Multiplex lexical networks reveal patterns in early word acquisition in children

Massimo Stella[1,+], **Nicole M. Beckage**[2], **and Markus Brede**[1]

[1]Institute for Complex System Simulations, University of Southampton, UK
[2]Department of Electrical Engineering and Computer Science, University of Kansas, USA
[+]Corresponding author: massimo.stella@inbox.com

ABSTRACT

Network models of language have provided a way of linking cognitive processes to language structure. However, current approaches focus only on one linguistic relationship at a time, missing the complex multi-relational nature of language. In this work, we overcome this limitation by modelling the mental lexicon of English-speaking toddlers as a multiplex lexical network, i.e. a multi-layered network where N=529 words/nodes are connected according to four relationships: (i) free association, (ii) feature sharing, (iii) co-occurrence, and (iv) phonological similarity. We investigate the topology of the resulting multiplex and then proceed to evaluate single layers and the full multiplex structure on their ability to predict empirically observed age of acquisition data of English speaking toddlers. We find that the multiplex topology is an important proxy of the cognitive processes of acquisition, capable of capturing emergent lexicon structure. In fact, we show that the multiplex structure is fundamentally more powerful than individual layers in predicting the ordering with which words are acquired. Furthermore, multiplex analysis allows for a quantification of distinct phases of lexical acquisition in early learners: while initially all the multiplex layers contribute to word learning, after about month 23 free associations take the lead in driving word acquisition.

Introduction

Language consists of a multi-level mapping of meanings onto words[1,2]. In order to communicate, humans must learn how to use linguistic structures to express thoughts as words. The cognitive processes behind language learning may organise the components of language into a so-called *mental lexicon* (ML)[2]. This lexicon can be described as a network structure of interacting lexical items (e.g. word representations). Empirical studies in psycholinguistics suggest that, rather than providing exact word definitions (as in common dictionaries), the ML stores word meanings as multi-relational or multiplex word patterns[3–8]. How the multiplex organisation of the ML relates to and influences language learning is still poorly understood but new techniques related to multiplex networks[9–12] allow us to explore patterns of word acquisition within the mental lexicon. We achieve this by constructing an edge-coloured multiplex network[9] based on relational features of phonology, semantics, and syntax. Going beyond the topological description of the multiplex network, we exploit the multiplex structure to predict normative acquisition of young children.

Previous literature on modelling language learning and use through network science has largely focused on single-layer representations of networks[6,7,13]. We build on these works, that strongly suggest that many cognitive constraints and mechanisms affecting the use of language can be explored through network science. In fact, experimental results have shown strong correlations between network structure and human performance in various language related tasks. Measurements of retrieval times[4,5,14], age of acquisition[15–17], creativity[18] and even semantic degradation due to ageing[19] have been studied and modeled using single layer networks. Further, networks of phonological word similarities have highlighted an upper bound on the size of phonological neighbourhoods as well as a tendency to avoid local clustering[20–22]. Experiments with adults have indicated that these constraints relate to word confusability in identification tasks[20,23].

While single layer networks reveal aspects of the structure of language and language-related cognitive processes, it is clear that this approach cannot offer a unified view of language that simultaneously accounts for phonological, semantic and syntactic aspects of language, as required by increasingly sophisticated experimental set ups[8]. One approach that is capable of overcoming this limitation is to use multiplex network representations as originally introduced in the social sciences[9]. Within this approach, we project the complexity of an individual's mental lexicon (ML) onto a multiplex network, which we call *Multiplex Lexical Network* (MLN). Our MLN is composed of multiple network layers, with nodes representing words and layers capturing different relationships between words. Multiplex networks[9] are a specific type of multi-layer networks[10,12] where nodes represent the same set of items on all layers. In our MLN we do not consider inter-layer connections. Instead we focus on intra-layer relationships that have previously been shown to influence children's lexical learning in single-layer network

studies: 1) free-associations[15,24], 2) shared features[16,25], 3) co-occurrence in child directed speech[17,26] and 4) phonological similarity[20,27]. We quantitatively show that by understanding the structure of multiple layers of the ML we can gain further insight into human cognition as related to language acquisition in young toddlers.

The main advantage of the multiplex approach is that it allows for a more detailed representation and quantification of many real-world systems[28]. This added model complexity provides additional insights into a system's structure and dynamics. In the last few years multiplex modelling has provided novel insights in areas as diverse as social balance in on-line platforms[29,30], emergence and stability of multiculturalism[31], congestion in transportation networks[28], and ecosystems in ecology[32,33], cf.[10,12] for a review on multi-layer and multiplex networks.

The multiplex network approach relates to other works in psycholinguistics investigating language via multi-layer networks. Liu et al.[34] analysed Chinese as a multi-layer network, composed of syntactic and phonemic layers. They found that almost half of the syntactic dependency relations are between phonologically similar words. A similar analysis was performed for English and Croatian establishing language specific relationships between syntax and phonology[35]. However, these studies consider only topological features of a given language representation. We extend this approach by explicitly considering the interaction of topological features with a cognitive dynamical process, i.e. for exploring language acquisition in young children.

