linguistic maturation in this model (*pace* Radford, 1988; Friedman et al., 2020, and cf. Bosch, in progress)

The flexible life of features

? How do features emerge in such a model?

- The approaches so far:
  1. **Innate features**: one-time selection from an inventory of formal features supplied by UG (Chomsky, 2001). So, the problem of their emergence either does not arise or its answer falls out for free.
  2. **Emergent features**: little discussion on their precise development. Tacit assumption that once a relevant feature is postulated, it "stays as is".
     - ★ This would mean the child would proceed from having no feature to acquiring it with its full steady-state specification (with possibly no intermediate stage). This may be feasible for some features, but perhaps not all.
     - ★ Futher, if one adopts a truly emergentist approach to FFs, it requires a stipulation to render them ”fixed” once the child has postulated them.

The flexible life of features

- I would like to suggest that DST also has something to contribute to this discussion.
- Recall we linked **control parameters** to **feature postulation**.
- Two important properties of dynamical systems become relevant here:
  - Feedback and structural homology → the output of some processes within the system is "recycled" and becomes a new input for the system. Later steps always elaborate on earlier structures.
  - Cues and weights for control parameters change in order to direct the systems to more language-relevant analyses (Tucker and Hirsh-Pasek, 1993; Hirsh-Pasek et al., 1996).
- FFs are **malleable** in acquisition → dynamical systems can **recycle** the use made of a FF and a given FF's specifications can get progressively **refined** during development (from general to specific)[2].
  - Hale (1986) - features as "semantically broad, ontologically flexible and category independent".

---

[2] Although the latter is fairly novel for FFs, it has been suggested for categories, see Fourtassi and Dupoux (2014).

The flexible life of features

- We can call this the *Formal Feature Adaptability Hypothesis* (or FFAH), formulated as a bipartite statement:

Formal Feature Adaptability Hypothesis

- The formal features of complex adaptive (linguistic) systems may get “recycled” throughout the course of acquisition and may take on new, but developmentally related, functions (following Biberauer, 2019).
- The substantive content of a feature may start out being comparatively underspecified and may undergo later refinement.

The flexible life of features

- The logic that results from FFAH: ”refine and recycle features” (aka. make maximal use of a feature).
  - Developmental interaction prosody vs head-directionality (Gervain and Werker, 2008):
    - ★ OV languages → strong(=complement)-weak(=head) prosodic outline ($[\mathbf{s}\mathrm{w}]_\phi$)
    - ★ VO languages → weak(=head)-strong(=complement) prosody ($[\mathrm{w}\mathbf{s}]_\phi$).
    - ★ Prosodic cues as the coarse-grained formal basis for later syntactic features.
  - Recycling features beyond the syntactic domains they were first postulated for:
    - ★ Harbour (2020) on ”Frankenduals”.
    - ★ Adger and Harbour (2007)’s treatment of the Person Case Constraint.
    - ★ Ritter and Wiltschko (2014) on INFL.
    - ★ Douglas (2018) on Maōri subject extraction.
  - ”Refining” predicts interim developmental stages (possibly unattested in the adult L1), where categories/features are ”mid-way” from being non-existent (unpostulated) to fully adult-like:
    - ★ Hanson (2000): in the acquisition of pronominal morphology, the defaults of each Organising node (e.g., Participant, Class) in Harley and Ritter (2002)’s feature geometry are acquired before the node is subdivided into additional features (Speaker, Addresse, Animate, etc.).

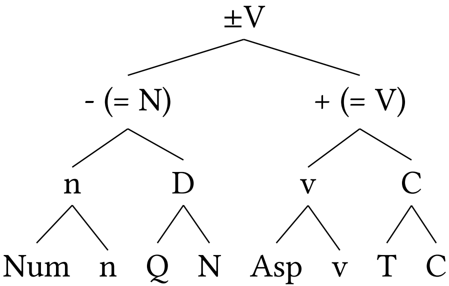
The flexible life of features

- Also provides a rationale for why UG-given features will be unhelpful to the child: UG-given information is steady-state (adult) information. But if FFAH is correct, there is much more to the life of a feature than what is readily perceptible in steady-state specifications.
- Flexible or ”polyvalent” features are expected in structurally homologous systems. If edge-of-chaos dynamics are a driving force, fundamentally adaptable features should abound, either during development or at steady-states.
  - ”Intermediate” or ”mid-way” developmental stages

