Linguistics Background

- Understand the causal links between complexities of child language and child-directed language
  - how does the evolution of child directed language affect the child's language and vice versa?
- CDL: language used by mother when talking to young children
- observed that lexical/syntactic complexity of CDL gradually increases as child develops
- Fine-tuning: the process by which caregivers adjust the complexity of their language according to the complexity of the child's language used or understood
  - weak: mother globally adapts based on her knowledge of child's abilities (mother causes child)
  - strong: mother reacts to specific and local cues (child causes mother)
  - strong remains the dominant view in the field, considered fact by many researchers

Dynamical Systems Background

- Problem with creating a dynamical system model is that not all contributing variables may be available or easily measurable, or even known to exist. this will be addressed using Taken's Theorem
- using a dynamical systems approach provides an interesting method of understanding linguistic growth, namely considering measurements of CL and CDL as time series trajectories from a high dimensional dynamical system
- we want to discover how these measured variables affect each other. this goes beyond correlation and requires the ability to ascertain asymmetrical causal relations

Previous Causality Techniques

- Pearl: If A and B are correlated, then at least one of the following must hold:
  1. A causes B
  2. B causes A
  3. A third variable C causes A and B
- Granger-causality: Y Granger-causes X if the predictability of X decreases when Y is removed from the universe U of all possible causative variables
  - The past of Y predicts the future of X over and above the past of X
- Shortcomings:
  - Detecting direct causal relation requires knowledge of all relevant variables
  - Granger assumes separability: the information about a causative variable is independently unique to it, and that information can be removed by eliminating that variable from the model
    - If Y is a cause for X, likely information about Y will be redundantly present in X itself and cannot be formally removed from U.
    - If there were a third variable Z causing both X and Y, Granger-causality would incorrectly infer causality (that Y causes X) in the non-separable case because Y would contain information about the evolution of X (so predictability would likely decrease with its removal) but this is simply because the information in Y is not unique to it - it it shared by both Y and Z.
    - purely stochastic systems often demonstrate separability. however it is not exhibited by deterministic non-linear dynamical systems
  - Difficult to detect weakly coupled variables

CCM

- Sugihara et al.* proposed a causality detection technique which is valid for non-separable systems called Convergent Cross Mapping.
- This powerful technique can identify weakly coupled variables in the presence of noise, and most importantly, can distinguish causal relations between variables from effects of shared driving variables.
  - CCM would not find causality if a variable is causing both variables under consideration

E. Lorenz's Dynamical System

- Lorenz's well-studied dynamical system consists of three coupled variables $X(t)$, $Y(t)$, and $Z(t)$ whose co-evolution is described by the system of differential equations:

- The first equation indicates a relation that Y causes X because the change in X, (ie, its future value) depends on the value of Y.
- Its strength is indexed by parameter σ.

Taken's Embedding Theorem

- However, in many situations, not all variables relevant to the system are available.
  - difficult to measure, not aware of their relevance, etc.
- Taken's Embedding Theorem* allows us to recover critical properties of the coupled dynamical system's attractor using only measurements from a single one of its variables.
- These reconstructed manifolds are called "shadow" manifolds, and maintain many properties important to the original system.
- Each point in the original manifold M maps onto points in its shadow manifolds, as seen in the points $m(t), x(t), y(t)$.
- Because the embeddings preserve topological properties of the underlying attractor, the points corresponding to close points in this manifold will remain close in the embeddings.
- This means that, for causally linked variables within the same dynamical system, the state of one variable can identify the states of the others.

CCM Causality Detection

- Sugihara et al. 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.
- fish time series can be used to estimate weather
- To test for causality, CCM looks for the effect of X in Y's time series by determining if nearby points on Y's shadow manifold can be used to identify nearby points on X's shadow manifold (X causes Y)
- To distinguish causation from correlation, CCM requires convergence, ie, that cross-mapped estimates improve in estimation accuracy with the sample size (library size) used for reconstructing the manifolds.
- as the library size increases, the manifold trajectories fill in (become denser), resulting in closer nearest neighbors and declining estimation error, reflected in a higher correlation coefficient between points in the neighborhoods of the shadow manifolds
- sugihara demonstrated that this technique successfully recovers true directional causal relations when they are present, and does not discover causal relation when a third variable causes both variables and no true direct causation exists between them
- CCM generally require relatively long time series. but you can instead obtain multiple short time series from the same dynamical system. multispatialCCM extends the CCM technique and can infer causal relations from multiple short time series using a bootstrapping technique called dewdrop regression.

Supplementary

- Children were clustered
  - Diana (DIvisive ANAlysis) with Manhattan Similarity Metric
  - For n points, algorithm splits (n - 1 times) the Cluster with the largest diameter
  - Once Dendrogram formed optimal number of clusters is chosen
- Also explored inflectional diversity (difference between entropy of unlemmatized words and entropy of lemmatized words)
  - No causal relationship was discovered.
  - This, however, still contributed to the FDR adjusting methods.
- There are well-studied prescriptive methods* to choose the time lag τ and embedding dimension E.
  - If τ is too small, the adjacent coordinates will be so close numerically that they are indistinguishable (and not independent).
  - If τ is too large, the resulting embedding will be a projection onto two completely unrelated directions.
  - τ is normally chosen to minimize the average mutual information, which yields coordinates that are independent but not probabilistically independent.

- E needs to be large enough so that points on the attractor can be unfolded to create the embedding without ambiguity.
  - Two points close in a certain dimension should be so because it is a property of the points, not because the dimension is small.
  - Taken's provides us with a sufficient dimension, but we still want a small E to minimize computation time.
  - R package multispatialCCM does this automatically, choosing the dimension to maximize predictive ability (rho).
- Multiple Comparisons: when one evaluates a set of statistical inferences as a whole, it is much more likely for hypothesis tests to yield a Type I error (incorrectly rejecting the null hypothesis, meaning the tests will incorrectly determine there is a causal relation between the variables)
- FDR: FDR controlling procedures are less conservative (they provide no direct bound on FWER - this change is insignificant to Bonferroni correction which provides a direct bound on the FWER), but they result in more Type 1 errors