Systematicity over the course of early development: an analysis of phonological networks

Catherine E. Laing[1]

[1] University of York, York, UK

Author Note

All code and associated data for this manuscript can be found at https://github.com/cathelaing/NetworkGraphs. Data provided in the PhonBank corpus was collected through support from grant number NIH-NICHHD RO1-HD051698.

Correspondence concerning this article should be addressed to Catherine E. Laing, Department of Language and Linguistic Science, University of York, Heslington, YO10 5DD.. E-mail: catherine.laing@york.ac.uk

Abstract

This paper explores the early lexicons of nine infants acquiring English or French to determine the extent of systematicity in the early vocabulary, and how this changes over time. Network graphs are generated from the point of first word production in the data set until age 30 months. Two measures of systematicity - mean path length and clustering coefficient - are analysed to establish the extent to which the early productive lexicon consists of closely-connected clusters of similar-sounding forms. Results show that early production is highly systematic when compared to random, prototypical and adult phonological networks, and that over time the infant network constitutes an increasing number of dense clusters of similar-sounding forms. Generalised additive mixed effects models support these findings.

Keywords: systematicity, phonological development, preferential attachment, networks analysis

Word count:

Systematicity over the course of early development: an analysis of phonological networks

Infants' early words are phonologically similar to one another, if not in the vocabulary items they choose to produce, but in the way they produce them. This has been well-documented in a number of previous studies (Szreder, 2013; Vihman, 2016; e.g. Waterson, 1971), and can be clearly observed in datasets of early productions. For example, an inspection of Deuchar and Quay's (2000) record of their Spanish-English bilingual child's vocabulary acquisition shows that many of her earliest words are produced with an open CV syllable, and she produces a number of identical forms to refer to a range of different (though phonologically-similar) words. This suggests that infants may be drawing on a systematic approach to early productions, whereby a small subset of simple phonological forms are used to produce a range of more varied and phonologically-challenging adult targets. We would expect, therefore, that newly-acquired productions 'cluster together' with existing forms, with high phonological similarity between new words and existing words in the lexicon. This paper will draw on network graphs of the early vocabulary to determine how similar infants' words are to one another, and how this changes over time.

Network analysis is an increasingly popular method of analysing lexical acquisition. Network models allow the analysis of connectivity within a system (in our case, the lexical or phonological system), and can track how that connectivity changes over time. In the case of language development, the nodes (individual items) in the network typically consist of words, and these are connected (or not) depending on how similar two nodes are in phonological or semantic space. Two words that are more similar to one another will be positioned closer together in the network, and less similar nodes further apart. Because this is a convenient way to think about language development over time, a number of studies have considered vocabulary acquisition within this framework. Analyses have considered infants acquiring their first language (e.g. Amatuni & Bergelson, 2017;

Fourtassi, Bian, & Frank, 2020) and adults acquiring a novel language (e.g. Mak & Twitchell, 2020; Siew & Vitevitch, 2020).

Systematicity in early phonological development is most comprehensively discussed in work by Vihman (e.g. Vihman, 2016, 2019; Vihman & Croft, 2007; Vihman & Keren-Portnoy, 2013). In more than four decades of work, incorporating data from a large number of infants acquiring an impressive range of languages, Vihman demonstrates a clear systematicity in infants' path to target-like word production. From the initial 'surprisingly accurate' forms that appear in the first stages of word production, infants are shown to draw on what they know: they generally choose, for first production, words that are simple in their target phonological form, with consonants that are already familiar from the most common syllables of canonical babble (McCune & Vihman, 2001). These forms are, in Vihman's terms, *selected* for first word production owing to their easily-producible features. As the vocabulary grows, infant must necessarily acquire forms that do not contain such accessible phonological or segmental properties. Here we begin to see regression in the accuracy of early production, as infants systematically adapt forms to fit the most common structures and segments in their repertoire. When words are systematically altered to fit a dominant pattern in the child's output, these forms are said to be *adapted.* Systematicity is apparent not just in the earliest words, but across the trajectory of acquisition as infants deal with the challenges of early word production by relying on well-rehearsed output forms. Over the first months of lexical development at least, infants' productions of newly-acquired words are likely to match their productions of existing words in the lexicon.

In recent work (Laing, under review), I draw on network analysis to analyse systematicity in the developing lexicon. I show that in the first three years of life, infants' production of new words can be predicted based on the words they already produce. That is, a word is more likely to be acquired if it is phonologically similar to existing words in the productive repertoire, particularly when the new word is similar to a cluster of existing phonologically-similar forms in the output. This effect becomes stronger over time, suggesting that systematicity is more relevant to later word learning (at least, upto age 30 months) than in the first few months of word production. Moreover, the phonological properties of infants' word productions are more similar to one another than their adult target forms would suggest. These findings support the more fine-grained analyses presented by Vihman and colleagues.

~~A number of other studies have found consistent results when analysing early vocabulary development.~~ Laing (under review) drew on corpus data of fortnightly word production across nine infants; ~~similar results have been observed by~~ Kalinowski and colleagues (in prep), who draw on vocabulary norms from >1000 Norwegian infants taken longitudinally at up to six individual time-points. Systematicity was also identified in early word learning by Siew and Vitevitch (2020), in an analysis of vocabulary norms of children aged 3-9 years acquiring English and Dutch. Fourtassi and colleagues (2020), on the other hand, found no support for systematicity in their analysis of infants acquiring a range of 10 different languages, instead showing that salient properties of the input, rather than previously-learned phonological properties, predicted learning.

