

# Explicit Causal Connections between the Acquisition of Linguistic Tiers: Evidence from Dynamical Systems Modeling

Anonymous ACL submission

## Abstract

In this study, we model the *causal* links between the complexities of different macroscopic aspects of child language. We consider pairs of sequences of measurements of the quantity and diversity of the lexical and grammatical properties. Each pair of sequences is taken as the trajectory of a high-dimensional dynamical system, some of whose dimensions are unknown. We use Multispatial Convergent Cross Mapping to ascertain the directions of causality between the pairs of sequences. Our results provide support for the hypothesis that children learn grammar through distributional learning, where the generalization of smaller structures enables generalizations at higher levels, consistent with the proposals of construction-based approaches to language.

## 1 Introduction

A crucial question in language acquisition concerns how (or, according to some, whether) children learn the grammars of their native languages. Some researchers, mainly coming from the generative tradition, argue that, although the grammatical rules are possibly ‘innate’ (Pinker, 1994), children still need to learn how to map the different semantic/grammatical roles onto the different options offered by Universal Grammar (e.g., ‘parameter-setting’). The evidence, however, does not seem to support this hypothesis. For instance, Bowerman (1990) notes that the type of semantic aspects learned by the child do not match well into the prototypical roles that would be required to map into hard linguistic rules (e.g., learning an AGENT category to map onto the SUBJECT syntactic role). Other researchers (Goldberg, 2003;

Tomasello, 1992; Tomasello, 2005) propose that there is a gradual increase in the generality of the structures learned by the child, which are slowly acquired through distributional learning. Such a picture is strongly supported by the remarkably little creativity exhibited by children, most of whose utterances are often literal repetitions of those that they have previously heard (Lieven et al., 1997; Pine and Lieven, 1993), with little or no generalization in the early stages. It appears as though children progressively and conservatively increase the level at which they generalize linguistic constructions, building from the words upwards, in what some have termed ‘lexically-based positional analysis’ (Lieven et al., 1997).

The Theory of Dynamical Systems offers powerful tools for modeling human development (e.g. Smith and Thelen, 2003; van Geert, 1991).. It provides a mathematical framework for implementing the principle that development involves the mutual and continuous interaction of multiple levels of the developing system, which simultaneously unfold over many time-scales. Typically, a dynamical system is described by a system of coupled differential equations governing the temporal evolution of multiple parts of the system and their interrelations. One difficulty that arises when trying to model a dynamical system as complex as the development of language is that many factors that are important for the evolution of the system might not be available or might not be easily measurable or –even worse– there are additional variables relevant for the system of which the modeler is not even aware. In this respect, a crucial development was the discovery that, in a deterministic coupled dynamical system –even in the presence of noise– the dynamics of the whole system can be satisfactorily recovered using measurements of a single of the system’s variables (Takens’ Embedding Theorem; Takens, 1981).

The finding above opens an interesting avenue for understanding the processes involved in language acquisition (perhaps more suitably termed language *growth*, following van Geert, 1991). In the same way that systems of differential equations can be used to model the evolution of ecosystems (e.g., predator-prey systems), one could take measurements of the detailed properties of CL and CDL, and build a detailed system of equations capturing the macroscopic dynamics of the process. However, in order to achieve this, it is necessary to ascertain the ways in which different measured variables in the system affect each other. This problem goes beyond estimating correlations (as could be obtained, for instance, using regression models), as one needs to detect asymmetrical *causal* relations between the variables of interest, so that these causal influences can be incorporated into the models.

In this study, we investigate the causal relations different macroscopic-level measures characterizing the level of development of different tiers child language (i.e., number of words produced, lexical diversity, inflectional diversity and mean length of utterances), using the longitudinal data provided in the Manchester Corpus (Theakston et al., 2001). In order to detect causal relations between the different measures, we make use of state space reconstruction relying on Takens (1981)’s Embedding Theorem, and recently developed techniques for assessing the strength of causal relations in dynamical systems (Multispatial Convergent Cross Mapping; Clark et al., 2015). Here, we provide a detailed picture of the causal connections between the development of different aspects of a child’s language (while acquiring English). Our result provide support for theories that advocate distributional learning of linguistic constructions by gradual generalizations from the level of words to larger scale constructions.

## 2 Causality Detection in Dynamical Systems

Whenever two variables are correlated, there *must* exist some causal link between them. Namely, if variables A and B are found to be correlated, then one of four possibilities must be true: (a) A causes B, (b) B causes A, (c) A and B form a feedback loop, each causing the other, or (d) there is a third variable C causing both A and B. For studying the interactions of species within ecosystems, Sugi-

hara et al. (2012) introduced *Convergent Cross Mapping* (CCM), a causality-detection technique that is valid for non-separable systems, is capable of identifying weakly coupled variables even in the presence of noise, and –crucially– can distinguish direct causal relations between variables from effects of shared driving variables (i.e., in possibility (d) above, CCM would *not* find causality).