Flexibility and homology beyond features

- FFAH is a specific hypothesis emerging from the broader pattern of **structural homology** observed in these systems. The 'refinement' and 'recycling' duo is thus not restricted to features.
    - Biberauer and Roberts (2015)
    - Mitrofanova (2018)'s Underspecified P Hypothesis → Russian children acquiring PPs go through a stage where a coarse-grained Prep category has been postulated, but not individual cartographic heads encoding fine-grained meaning distinctions (e.g., AxPartP).
    - My MPhil thesis (ongoing): detecting the stages in the emergence of syntactic categories in child production data

Emergent categories



±V

- (= N)

+ (= V)

n

D

v

C

Num n Q N Asp v T C

Roadmap

1. Introduction and theoretical background
   - Neo-emergentism and biolinguistics
2. Dynamical Systems Theory
   - General overview
   - Properties of Complex Adaptive Systems (CASs)
   - DST in cognition
3. A DST-based model of grammar construction
   - Emergence and the role of Universal Grammar
   - Attractors and life at the edge of chaos: Goldilocks, MMM and fractals
   - Control parameters: features and contrast
   - Dynamics *vs* representations
4. Novel predictions
   - Emergence in a dynamic and representationalist system
   - Shape and substance in learning paths
   - The flexible life of features
5. Conclusions

Conclusions

- We tried to pursue three different research questions here:
  1. How do grammars emerge and reach the levels of complexity they do?
  2. Can DST help us understand (1)? And what consequences would it have for learnability theory?
  3. Can we set out the foundations of what a DST-based and neo-emergentist model of grammar construction would look like?

Conclusions

- CASs and neo-emergentist systems see eye to eye:
  - Minimal starting conditions needed ↔ Poor UG
  - Structural homology, sensitivity to initial conditions
  - Edge of chaos and strange attractors ↔ MMM, Goldilocks, fractals
  - Contrast and discrepancies ↔ features/categories piggyback on contrast.
  - Representations are fully emergent, predicting resulting idiosyncrasies (e.g., ”refining”, ”recycling” and ”interim” developmental stages).

→ CASs thus sketch a broad developmental picture which aligns particularly well with current neo-emergentist proposals.

→ Existent theoretical proposals and preliminary empirical evidence (from development, but also synchrony and diachrony) align with these predictions.

- Highly programmatic proposal, but it provides tools with which to understand UG, the Three Factors and linguistic/cognitive representations.
  - 💡 Effectively, it probes how much mathematical systems ↔ learnability ↔ language say about each other.

Thank you!

Acknowledgements

Huge thanks go to Theresa Biberauer, for supervising this work and for all the support throughout my time at Cambridge. Thanks too to St John's College and Cambridge Trust for generously supporting part of this research.

References I

Abler, W. L. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, 12(1):1–13.

Abney, S. P. (1987). *The English noun phrase in its sentential aspect.* PhD thesis, Massachusetts Institute of Technology.

Adger, D. and Harbour, D. (2007). Syntax and syncretisms of the Person Case Constraint. *Syntax*, 10(1):2–37.

Aisbett, J. and Gibbon, G. (2001). A general formulation of conceptual spaces as a meso level representation. *Artificial Intelligence*, 133(1-2):189–232.

Baym, M. and Hübler, A. W. (2006). Conserved quantities and adaptation to the edge of chaos. *Physical Review E*, 73(5):056210.

Berwick, R. C. and Chomsky, N. (2016). *Why only us: Language and evolution.* MIT press.

Biberauer, T. (2018). Less is More: On the Tolerance Principle as a manifestation of Maximize Minimal Means. *Linguistic Approaches to Bilingualism*, 8(6):707–711.

Biberauer, T. (2019). Factors 2 and 3: Towards a principled approach. *Catalan Journal of Linguistics*, pages 45–88.

References II

Biberauer, T. and Bosch, N. (2021). The case for Maximise Minimal Means: linguistic and general-cognitive perspectives. Talk presented at *SyntaxLab* (University of Cambridge), 16 November.

Biberauer, T. and Roberts, I. (2015). Rethinking formal hierarchies: A proposed unification. *Cambridge Occasional Papers in Linguistics*, 7(1):1–31.