This previous work uses network growth algorithms to test whether learning can be predicted based on the words that infants already know or produce. This method works on the assumption that, if infants' early word acquisition and production is phonologically systematic, then it should be possible to predict learning based on the known words in the infant's existing lexicon. This is a compelling computational approach to observing systematicity in the developing lexicon, but does not give us a clear view into the extent to which early-acquired forms are produced in a phonologically similar way. That is, network growth models analyse connectivity (are two words similar, yes or no? if yes then they are connected in the network), rather than phonological distance (*how* similar are two words in the network?). This study expands on previous work, and builds on findings from Laing (under review), by analysing network graphs of infants' early lexicons. This provides an insight into the properties of the network, including how densely-connected the network is, and the phonological distance between items in the network. It also allows us to make predictions about network properties over time that reflect word selection and adaptation. Following Laing (under review), this paper will analyse infants' actual productions as well as the adult target forms, to test the extent to which systematicity is present in production, in terms of both the words infants select in early development, and how they produce them.

Research questions and predictions

This paper analyses network graphs ~~instead of growth algorithms~~ to look more closely at the phonological distance between individual words in the developing lexicon. In doing so, it attempts to address the following questions:

1. How systematic are early word productions, and (how) does this change over time?
2. Are the phases of word selection and adaptation identifiable in the dataset?

To test these questions, network graphs will be generated using the *igraph()* package (Csardi & Nepusz, 2006) in R (R Core Team, 2020). To address the first question, properties of the graphs will be analysed to determine 1) how closely connected individual words are to one another; 2) how dense the overall distribution of words is in the network; and 3) how/whether this changes over time. Following Vihman's work, and findings presented by Laing (under review), it is expected that the early vocabulary will become increasingly systematic over time. This would be reflected in denser clusters of phonologically-similar forms and ~~lower~~ distance between words. Simulated networks will be used to compare the real networks against both highly systematic and random networks to determine the extent of systemacitity present in the data.

To address the second question, network graphs of infants' actual productions will be compared with those of the target form, to trace the 'target-likeness' of individual productions, and how this changes over time. Following Vihman once again, early word selection would be reflected in early similarities between Actual and Target network properties, as target forms are selected to match the structures and segments that infants are able to produce, meaning they should be produced with relative accuracy. Over time, Actual and Target forms are expected to diverge, such that Actual forms show more systematicity in the data than Target forms. Approaches used to test these predictions are outlined in detail below.

Methods

Data extraction and preparation

This study draws on the same data as that analysed by Laing (under review). This was drawn from two corpora on PhonBank (Rose & MacWhinney, 2014): Providence (American English, Demuth, Culbertson, & Alter, 2006) and Lyon (French, Demuth & Tremblay, 2008). These were selected due to their equivalent data collection methods and the fact that the infants' productions, as well as the corresponding target forms, are phonetically transcribed. Nine infants' (5 English, 4 French, 4 boys overall) data were extracted using Phon (Hedlund & Rose, 2020), from the transcript with their first-recorded word to the final transcript taken at age 2;6. Infants were recorded in the home on a fortnightly basis, participating in naturalistic interactions with their caregivers. Two of the American infants were recorded weekly during some periods of data collection, but this is not an issue for this analysis since no between-child comparisons will be made. See Demuth et al. (2006) and Demuth and Tremblay (2008) for full details of data collection.

Extracted data was filtered to include only words featuring on the communicative development inventory (CDI) of the respective language, including all variants of a given form. For example, plurals were included alongside the singular noun form, and variable verb conjugations alongside their corresponding infinitive verb form. These data were used to create a network for each infant for each month of data, whereby all words in a given month that hadn't been produced in previous months were included as new items in the network. When a word had been produced in previous months, it was not included. While this means that the data does not capture change in the production of a single form over time, it allows us to observe network growth at the point of acquisition for each word form.

To determine the structure of the network, the first step was to create distance values between each word and each other word in the network. This was first done using a 'global' network of all forms produced by the infant up to the final session at 2;6, to create a large distance matrix for each infant that incorporated all word productions. Essentially, this global network reflects the distance between every word and every other word in each child's productive vocabulary at 2;6. Distance values were established using methods set out in Monaghan et al. (2010), using distinctive features to generate a set of phonetic values for each word that could then be compared with all other words [note that only consonants were analysed, given that vowels are highly variable in early production and also very difficult to transcribe accurately; Donegan (2013); Kent and Rountrey (2020)]. Euclidean distance between the values of each word and each other word in each infant's global network was then used to determine how close/distant words were from one another. Often, infants produced multiple tokens of the same word type in a given month, often with high variability across tokens. Because it was not possible to generate networks with all tokens included (even with only single word types included, the final dataset for all nine infants includes over 5.5 million data points, once distance between each word and each other word is calculated), a mean value for each distinctive feature was established across tokens, meaning that each word's distinctive feature value represents the variability of the infant's production of a given word. This may not be a perfect measure, but it is more representative than taking, for example, the first instance of each word type.

Distance scores were generated between each word and each other word in each child's dataset, for both Target and Actual forms. These scores were then normalised, and a normalised distance of 0.25 was chosen to indicate connectivity. That is, words were said to be connected in the network if their distance score was 0.25 or less. This accounted for the lower quartile of connectivity across the dataset. See Supplementary Materials for an overview of how this measure was established.