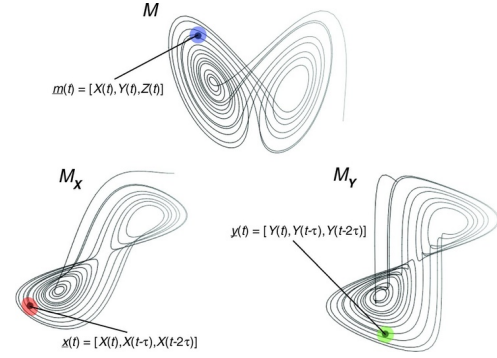


Figure 1: Reconstructed manifold for Lorenz’s system ( $M$ ; top), as well as the shadow manifolds reconstructed considering only  $X$  ( $M_X$ ; bottom-left) and  $Y$  ( $M_Y$ ; bottom-right) (reprinted with permission from Sugihara et al., 2012).

For instance, consider E. Lorenz’s often studied dynamical system including three coupled variables  $X(t)$ ,  $Y(t)$ , and  $Z(t)$  whose co-evolution is described by the system of differential equations

$$\begin{cases} \frac{dX}{dt} = \sigma(Y - X) \\ \frac{dY}{dt} = X(\rho - Z) - Y \\ \frac{dZ}{dt} = XY - \beta Z \end{cases} \quad (1)$$

The first equation in this system indicates that there is a relation by which  $Y$  causes  $X$ , as the change in  $X$  (i.e., its future value) depends on the value of  $Y$  (i.e., the future of  $X$  depends on the past of  $Y$  even after the past of  $X$  itself has been considered), a causal relation whose strength is indexed by parameter  $\sigma$ . The manifold defined by these three variables (Lorenz’s famous strange attractor), which we can denote by  $M$ , is plotted in the top of Fig. 1. In many circumstances, however, not all variables of the system are available (some might be difficult to measure, or we might not even be aware of their relevance). It is at this point that Takens (1981)’s Embedding

Theorem comes into play. Informally speaking, the theorem states that the properties of a coupled dynamical system’s attractor can be recovered using only measurements from a single one of its variables. This is achieved by considering multiple versions of the same variable lagged in time, that is, instead of plotting  $(X[t], Y[t], Z[t])$ , when only measurements of  $X$  are available, we can plot  $(X[t], X[t + \tau], \dots, X[t + (E - 1)\tau])$ . These reconstructed manifolds are termed “shadow” manifolds.  $M_X$  denotes the shadow manifold of  $M$  reconstructed on the basis of  $X$  alone. There are well-studied techniques for finding the appropriate values for the parameters for the lag  $\tau$  and the number of dimensions  $E$  (c.f., Abarbanel et al., 1993) so that the properties of the original manifold  $M$  are recovered by the shadow manifold  $M_X$ . Fig. 1 illustrates this point by plotting the shadow manifolds  $M_X$  (bottom-left) and  $M_Y$  (bottom-right) for the Lorenz system. Notice how both shadow manifolds recover much of the original’s structure, using only knowledge of one of its three variables.

Each point in the original manifold  $M$  maps onto points in its shadow manifolds, as is illustrated by the points labelled  $m(t)$ ,  $x(t)$ , and  $y(t)$  in Fig. 1. The preservation of the topological properties of the original manifold in its shadow manifolds entails that points that are close-by in the original manifold will also be close-by in its shadow versions. This implies that, for causally linked variables within the same dynamical system, the state of one variable can identify the states of the others. Sugihara et al. (2012) noticed that, when one variable  $X$  stochastically drives another variable  $Y$ , information about the states of  $X$  can be recovered from  $Y$ , but not vice-versa. This is the basic insight of the CCM method. To test for causality from  $X$  to  $Y$ , CCM looks for the signature of  $X$  in  $Y$ ’s time series by seeing whether the time indices of nearby points on  $M_Y$  can be used to identify nearby points on  $M_X$ . Crucially, in order to distinguish causation from mere correlation, CCM requires *convergence*, that is, that cross-mapped estimates improve in estimation accuracy with the sample size (i.e., “library size”) used for reconstructing the manifolds. As the library size increases, the trajectories defining the manifolds fill in, resulting in closer nearest neighbors and declining estimation error, which is reflected in a higher correlation coefficient between the points in

the neighborhoods of the shadow manifolds. Convergence then becomes the necessary condition for inferring causation. Using both artificial systems and ecological time-series with known dynamics, Sugihara and his colleagues demonstrated that this technique successfully recovers true directional causal relations when these are present, and –crucially– is able to discard spurious causation in the case when both variables are causally driven by a third, unknown, variable, but there is no true direct causation between them.