Bosch, N. (in progress). *Emergent Syntax and Maturation: rethinking maturation from a neo-emergentist lens.* MPhil thesis, University of Cambridge.

Brochhagen, T. and Boleda, G. (2022). When do languages use the same word for different meanings? The Goldilocks principle in colexification. *Cognition*, 226:105179.

Chialvo, D. R., Balenzuela, P., and Fraiman, D. (2008). The brain: What is critical about it? *AIP Conference Proceedings*, 1028(1):28–45.

Chomsky, N. (2001). Derivation by phase. In Kenstowicz, M., editor, *Ken Hale: a life in language*, pages 1–50. MIT Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1):1–22.

Cowper, E. and Hall, D. C. (2014). Reductio ad discrimen: Where features come from. *Nordlyd*, 41(2):145–164.

References III

Crutchfield, J. P. and Mitchell, M. (1995). The evolution of emergent computation. *Proceedings of the National Academy of Sciences*, 92(23):10742–10746.

Cui, A. (2020). *The Emergence of Phonological Categories*. PhD thesis, University of Pennsylvania.

Dale, R., Dietrich, E., and Chemero, A. (2009). Explanatory pluralism in cognitive science. *Cognitive science*, 33(5):739–742.

Dale, R. and Spivey, M. J. (2005). From apples and oranges to symbolic dynamics: A framework for conciliating notions of cognitive representation. *Journal of Experimental & Theoretical Artificial Intelligence*, 17(4):317–342.

Dautriche, I. and Chemla, E. (2016). What homophones say about words. *PloS one*, 11(9):e0162176.

Douglas, J. (2018). Māori subject extraction. *Glossa: a journal of general linguistics*, 3(1):110.

Dresher, B. E. (2009). *The contrastive hierarchy in phonology*. Cambridge University Press.

Dresher, B. E. (2014). The arch not the stones: universal feature theory without universal features. *Nordlyd*, 41(2):165–181.

References IV

Dresher, B. E. (2021). Contrastive hierarchies and phonological primes. *Perspectives on Element Theory*, 143:33.

Ehresmann, A. C. and Vanbremeersch, J.-P. (2007). *Memory Evolutive Systems: Hierarchy, Emergence, Cognition (Studies in Multidisciplinarity)*. Elsevier.

Elder, J. H. and Sachs, A. J. (2004). Psychophysical receptive fields of edge detection mechanisms. *Vision Research*, 44(8):795–813.

Evers, A. and van Kampen, J. (2008). Parameter setting and input reduction. In Biberauer, T., editor, *The Limits of Syntactic Variation*, pages 483–515. John Benjamins.

Fekete, T. (2010). Representational systems. *Minds and Machines*, 20(1):69–101.

Fourtassi, A. and Dupoux, E. (2014). A rudimentary lexicon and semantics help bootstrap phoneme acquisition. In *Proceedings of the Eighteenth Conference on Computational Natural Language Learning*, pages 191–200.

Friedman, N., Belletti, A., and Rizzi, L. (2020). Growing Trees. The acquisition of the Left Periphery. *lingbuzz/005369*.

Gagliardi, A. C. (2012). *Input and intake in language acquisition*. PhD thesis, University of Maryland, College Park.

References V

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* MIT Press.

Gärdenfors, P. (2014). *The geometry of meaning: Semantics based on conceptual spaces.* MIT Press.

Gerken, L., Balcomb, F. K., and Minton, J. L. (2011). Infants avoid ‘labouring in vain’ by attending more to learnable than unlearnable linguistic patterns. *Developmental science*, 14(5):972–979.

Gervain, J. and Werker, J. F. (2008). How infant speech perception contributes to language acquisition. *Language and Linguistics Compass*, 2(6):1149–1170.

Giunti, M. (2006). Is being computational an intrinsic property of a dynamical system? In Minati, G., Pessa, E., and Abram, M., editors, *Systemics of Emergence: Research and Development*, pages 683–694. Springer.

Hale, K. (1986). Notes on World View and Semantic Categories: some Warlpiri Examples. In Muysken, P. and van Riemsdijk, H., editors, *Features and projections*, pages 233–256. De Gruyter Mouton.

Hall, D. C. (2007). The role and representation of contrast in phonological theory. *Toronto Working Papers in Linguistics.*

References VI

Hanson, R. (2000). Pronoun acquisition and the morphological feature geometry. *Calgary Working Papers in Linguistics*, 22(Winter):1–14.