The final dataset includes 3223 word types in total, with 1927 in the English corpus and 1296 in the French corpus. On average, infants produced 47 tokens of each word type in a single session (SD = 170; mean English tokens = 40, SD = 143; mean French tokens = 58, SD = 202).

Data analysis

Network graphs. The prepared data was then used to generate a series of network graphs for each infant (for both Target and Actual data) using the *igraph()* package in R (Csardi & Nepusz, 2006). One network was generated per month, for each month in the dataset, based on all words produced in the given month and all months prior. The network at time-point n+1 thus included all words produced upto and including time-point n, plus all additional words produced for the first time at n+1. The *igraph()* package generates graphs that include all nodes (whether or not they are connected to other nodes[1]), and measures the distance between all connected nodes, as well as the clustering of nodes in graphical space.

Two key variables will be explored through an analysis of network graphs: *mean path length* and *clustering coefficient*. Path length is a measure of distance between nodes, and mean path length indexes the average phonological distance (of all connected nodes) within a network; by this measure, we would expect that systematicity in early phonological development would be reflected in low mean path length. Clustering coefficient is an indication of network density: a higher density of nodes in the network indicates denser clusters of similar forms; again, this is what we would expect to see in a network of early

[1] Recall that any two nodes that have a scaled phonological distance of >.25 will not be connected.

phonological development. See Goldbeck (2013) for a full overview of network structures and measures.

Simulated networks. Networks with high phonological systematicity should exhibit properties of prototypical “small-world” network growth, namely a low mean path length and a high clustering coefficient (Amaral, Scala, Barthelemy, & Stanley, 2011; Steyvers & Tenenbaum, 2005; Watts & Strogatz, 1998): words should be more densely connected, with shorter connections between words. To test this, mean path length and clustering coefficient values were generated for both the Target and Actual networks, as described above. Data were compared to the growth of a simulated small-world network of equal size, known as a Watts-Strogatz network (Watts & Strogatz, 1998); if no statistical differences are observed between the real and simulated data over the trajectory of development, this would reflect strong systematicity in growth of the real network. The real data were also compared to a similarly-sized but randomly-generated network known as a Erdős–Rényi model. Similarly, if the real network grows in a systematic way, then we would expect the real data to differ significantly from the randomly-generated Erdős–Rényi network. To run these analyses, mean path length and clustering coefficient were calculated for each monthly graph, and then compared to the two kinds of simulated data, matched for network size: a Watts-Strogatz network (prototypical systematic network) and an Erdős–Rényi network (random network). Watts-Strogatz simulations were generated to match both the mean connectivity and the network size of the network in each month; Erdős–Rényi networks require only network size to be specified.

Results

RQ1: How systematic are early word productions, and (how) does this change over time?

To test RQ1, network graphs were compared to simulated small world and random networks of equivalent size to determine whether the data (the *Real* network, drawing from infants' Actual productions) differed from the *Simulated* networks. For both mean path length and clustering coefficient, we would expect the Real data to differ significantly from the Simulated Erdős–Rényi (random) network, and for the Real network to show similar properties to the Simulated Watts-Strogatz (small-world) network. Note that the extent of the expected statistical difference is not easy to predict here: if the Real network is very similar to the Watts-Strogatz network then no statistical difference would be expected, but this relies on the Real data being highly systematic, which may not be realistic. In order to fully understand the nature of the data, figures and model outputs will be inspected in relation to these predictions.

The two measures will be discussed in turn. Models include mean path length or average clustering coefficient as dependent variables, respectively, each with data type (Real vs. small-world vs. random), corpus (English vs. French) and age as fixed effects, and subject as a random effect with a by-subject random slope for the effect of age. Initial model comparisons showed that including network size, alongside age, improved fit in the model testing clustering coefficient, but not mean path length. Network size is thus included as a fixed effect in the clustering coefficient model only.

Mean Path Length. Nested model comparisons revealed a significant effect for data type on mean path length. See Table 1. As shown in Table 2, the Real data had a significantly lower mean path length than the random Simulated data, as predicted. However, the Real data also differed significantly from the small-world Simulated data, which had a significantly shorter mean path length than that of the Real data. Model outputs revealed no effect for corpus or age on the data. Contrary to predictions, there was no change in systematicity over time (i.e. no effect of Age was observed), at least in terms of mean distance between words.

Clustering coefficient. The same main outcomes were found when clustering coefficient was tested in the model. See Tables 1 and 2. The Real data had a significantly higher clustering coefficient than the random Simulated data, but this was significantly lower than that of the small-world Simulated data. Again there was no effect for corpus on the data, but this time there was a significant effect for age: clustering coefficient values increased over time, suggesting that systematicity became stronger in the data over time. However, a significant effect for network size contradicts this; with each additional word added to the network, clustering coefficient decreases by 0.10%.

Actual vs. Target data. The differences between the Real data and the small-world Simulated data are difficult to interpret, given that there is no clear model of what phonological systematicity would look like in a highly systematic small-world network. To further interrogate systematicity within the data, network properties of the Real (Actual) data were compared to the Real Target data. Target data serves as an appropriate proxy for connectivity and clustering within a "standard" phonological network, albeit a network that is constrained by words produced in early acquisition, and further constrained by the fact that only a subset of these (i.e. CDI words) are included in the data set. We would expect that mean path length and clustering coefficient would each show higher systematicity in the Actual network than the Target network[2]. Basic model structure was the same as reported above, but with only Real data (Actual vs. Target) considered as a fixed effect. Network size improved model fit for both dependent variables and so was included in both models.