An inconvenience of CCM, and in general of techniques that rely on manifold reconstruction, is that they generally require that relatively long time-series of the behavior of the system are available. Such long series are, however, very difficult, if not impossible, to obtain in many fields, including of course language acquisition. One can however obtain multiple short time series from different instances of a similar dynamical system. In ecology, for instance, one can obtain short sequences of measurements of the population densities of a group of species measured at different places and times. In language acquisition, we might have multiple, relatively short longitudinal sequences of measurements from different children. With this in mind Clark et al. (2015) developed Multispatial CCM (mCCM), an extension of CCM able to infer causal relations from multiple short time-series measured at different sites, making use of dewdrop regression (Hsieh et al., 2008) to take the additional heterogeneity into account.

### 3 Materials and Methods

#### 3.1 Materials

We obtained from the CHILDES database (MacWhinney, 2000) the transcriptions contained in the Manchester Corpus (Theakston et al., 2001). This corpus contains annotated transcripts of audio recordings from a longitudinal study of 12 British English-speaking children (6 girls and 6 boys) between the ages of approximately two and three years. The children were recorded at their homes for an hour while they engaged in normal play activities with their mothers. Each child was recorded on two separate occasions in every three-week period for one year. Each recording session is divided into two half-hour periods. The annotations include the lemmatized form of the words produced by the children (incomplete words and small word-internal errors were manually cor-

rected in the lemmatization).

In order to increase the sample size in each period, we followed a sliding window technique of (Irvin et al., in press): We computed measures for the samples contained in overlapping windows of three consecutive corpus files. In this way, at each point we obtained samples originating from two files from the same recording session, and a file from either the previous or the next recording session.

### 3.2 Measures of Linguistic Development

As in previous studies (Irvin et al., in press; Moscoso del Prado Martín, in press), in order to measure the overall amount of speech produced by each child, we counted the total number of word tokens produced by each child in each temporal window. We refer to this measure as the child's *loquacity*.

In order to measure the diversity of the words used by the children, we use the *lexical diversity* measure (Irvin et al., in press; Moscoso del Prado Martín, in press). This is just the information entropy (Shannon, 1948) of the probability distribution of word lemmas found in the sample,

$$H[L] = \sum_{\ell \in L} p(\ell) \cdot \log \frac{1}{p(\ell)}, \quad (2)$$

where  $L$  refers to the set of word lemmas found in a sample, and  $p(\ell)$  is the probability with which the particular lemma  $\ell$  is found in that sample. Entropy estimates obtained using Eq. 2 using maximum likelihood estimates of the probabilities are known to be strongly biased (Miller, 1955), with the bias magnitude correlating with the size of the sample used. Importantly, the sample size is nothing else than the *loquacity* measure described above. Therefore, using this plain maximum likelihood method would result in spurious correlations. For this reason, Moscoso del Prado Martín (in press) recommends using the bias-adjusted entropy estimator (Chao et al., 2013, see Appendix A) instead of Eq. 2.

In order to measure the acquisition of inflectional morphological paradigms, we make use of the *inflectional diversity* measure (Moscoso del Prado Martín, in press). This is a macroscopic generalization of inflectional entropy (Moscoso del Prado Martín et al., 2004), a measure that is known to index morphological influences on adult lexical processing (Baayen and

Moscoso del Prado Martín, 2005; Moscoso del Prado Martín et al., 2004) as well as in child language acquisition (Stoll et al., 2012). The inflectional entropy of a lemma  $\ell$  ( $H[W|\ell]$ ) is the information entropy of the inflected variants of that lemma. Our inflectional diversity is just the average value of inflectional entropy across all lemmas,

$$H[W|L] = H[W, L] - H[L], \quad (3)$$

where  $H[L]$  is the lexical diversity measure described above, and  $H[W, L]$  is the joint entropy between the inflected word forms and their corresponding lemmas,

$$H[W, L] = \sum_{\ell \in L} \sum_{w \in W} p(w, \ell) \cdot \log \frac{1}{p(w, \ell)}, \quad (4)$$

where  $L$  denotes the set of all distinct lemmas encountered in the sample,  $W$  is the set of all distinct inflected word forms encountered, and  $p(w, \ell)$  is the joint probability with which lemma  $\ell$  occurs as the specific inflected form  $w$ . Inflectional diversity takes non-negative values, measuring how large are the average inflectional paradigms used in the language sample. Estimating  $H[W, L]$  using Eq. 4 is subject to the same estimation biases that were described for lexical diversity. Therefore, we also follow Moscoso del Prado Martín (in press) in using the bias-adjusted estimate (Chao et al., 2013, see Appendix A) for this magnitude, and then combining it with the lexical diversity using Eq. 3 to obtain our inflectional diversity estimates.