Harbour, D. (2020). Frankenduals: Their typology, structure, and significance. *Language*, 96(1):60–93.

Harley, H. and Ritter, E. (2002). Person and number in pronouns: A feature-geometric analysis. *Language*, 78(3):482–526.

Haselager, W. F., Bongers, R. M., and Van Rooij, I. (2003). Cognitive science, representations and dynamical systems theory. In Tschacher, W. and Dauwalder, J.-P., editors, *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination Dynamics*, pages 229–241. World Scientific.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298(5598):1569–1579.

Heffner, C. C., Idsardi, W. J., and Newman, R. S. (2019). Constraints on learning disjunctive, unidimensional auditory and phonetic categories. *Attention, Perception, & Psychophysics*, 81(4):958–980.

References VII

Hirsh-Pasek, K., Tucker, M., and Golinkoff, R. M. (1996). Dynamic Systems Theory: Reinterpreting" prosodic bootstrapping" and its role in language acquisition. In Morgan, J. and Demuth, K., editors, *Signal to syntax*, pages 461–478. Psychology Press.

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2012). The Goldilocks Effect: Infants' preference for stimuli that are neither too predictable nor too surprising. *PloS ONE*, 7(5).

Kidd, C., Piantadosi, S. T., and Aslin, R. N. (2014). The Goldilocks Effect in infant auditory attention. *Child development*, 85(5):1795–1804.

Lass, R. (1997). *Historical linguistics and language change*, volume 81. Cambridge University Press.

Lidz, J. and Gagliardi, A. (2015). How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annu. Rev. Linguist.*, 1(1):333–353.

Mandelbrot, B. (1965). Information theory and psycholinguistics. *BB Wolman and E.*

Mitrofanova, N. (2018). Early underspecification of functional categories: Evidence from the acquisition of locative pps in russian. *Language Acquisition*, 25(4):341–365.

References VIII

Ochs, K. (2011). A comprehensive analytical solution of the nonlinear pendulum. *European Journal of Physics*, 32(2):479.

Oudeyer, P.-Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation*, 11(2):265–286.

Oudeyer, P.-Y. and Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, 8(2):492–502.

Plunkett, K., Hu, J.-F., and Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, 106(2):665–681.

Poli, F., Meyer, M., Mars, R. B., and Hunnius, S. (2022). Contributions of expected learning progress and perceptual novelty to curiosity-driven exploration. *Cognition*, 225:105119.

Poli, F., Serino, G., Mars, R. B., and Hunnius, S. (2020). Infants tailor their attention to maximize learning. *Science Advances*, 6(39):eabb5053.

Radford, A. (1988). Small children's small clauses. *Transactions of the philological society*, 86(1):1–43.

References IX

Ritter, E. and Wiltschko, M. (2014). The composition of infl. *Natural Language & Linguistic Theory*, 32(4):1331–1386.

Roberts, I. (2021). *Diachronic Syntax*. Oxford University Press, 2nd edition.

Roberts, I. and Roussou, A. (2003). *Syntactic change: A minimalist approach to grammaticalization*. Cambridge University Press.

Schöner, G. (2012). Dynamical systems approaches to cognition. In Sun, R., editor, *The Cambridge Handbook of Computational Psychology*, pages 101–126. Cambridge University Press.

Siegelmann, H. T. and Fishman, S. (1998). Analog computation with dynamical systems. *Physica D: Nonlinear Phenomena*, 120(1-2):214–235.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and brain sciences*, 11(1):1–23.

Thelen, E. and Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Tucker, M. and Hirsh-Pasek, K. (1993). Systems and language: Implications for acquisition. *A dynamic systems approach to development: Applications*, pages 359–384.

References X

van Gelder, T. (1995). What might cognition be, if not computation? *The Journal of Philosophy*, 92(7):345–381.

van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioral and brain sciences*, 21(5):615–628.

von Humboldt, F. W. H. A. (1836). *Über die Verschiedenheit des menschlichen Sprachbaues und ihrer Einfluss auf die Geistige Entwickelung des Menschengeschlechts*. Dümmler.

Westergaard, M. (2009). *The Acquisition of Word Order*. John Benjamins, Amsterdam.

Xu, F. and Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2):245.