[2] Note that this result was observed for network growth models in Laing (under review), whereby the Actual network was found to be a better predictor of learning based on the known network of each child.

There was a significant effect for data type on mean path length. See Table 1. Model outputs revealed that Target data had a significantly higher mean path length than Actual data (see Table @ref(tab:table-actual-target and Figure 1. This time there was a significant effect for corpus, whereby the French data had a higher mean path length than the English data overall. Network size, but not age, significantly affected mean path length, which increased as network size increased, indicating a decrease in systematicity over time.

The effect of data type on clustering coefficient was also significant. Mean clustering coefficient was significantly lower in Target compared with Actual data. See Figure 2. The French data had a significantly lower mean clustering coefficient than the English data. Again, there was a significant effect for network size, but this was in the opposite direction as reported above for mean path length. As new words were acquired in the network, mean clustering coefficient increased by 0.05%, indicating an increase in systematicity over time.

RQ2. Is there evidence of word selection and adaptation in the dataset?

To address the second research question, the phonological distance between Target and Actual forms was taken as a proxy of word selection and adaptation. That is, if a word is produced in a target-like way (i.e. assumed to be selected[3]), then the phonological distance between the Target form and the way it is produced (Actual form) should be low. The opposite is true for adapted forms, as we expect, by definition, a non-target-like production and thus a higher distance between Target and Actual form. This measure is not perfect, but coding selected/adapted forms would otherwise have to be done by hand, which is not feasible across such a large dataset. Following Vihman's (2019) framework, we would expect low distance been Actual and Target forms earlier on in development, and higher distance later. To test this, I drew on generalised additive mixed effects models (GAMMs), since these allow the analysis of non-linear change over time, and can account

[3] though note that, while a selected form is, by definition, target-like in phonological form, a target-like form isn't necessarily a selected form.

for statistical differences between two non-linear trajectories of data that may differ in non-linear ways (Sóskuthy, 2017; Wieling, 2018).

First, GAMMs were used to examine connectivity of the infants' Actual and Target networks and how these changed over time. These were run using the *mgcv()* package in R (Wood, 2011). These models analyse the extent to which the two networks differ (or not) from one another across infants, and how this changes non-linearly month-by-month. Fixed effects in the model can include parametric terms, as is typical in regression modelling, and also *smooth terms*, or non-linear fixed effects. Much like mixed-effects linear models, GAMMs can account for random effects in the data; in this case by-subject random effects were included through the addition of *random smooths* in the model.

The model tested number of connections in the network (*mean k*) as the dependent variable, working on the assumption that connectivity in the Target vs. Actual networks would be similar during periods of word selection, and would differ during periods of adaptation. Specifically, periods of adaptation should lead to higher connectivity in the Actual network than the Target network, since we expect productions to be more similar (and thus more well-connected) in Actual forms; we would expect connectivity across data types to diverge at the point that word adaptation begins to take hold. A higher number of connections (higher mean k) for Actual vs. Target networks is expected during periods of adaptation, and no difference in connectivity is expected for periods of selection. Data type (Actual vs. Target) and corpus (English vs. French) were included as parametric terms, with data type being the variable of interest in the model. Network size and age were included as smooth terms, as well as by-subject and by-data type random smooths for the effect of age, which account for by-subject and by-data type differences in the data over time.

To account for the fact that adjacent values (i.e. connectivity at month n and month $n{+}1$) are likely correlated, GAMM modelling includes an autocorrelation parameter; see

Sóskuthy (2017) and Wieling (2018) for full details. Additionally, the start point for each infant's data (i.e. their first recording session) was indexed in the model. To test for an effect of data type, model comparisons were run using the *compareML()* function from the *itsadug()* package (Rij, Wieling, Baayen, & Rijn, 2022): the full model including the effect of data type and the by-data type random smooth was compared to a model without these terms. Because model summaries for GAMM smooths may be non-conservative (*Sóskuthy, 2017*), smooth plots will be observed alongside any significant effects to determine relevant trends in the data.

Model comparisons revealed a significant effect for data type on connectivity in the networks over time. See Table 4. Figure 3 shows the difference between Actual and Target connectivity over the course of development. The red line indicates periods of significant difference, showing that Actual vs. Target connectivity was significantly different throughout the period of analysis; Actual forms were always more well-connected than Target forms. This contrasts with the expectations set out above. However, Figure 3 clearly shows an increase in the difference in connectivity between Actual and Target forms over the period of data collection, supporting the expectation that Target and Actual forms are more similar earlier on in development, and implying that word selection takes place earlier on in the data, likely with some words being produced in a more target-like way than others.

Discussion

This paper set out to test the presence of systematicity in infants' developing lexicon by analysing phonological network graphs from nine infants acquiring French or English. Network graphs allow a close-up view of phonological similarity between forms within the network (via mean path length) and the extent to which groups of phonologically similar forms cluster together (via average clustering coefficient). Systematicity in the network would be reflected in shorter distance between forms and higher clusters of phonologically similar forms.