Finally, in order to measure the degree of syntactic development of the children we used their *mean length of utterances* (MLU). Instead of measuring MLU in morphemes (Brown, 1973), we used the simpler, but equally accurate measure in number of words (c.f., Parker and Bronson, 2005). In these ages, MLUs are well known to provide an accurate measure of the syntactic richness of the utterances produced (Brown, 1973), and in fact correlate almost perfectly with explicit measurements of grammatical diversity (Moscoso del Prado Martín, in press).

### 3.3 Reconstruction of Shadow Manifolds

Using the windowing technique, for each child we obtained four time series, one corresponding to each of the four measures described above: *loquacity*, *lexical diversity*, *inflectional diversity*, and *MLU*. These time series are plotted in Fig. 2

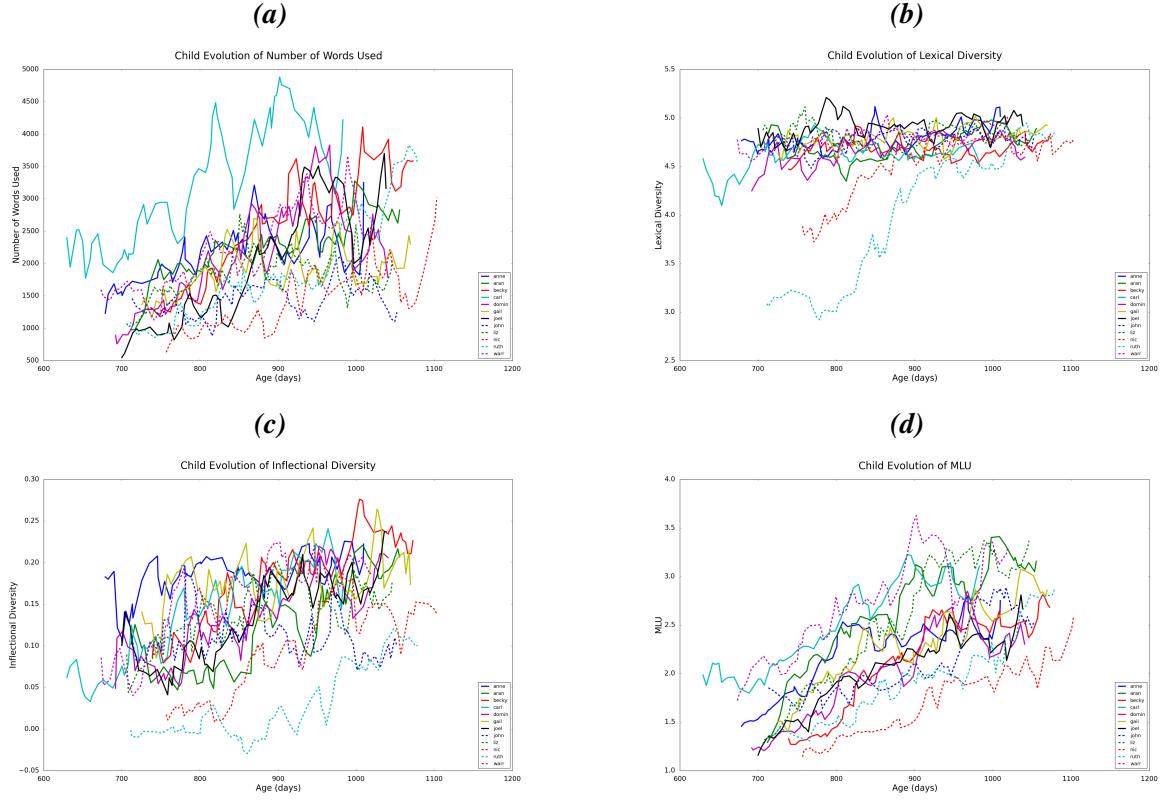


Figure 2: Evolution of the measures studied as a function of the children’s ages (in days). (a) Evolution of the loquacity (measured in number of word tokens produced) for each of the twelve children. (b) Evolution of the lexical diversity (measured in nats per word) for each of the twelve children. (c) Evolution of the inflectional diversity (measured in nats per word) for each of the twelve children. (d) Evolution of the MLU (in number of words per utterance) for each of the twelve children.

Parameter	Loquacity	Lexical Diversity	Inflectional Diversity	MLU
$\tau$	3	2	3	3
$E$	3	3	2	4

Table 1: Parameter values used in the reconstruction of the shadow attractors based on each of the four measures.

In order to ensure that applying the non-linear dynamics techniques on these time series was sensible, the series were checked to ensure that they contained non-linear signal not dominated by noise. This was achieved using a prediction test (Clark et al., 2015): We ensured that, for all four variables, the ability to predict future values significantly decreased as one increases the distance in the future at which the predictions are being made. This increasing unpredictability is the hallmark of non-linear dynamical systems. Therefore, we could safely proceed to reconstruct the shadow attractors.

Following Clark et al. (2015), we reconstructed

the shadow attractors from each of these collections of time series. The optimal time-lags ( $\tau$ ) for constructing the shadow manifolds were estimated as the first local minimum of the lagged self-information in each of the time series (c.f., Abarbanel et al., 1993). The optimal embedding dimensionalities ( $E$ ) were estimated by optimizing next-step prediction accuracy. The estimates were not found to differ significantly across children, and therefore for each measure, we used a single estimate of ( $\tau, E$ ) for all children. The estimated optimal parameter values used for the reconstruction of each shadow attractor are given in Table 1.

### 3.4 Detection of Causal Relationships

The presence of directional causal relations was tested for each of the six possible pairs of variables using mCCM. We performed 1,000 bootstrapping iterations for assessing the  $p$ -values of

the relations.<sup>1</sup> Finally, to account for our lack of *a priori* predictions on the causal directions to be tested, the *p*-values were adjusted for multiple comparisons using the false discovery rate procedure for correlated data (FDR; Benjamini and Yekutieli, 2001).

## 4 Results

As plotted in Fig. 2, the four groups of time-series considered here exhibit different patterns of development. On the one hand, the loquacity, inflectional diversity, and MLU series show evidence of a more or less linear increase along the child’s development, with their values towards the end of the studied interval being close to what was found for their mothers in those same conversations. On the other hand, the lexical diversity measure exhibits quite constant patterns across all children, with their values being pretty much indistinguishable from those observed for their mothers. This latter pattern is slightly different in two children (Ruth and Nick), who seem to be experiencing their ‘vocabulary burst’ later than the rest of the children did. In fact, if one examines panel (c) in detail, one sees that the linear increase in the inflectional diversity curves for these two children, only beginning its linear increase *after* they have experienced the vocabulary burst. A similar pattern can be seen in the MLU curves (panel (d)) for these two particular children, with syntactic development apparently being delayed by their late vocabulary bursts. These two patterns suggest that the development of both grammatical components of their language (inflectional morphology and syntax) depends on having attained a certain degree of vocabulary richness. However, just examining these curves does not provide evidence of whether these relationships are actually reliable, or they are just statistical mirages. The mCCM method addresses such question directly.

Fig. 3 plots the results of the mCCM method for each pair of reconstructed shadow manifolds. The curves plot how the correlations between nearest neighbors across shadow attractors evolve as one considers increasingly larger library sizes. The *p*-values report whether these correlation values are significantly increasing (the *p*-values are obtained by a Monte Carlo method with 1,000 resamplings, and further corrected for the twelve comparisons

using the FDR procedure).

Using the *p*-values in Fig. 3 enables the reconstruction of the network of causal relations depicted in Fig. 4. In this graph, the causal relation between the loquacity and the lexical diversity is considered weaker than the rest. The reason for this is that the comparisons reported here are in fact part of a larger study considering many more comparisons (including many factors of the mothers as well), on which we did not have any clear *a priori* predictions on the relations that would be found. When applying the FDR method on the whole set of 56 comparisons that we actually considered, the relation plotted by the dashed arrow is in fact not significant. In short, one should not trust the reliability of that particular relation.

Considering only the fully reliable relations, one finds that, as was suspected from the curves in Fig. 2, there is an explicit causal relation between the development of vocabulary richness (i.e., lexical diversity) and the acquisition of inflectional paradigms (i.e., inflectional diversity). The increase in lexical diversity indeed *causes* the development of inflectional paradigms. In turn, that the inflectional paradigms begin to be in place enables the child to begin generalizing more syntactic relationships (as is reflected by the strong bidirectional coupling of the inflectional diversity and MLU manifolds). Importantly, that the inflectional paradigms are developed is also strongly coupled with the increase the children’s overall loquacity; once the children begin to get a hold of grammar (inflection and syntax) they are enabled to speak more, which furthers their ability to generalize morphological relations and –by association– syntactic relations.