In previous work (Laing, under review) I found that network growth algorithms ~~could~~ predict word learning; the model successfully predicted which words would be produced next, based on how similar a new word was to the most well-connected words in the existing network. This outcome lends support to the hypothesis that early word production is phonologically systematic. Here, an analysis of the networks themselves reveals complimentary findings: when compared with a randomly-generated network of equal size, the phonological networks were found to be structured in a significantly more systematic way, in terms of both mean path length and clustering coefficient. Words were phonologically more similar and formed denser clusters than we would expect by chance.

When the phonological networks were compared with a prototypical ‘small-world’ network, which possesses a highly systematic structure by design (Watts & Strogatz, 1998), both mean path length and clustering coefficient were found to be less systematic than the prototypical model. To interrogate this further, a follow-up analysis comparing infants’ Actual productions with the Target forms was conducted. This allowed a consideration of network properties based on an ‘adult’ model, as opposed to a prototypical model of systematicity that does not necessarily reflect real-world linguistic phenomena (or indeed a random network, that may also not be a suitable comparison for the non-random distributions that occur in language). In this follow-up analysis, infants’ productions were found to be more systematic across both measures than we would expect from an analysis of Target forms alone, consolidating findings reported above and lending further support to initial predictions.

When systematicity in the networks was observed over time, findings were less clear, and differed between the two measures. Mean path length was not affected by changes in age, but it did increase with network size, at least in the model testing only Real (Actual vs. Target) data. This is opposite to what was predicted above, and contrasts with findings presented in Laing (under review), given that an increase in mean path length indicates decreasing systematicity, as nodes are more widely dispersed across the network. However, the opposite was true for clustering coefficient; both age and network size affected this measure in the initial model comparing Real vs. Simulated data, and in opposite directions. Increasing age led to a decrease in mean clustering coefficient, while increasing network size led to a decrease in this measure. This was consistent in the model testing Actual vs. Target data, as well. This opposite effect is likely due to age-related changes in network size over time: earlier on in the data set, network size increases rapidly, while over time this plateaus as new CDI words are acquired less frequently. Increases in age may thus not lead to any change in network size or distribution, while changes in network size likely always lead to changes in distribution.

The differences in findings between mean path length and clustering coefficient make sense in terms of systematicity in a growing lexicon. As new words are acquired, they likely connect with one of a number of dense clusters of phonologically-similar words. These clusters, particularly over time, may become more distinct from one another, and thus more dispersed in the network. A decrease in mean path length would indicate that all words were similar to one another, ~~which~~ would not be an efficient way to be understood by other speakers as an increasing number of words are acquired. Instead, having a number of systematic but distinct patterns allows infants to navigate the challenges of early word production, maintaining dense clusters of similar forms that, on average, are increasinly dispersed from one another in network space. This outcome aligns well with care study accounts of infants' early words, where we see the establishment of different production patterns, or templates (Vihman, 2019) over time; for example Waterson's (1971) case study of her son's production, where five distinct systematic structures are identified in his data

The second analysis tested whether or not word selection and adaptation could be observed in the data using GAMMs. Overall, this was not found to be the case, at least not in terms of the predicted changes in network connectivity. Instead, the argument for systematicity across development was strengthened, as Actual networks were found to have consistently higher connectivity (and thus more forms that are phonologically similar to one another) than Target networks over the full period of data collection. As observed in Figure 3, we do see that the difference in Actual vs. Target connectvity changes over time, suggesting that selection and adaptation are taking place in the dataset. It may be the case that the measures used here (specifically, mean k as a dependent variable) were not able to capture these changes, or that changes were too nuanced, and too variable between infants, to show meaningful results in the model.

While this analysis presents a more nuanced view of systematicity than in previous studies that draw entirely on network growth models, including Laing (under review), Fourtassi and colleagues (2020) and Siew and Vitevitch (2020), closer inspection of the network graphs themselves may have been useful in supporting and explaining the findings in more detail. Future work could combine computational analyses of networks with a more impressionistic analysis of infants' early word productions to bring together these two very different but equally valuable methodologies. Indeed, drawing on modelling to understand large-scale data makes it possible to quantify findings in a more rigorous way, but it abstracts away for the nature of early productions and means we miss out on the detail of the patterns that are being drawn upon. It also undervalues the extent to which infants take very individualised paths in phonological development (Vihman, 1993), and the possibility that broad-scale generalization is simply not wreflective of the reality of early production.

Overall, these findings support a case for systematicity in early development. The analysis of network graphs supports and builds on existing studies in this area - those that present a 'close-up' case-study analysis (e.g. Szreder, 2013; Waterson, 1971) and those that draw on computational methods to analyse large data sets in a generalised way (e.g. Fourtassi et al., 2020; Laing, under review) - to present a nuanced evaluation of a large-scale early production data set. Findings suggest that the developing network is characterised by an increasing number of dense clusters of similar-sounding words, and that systematicity is present in early production from the beginning. Future studies could combine computational findings with a close-up analysis of a sample of early words, in order to better understand how generalised findings are represented in real production data.

Amaral, L. A. N., Scala, A., Barthelemy, M., & Stanley, H. E. (2011). Classes of small-world networks. In *The structure and dynamics of networks* (pp. 207–210). Princeton, NJ: Princeton University Press.

Amatuni, A., & Bergelson, E. (2017). *Semantic Networks Generated from Early Linguistic Input*. https://doi.org/10.1101/157701

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. Retrieved from https://igraph.org

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of english. *Language and Speech*, *49*(2), 137–174.

Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of french determiners. *Journal of Child Language*, *35*(1). Retrieved from https://www-cambridge-org.abc.cardiff.ac.uk/core/journals/journal-of-child-language/article/prosodicallyconditioned-variability-in-childrens-production-of-french-determiners/B34D994A1EA63DFF841D1C2C7974D6FE

Deuchar, M., & Quay, S. (2000). *Bilingual acquisition: Theoretical implications of a case study*. Oxford, UK: Oxford University Press.

Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbons (Eds.), *Handbook of vowels and vowel disorders* (pp. 24–60). Psychology Press.

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847. https://doi.org/10.1111/cogs.12847

Golbeck, J. (2013). *Analyzing the Social Web*. San Francisco, UNITED STATES: Elsevier Science & Technology. Retrieved from http://ebookcentral.proquest.com/lib/york-ebooks/detail.action?docID=1152671

Hedlund, G., & Rose, Y. (2020). *Phon 3.1 [computer software]*. Retrieved from https://phon.ca

Kalinowski, J., Hansel, L., Vystrčilová, M., Ecker, A., & Mani, N. (in prep). *The development of early phonological networks: An analysis of individual longitudinal vocabulary growth.*

Kent, R. D., & Rountrey, C. (2020). What acoustic studies tell us about vowels in developing and disordered speech. *American Journal of Speech-Language Pathology*, *29*(3), 1749–1778. https://doi.org/10.1044/2020_AJSLP-19-00178

Laing, C. (under review). *Phonological Networks and Systematicity in Early Lexical Acquisition* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/z8pyg

Mak, M. H. C., & Twitchell, H. (2020). Evidence for preferential attachment: Words that are more well connected in semantic networks are better at acquiring new links in paired-associate learning. *Psychonomic Bulletin & Review*, *27*(5), 1059–1069. https://doi.org/10.3758/s13423-020-01773-0

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research*, *44*(3), 670–684. https://doi.org/10.1044/1092-4388(2001/054)

Monaghan, P., Christiansen, M. H., Farmer, T. A., & Fitneva, S. A. (2010). Measures of phonological typicality. *The Mental Lexicon*, *5*(3), 281–299. https://doi.org/10.1075/ml.5.3.02mon

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rij, J. van, Wieling, M., Baayen, R. H., & Rijn, H. van. (2022). *itsadug: Interpreting time series and autocorrelated data using GAMMs.*

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and software-assisted methods for the study of phonology and phonological development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford handbook of corpus phonology* (pp. 380–401). Oxford, UK: Oxford University Press.

Siew, C. S. Q., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000876

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. arXiv. Retrieved from http://arxiv.org/abs/1703.05339

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3

Szreder, M. (2013). The acquisition of consonant clusters in polish: A case study. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence* (pp. 343–361). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511980503.016

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, *21*, 61–82.

Vihman, M. M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition*, *19*(1), 69–88. https://doi.org/10.1017/S1366728914000790

Vihman, M. M. (2019). *Phonological templates in development*. Oxford, UK: Oxford University Press.

Vihman, M. M., & Croft, W. (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, *45*(4). https://doi.org/10.1515/LING.2007.021

Vihman, M. M., & Keren-Portnoy, T. (2013). The emergence of phonology: Whole-word approaches, cross-linguistic evidence. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology: Whole-word approaches and cross-linguistic evidence*.

Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511980503.002

Waterson, N. (1971). Child phonology : A prosodic view. *Journal of Linguistics*, *7*(2), 179–211. https://doi.org/10.1017/S0022226700002917

Watts, D. J., & Strogatz, S. H. (1998). *Collective dynamics of "small-world" networks. 393*, 3.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116. https://doi.org/10.1016/j.wocn.2018.03.002

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36.

Table 1

Outputs from nested model comparisons testing the effect of data type (Real vs. Simulated and Actual vs. Target on mean path length and clustering coefficient.

Model	Df	Chisq	p
Mean Path Length (Real vs. Simulated)	2	525.39	<0.001
Mean Path Length (Actual vs. Target)	1	359.58	<0.001
Clustering Coefficient (Real vs. Simulated)	2	1176.40	<0.001
Clustering Coefficient (Actual vs. Target)	1	124.47	<0.001

Table 2

Outputs from linear regression models testing comparisons of Real vs. Simulated data on mean path length and clustering coefficient.

	Mean path length				Clustering coefficient			
Effect	beta	SE	t	p	beta	SE	t	p
Intercept	1.350	0.12	11.225	<0.001	0.747	0.02	29.924	<0.001
Real vs. Erdos–Rényi	1.210	0.05	23.605	<0.001	-0.666	0.01	-60.302	<0.001
Real vs. Watts-Strogatz	-0.373	0.05	-7.454	<0.001	0.159	0.01	14.745	<0.001
Corpus	0.032	0.04	0.725	0.471	-0.014	0.01	-1.292	0.225
Age	0.007	0.00	1.343	0.213	0.004	0.00	2.746	0.009
Network size	NA	NA	NA	NA	-0.001	0.00	-11.128	<0.001

Table 3

Outputs from linear regression models testing comparisons of Actual vs. Target data on mean path length and clustering coefficient.

	Mean path length				Clustering coefficient			
Effect	beta	SE	t	p	beta	SE	t	p
Intercept	1.17	0.08	15.35	<0.001	0.88	0.03	30.08	<0.001
Actual vs. Target	0.30	0.01	27.07	<0.001	-0.10	0.01	-12.48	<0.001
Corpus	0.07	0.02	3.64	0.006	-0.04	0.01	-4.68	<0.001
Age	0.00	0.00	1.08	0.294	0.00	0.00	-1.61	0.121
Network size	0.00	0.00	17.23	<0.001	0.00	0.00	-11.91	<0.001

Table 4

Outputs from nested model comparisons of GAMMs testing the effect of data type (Actual vs. Target) on number of connections in the network. Model comparisons compared full models against those without parametric and smooth terms that included data type.

Model	Df	Difference	p
Mean connectivity (Actual vs. Target)	3.000	173.708	$< 2e\text{-}16$

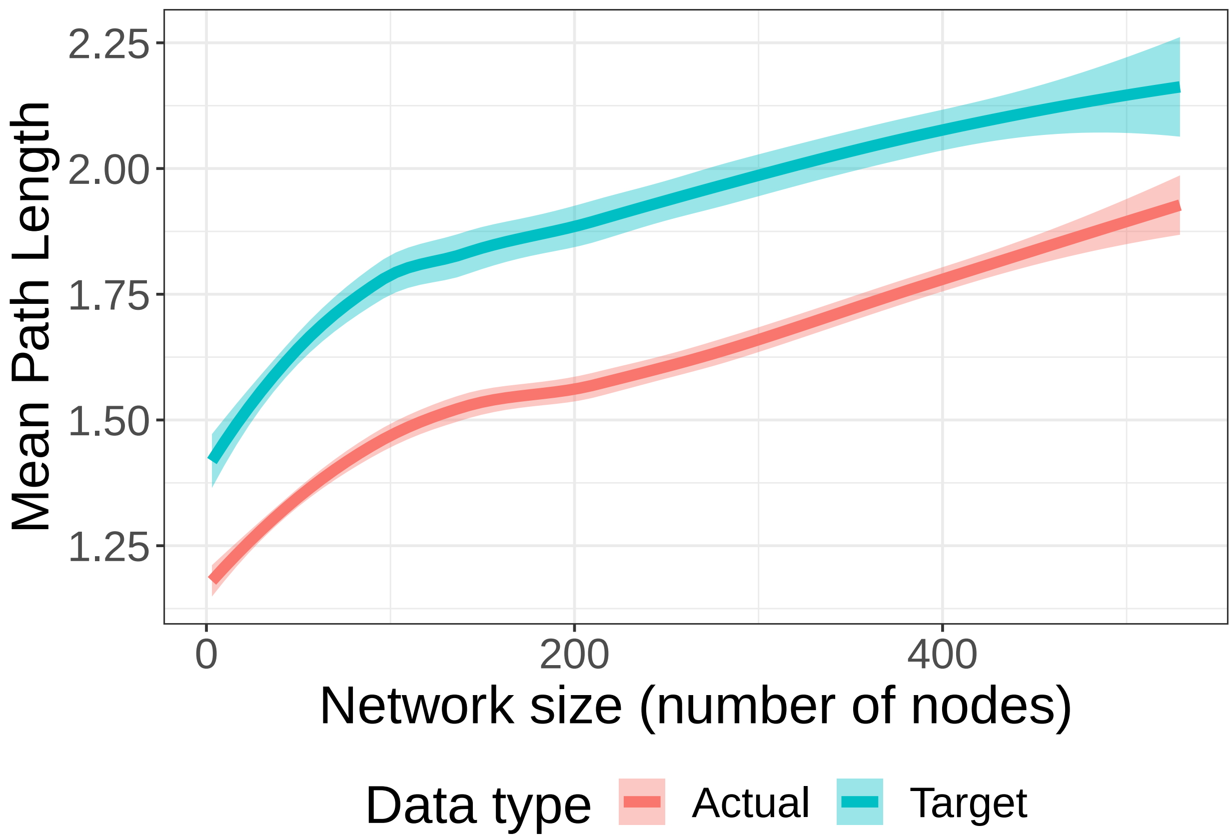

Figure 1. Change in mean path length as network size increases, in Actual vs. Target data. Coloured lines represent Data type; coloured bands represent 95% CIs.