## 5 Discussion

In this study, we have –for the first time– documented the explicit *causal* relations between different tiers of children’s linguistic development. At a macroscopic level, we find strictly causal relations between the acquisition of vocabulary, inflectional paradigms, and syntactic relationships. As schematized in Fig. 4, the development of a sufficiently large vocabulary is a crucial trigger for the successful acquisition of the grammatical aspects of language, which are in turn necessary for children to be able to speak more. These results are consistent with what theories advocating the importance distributional learning for the acqui-

<sup>1</sup>All computations were done using R package `multispatialCCM` (Clark et al., 2015).



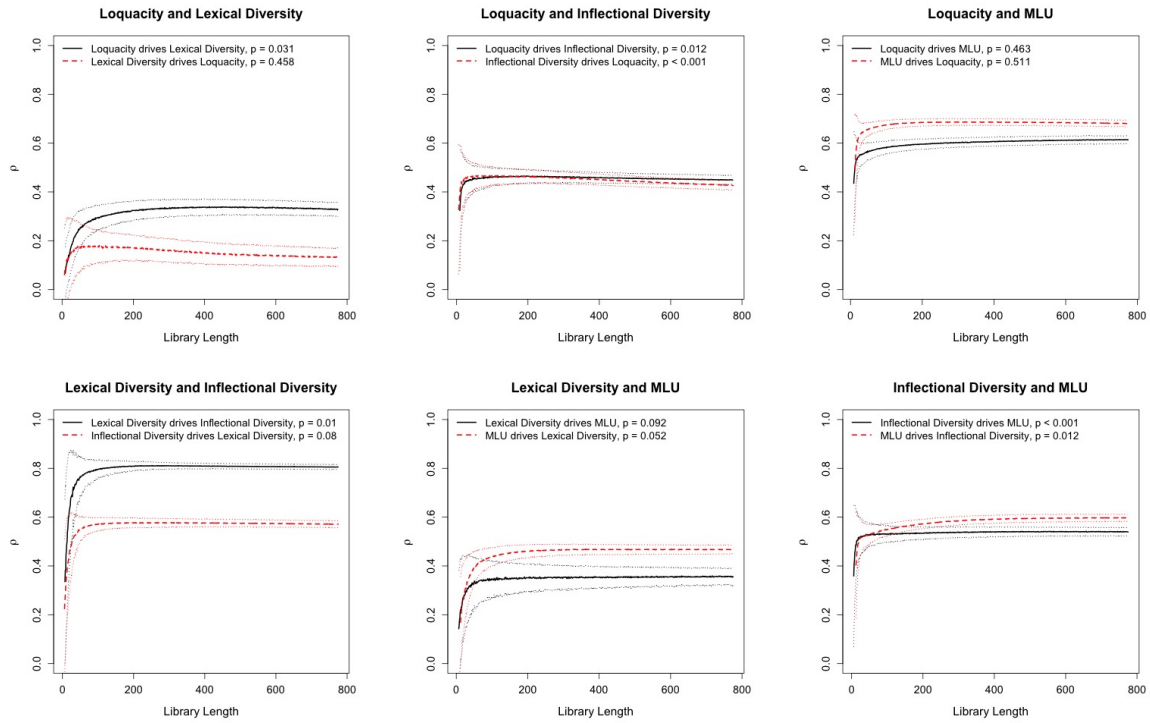


Figure 3: Results of mCCM between each pair of variables. The  $p$ -values are FDR-corrected considering the twelve comparisons reported here.

sition of constructions (Lieven et al., 1997; Pine and Lieven, 1993; Tomasello, 1992; Tomasello, 2005). It shows how the level of generality of the constructions is progressively increased (Goldberg, 2003) by the use of ‘lexically-based positional analysis’ (Lieven et al., 1997) to achieve early grammatical generalizations.

The picture of causal relations observed here could be put into an informal narrative as follows: The acquisition of sufficient lexical forms enables children to generalize their relations into inflectional paradigms. When a sufficient command of the language’s inflectional morphology has arisen, children are able to begin generalizing syntactic relations. The presence of these early syntactic developments in turn serves to increase the child’s awareness of the functional roles served by different paradigm members. From this point, one observes the strong bidirectional coupling between the development of syntax and inflectional morphology. An increasing awareness of the functional roles of the individual forms within these paradigms, and noticing the formal relations between them, in turn trigger further generalizations of the paradigms into *inflectional classes* (Milin et al., 2009), further increasing the productivity of the inflectional morphology system.