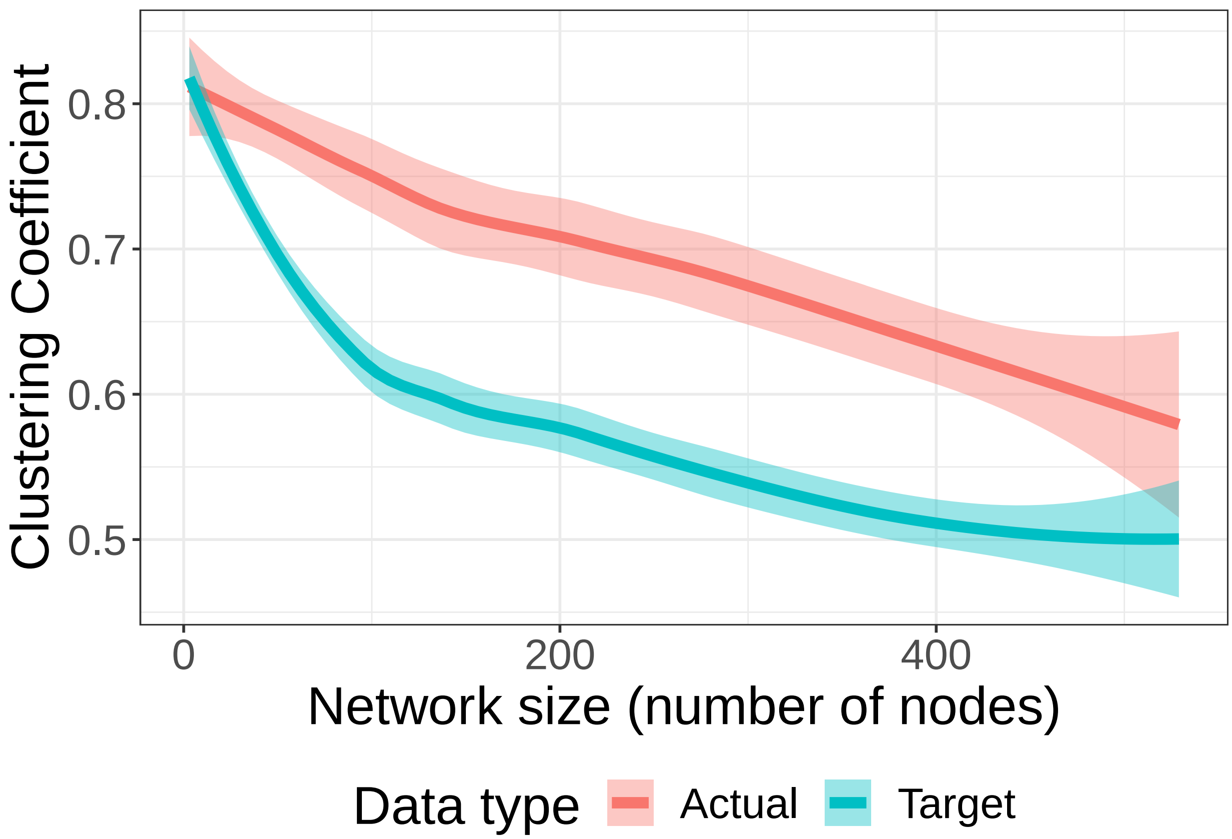

Figure 2. Change in mean clustering coefficient as network size increases, in Actual vs. Target data. Coloured lines represent Data type; coloured bands represent 95% CIs.

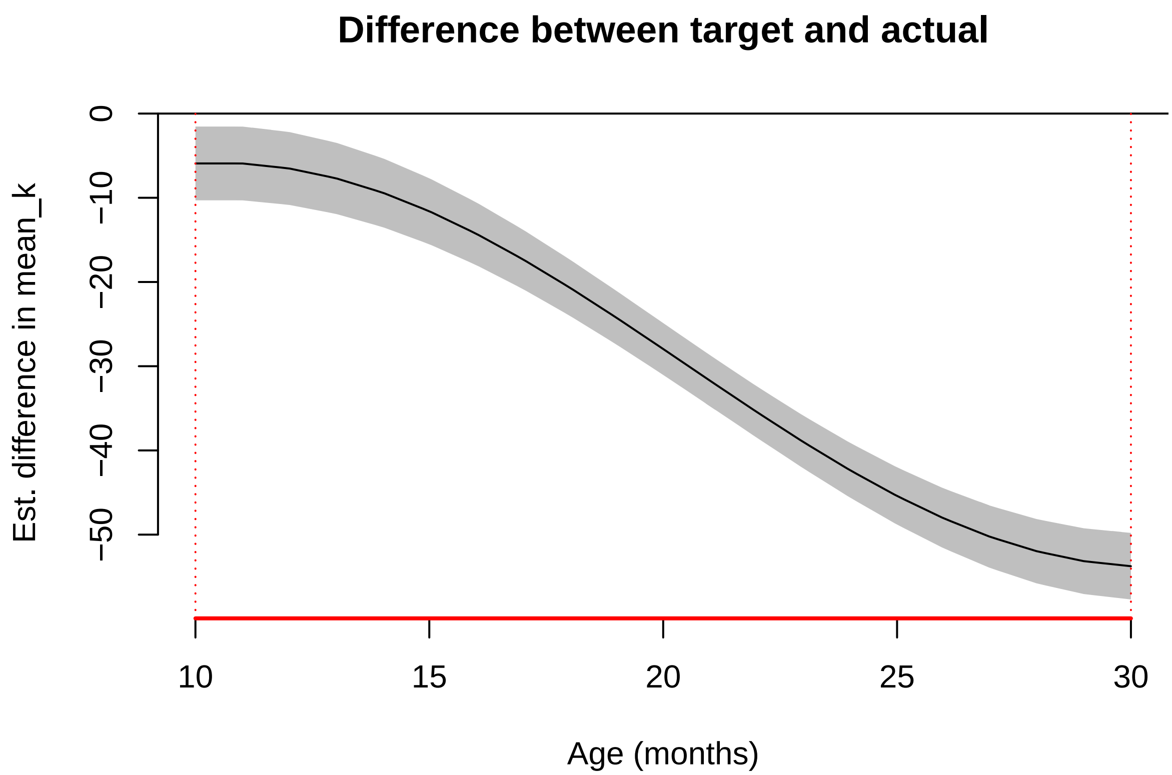

Figure 3. Difference smooth plot showing difference between connectivity (mean k) in Actual vs. Target forms from the GAMM model specified above. Shaded area shows 95% confidence intervals, red line along x-axis indicates months in which the difference between Actual and Target forms was significant.