This study also stresses the importance of *macroscopic* level linguistic analyses. Whereas much research in language acquisition has focused on the acquisition of specific individual constructions (*microscopic* level) or groups thereof (*mesoscopic* level), the investigation of the properties of the whole lexicon, inflectional and syntactic systems uncovers relations which are difficult to pinpoint at the other levels. This fits in well with the multiscale investigation of language development proposed from the point of view of the Theory of Dynamical Systems (van Geert, 1991). Indeed, one can see, at the mesoscopic level, that –also consistent with the distributional learning hypothesis– there is a causal chain by which the development of single word utterances triggers the development of two-word utterances, which in turn trigger three-word utterances, and so forth (Bassano and van Geert, 2007). The macroscopic analyses provided here complement that picture by indicating how that evolution of utterance lengths is strongly coupled with the development (or ‘growth’ in van Geert’s terms) of the grammatical knowledge.

An innovative aspect of the methods we have developed in this paper provide is that they provide an explicit procedure for testing whether there are

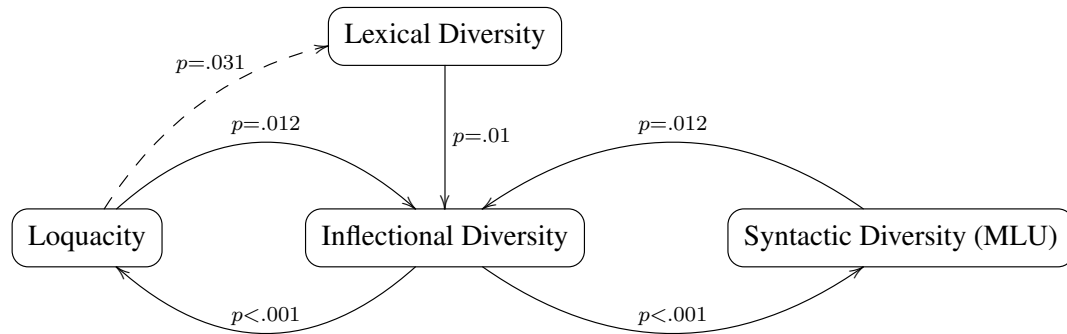


Figure 4: Reconstructed network of causal relations between the different measures of the children's linguistic performances. The  $p$ -values indicated on the causal arrows are FDR-corrected. The dashed-line denotes a relationship that does not survive FDR correction considering a larger set of variables.

explicit causal relations between the development of different aspects of language. Here we have used the methods at a macroscopic level, but it would be equally possible to apply them to both microscopic- or mesoscopic-level time series. Previous research on dynamical systems on language acquisition (e.g., Bassano and van Geert, 2007; Steenbeek and van Geert, 2007; van Geert, 1991) rely on proposing different candidate models in terms of systems of differential equations, each including different sets of causal relations and couplings between time series. Our methods, using techniques for explicitly testing causal relations borrowed from ecology (a field whose study bears uncanny similarities with the study of human development), complement the curve-fitting by explicitly testing which couplings and causalities should be included in the models, thus significantly reducing the model space that needs to be explored.

## References

- Henry D. I. Abarbanel, Reggie Brown, John J. Sidorowich, and Lev Sh. Tsimring. 1993. The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65:1331–1392.
- R. Harald Baayen and Fermín Moscoso del Prado Martín. 2005. Semantic density and past-tense formation in three Germanic languages. *Language*, 81:666–698.
- Dominique Bassano and Paul L. C. van Geert. 2007. Modeling continuity and discontinuity in utterance length: a quantitative approach to changes, transitions and intraindividual variability in early grammatical development. *Developmental Science*, 10:588–512.
- Yoav Benjamini and Daniel Yekutieli. 2001. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29:1165–1188.
- Melissa Bowerman. 1990. Mapping thematic roles onto syntactic functions: are children helped by innate linking rules? *Linguistics*, 28:1253–1289.
- Roger Brown. 1973. *A first language: the early stages*. Harvard University Press, Cambridge, MA.
- Anne Chao, Y. T. Wang, and Lou Jost. 2013. Entropy and the species accumulation curve: a novel entropy estimator via discovery rates of new species. *Methods in Ecology and Evolution*, 4:1091–1100.
- Adam Thomas Clark, Hao Ye, Forest Isbell, Ethan R. Deyle, Jane Cowles, G. David Tilman, and George Sugihara. 2015. Spatial convergent cross mapping to detect causal relationships from short time series. *Ecology*, 96:1174–1181.
- Adele E. Goldberg. 2003. Constructions: a new theoretical approach to language. *TRENDS in Cognitive Sciences*, 7:219–224.
- Chih-hao Hsieh, Christian Anderson, and George Sugihara. 2008. Extending nonlinear analysis to short ecological time series. *The American Naturalist*, 171:71–80.
- Jeremy Irvin, Daniel Spokoyny, and Fermín Moscoso del Prado Martín. in press. Dynamical systems modeling of the child–mother dyad: Causality between child-directed language complexity and language development. In Anna Papafragou, John Trueswell, Dan Grodner, and Dan Mirman, editors, *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, TX.
- Elena V. M. Lieven, Julian M. Pine, and Gilliam Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language*, 24:187–219.



- Brian MacWhinney. 2000. *The CHILDES Project: Tools for analyzing talk*, volume 2: The database. Lawrence Erlbaum Associates, Mahwah, NJ, 3rd edition.
- Petar Milin, Dušica Filipović Đurđević, and Fermín Moscoso del Prado Martín. 2009. The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language*, 60:50–64.
- George Miller. 1955. Note on the bias of information estimates. In Henry Quastler, editor, *Information Theory in Psychology: Problems and Methods*, pages 95–100. Free Press, Glencoe, IL.
- Fermín Moscoso del Prado Martín, Aleksandar Kostić, and R. Harald Baayen. 2004. Putting the bits together: an information theoretical perspective on morphological processing. *Cognition*, 94:1–18.
- Fermín Moscoso del Prado Martín. in press. Vocabulary, grammar, sex, and aging. *Cognitive Science*.
- Matthew D. Parker and Kent Brorson. 2005. A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, 25:365–376.
- Julian M. Pine and Elena V. M. Lieven. 1993. Re-analysing rote-learned phrases: individual differences in the transition to multiword speech. *Journal of Child Language*, 20:43–60.
- Steven Pinker. 1994. *The language instinct: How the mind creates language*. Harper-Collins, New York, NY.
- Claude E. Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.
- Linda Smith and Esther Thelen. 2003. Development as a dynamic system. *TRENDS in Cognitive Sciences*, 7:343–348.
- Hederien W. Steenbeek and Paul L. C. van Geert. 2007. A theory and dynamic model of dyadic interaction: Concerns, appraisals, and contagiousness in a developmental context. *Developmental Review*, 27:1–40.
- Sabine Stoll, Bathasar Bickel, Elena Lieven, Netra P. Paudyal, Goma Banjade, Toya N. Bhatta, Martin Gaenszle, Judith Pettigrew, Ichchha Purna Rai, Manoj Rai, and Novel Kishore Rai. 2012. Nouns and verbs in Chintang: Children’s usage and surrounding adult speech. *Journal of Child Language*, 39:284–321.
- George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan R. Deyle, Michael Fogarty, and Stephan Munch. 2012. Detecting causality in complex ecosystems. *Science*, 338:496–500.
- Floris Takens. 1981. Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young, editors, *Dynamical Systems and Turbulence*, pages 366–381. Springer Verlag, Berlin, Germany.
- Anna L. Theakston, Elena V. M. Lieven, Julian M. Pine, and Caroline F. Rowland. 2001. The role of performance limitations in the acquisition of verb-argument structure: an alternative account. *Journal of Child Language*, 28:127–152.
- Michael Tomasello. 1992. *First Verbs: a Case Study of Early Grammatical Development*. Cambridge University Press, Cambridge, England.
- Michael Tomasello. 2005. Beyond formalities: The case of language acquisition. *The Linguistic Review*, 22:183–197.
- Paul L. C. van Geert. 1991. A dynamic systems model of cognitive and language growth. *Psychological Review*, 98:3–53.

## A Bias-Adjusted Entropy Estimator

The bias-adjusted entropy estimator (Chao et al., 2013) relies on properties of the accumulation curve of the number of distinct words observed (i.e., the species accumulation curve in the biological terms of the original paper). The estimator is given by

$$\hat{H} = \sum_{1 \leq F_i \leq n-1} \left[ \frac{F_i}{n} \left( \sum_{k=F_i}^{n-1} \frac{1}{k} \right) \right] - \frac{f_1}{n} (1-A)^{1-n} \left[ \log(A) + \sum_{r=1}^{n-1} \frac{1}{r} (1-A)^r \right],$$

where  $F_i$  are the word frequencies observed in the sample,  $n$  is the number of tokens in the corpus, and

$$A = \begin{cases} \frac{2f_2}{(n-1)f_1 + 2f_2} & \text{if } f_2 > 0, \\ \frac{2}{(n-1)(f_1-1) + 2} & \text{if } f_2 = 0, f_1 > 0, \\ 1 & \text{if } f_1 = f_2 = 0, \end{cases}$$

with  $f_1$  and  $f_2$  being the number of word types that were encountered exactly once or twice respectively (i.e., the numbers of *hapax legomena* and *dis legomena*). This estimator is demonstrated to be accurate and unbiased for word frequency distributions (Moscoso del Prado Martín, in press).