Below, we consider the ability of structural network features to account for normative acquisition trajectories. We evaluate performance using empirical parent reports of toddlers' productive speech, aggregated to capture the order of word learning for children between 16 and 30 months[36]. Our aim is to assess the predictive power of a multiplex lexicon representation to capture emerging features and possible mechanisms of word acquisition in young children. As it currently stands, our model is solely descriptive: it only approximates importance of words from observational data rather than inferring it from a generative Bayesian framework[37]. However, we demonstrate that our multiplex model is more accurate than any single layer representation in predicting the normative acquisition trajectories of young children. We find that the multiplex network can account for developmental trends, offering predictions and interpretability that analysis based on single layer networks or word specific measures such as frequency or length cannot.

Results

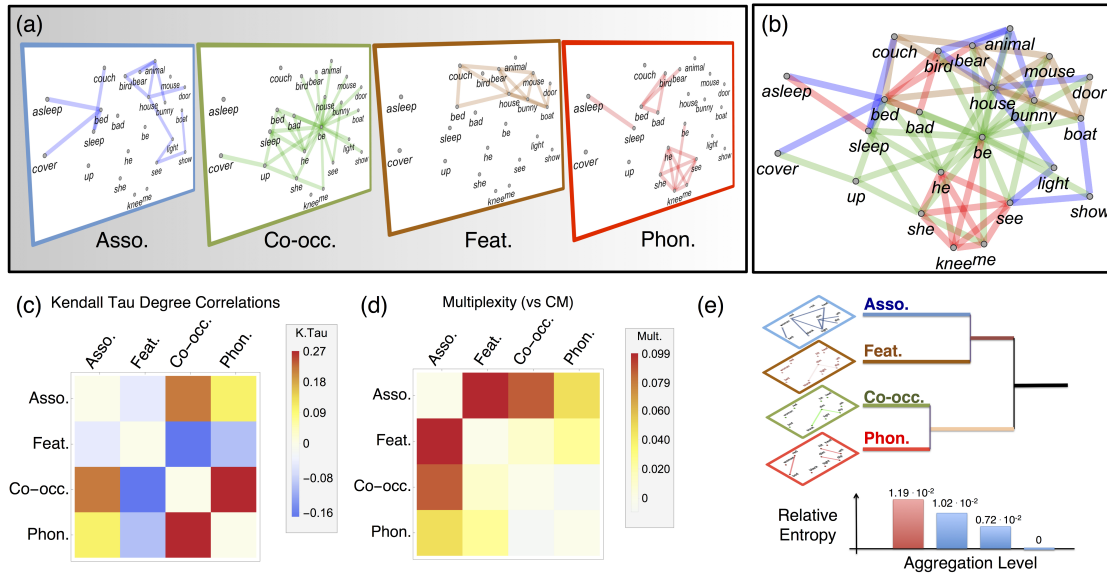## Multiplex network construction and analysis

Figure 1. (a) Visualisation of part of the multiplex network representing a toddler's lexicon. (b) Visualisation of an example of the MLN as an edge coloured network. Both (a) and (b) include only 24 of the 529 nodes/words in the whole MLN. (c) Degree correlations across layers quantified by the Kendall Tau. (d) Multiplexity or edge overlap among different layers relative to expectations from configuration models. (e) Reducibility dendrogram for the MLN. The MLN is irreducible: its layers are so different from each other that aggregating any of them would lead to a loss of information about topological patterns.

We first construct a multiplex lexical network composed of four layers capturing i) free associations from the South Florida association norms[24], (ii) feature sharing from the McRae et al. dataset[25], (iii) co-occurrence in child-directed speech from the CHILDES dataset[26], and (iv) phonological similarities from WordNet 3.0[20] (cf. Methods and SI Sect. 2). Hence the resulting

Empirical Network	$\langle k \rangle$	CC	a	$Conn.$	$\langle d \rangle$	$CC(CM)$	$a(CM)$	$Conn.$ (CM)	$\langle d \rangle$ (CM)
Associations (Asso.)	9.3	0.20	-0.1	99.6%	3.2	0.03(1)	$-0.01(6)$	99.6%	3.00(1)
Feature Norms (Feat.)	9.0	0.63	-0.01	24.2%	1.8	0.38(2)	$-0.07(6)$	24.2%	1.72(1)
Co-occurrences (Co-occ.)	8.1	0.69	-0.44	62.4%	2.2	0.75(3)	$-0.40(2)$	62.4%	2.19(1)
Phonological Sim (Phon.)	1.31	0.37	0.48	33.1%	7.7	0.02(1)	0.03(2)	45.5%(8)	5.31(7)
Multiplex Aggregate	26.5	0.33	-0.07	100%	2.4	0.18(5)	$-0.10(8)$	100%	2.25(1)

Table 1. Metrics for the MLN layers with $N = 529$ nodes, listing mean degree $\langle k \rangle$, mean clustering coefficient CC, assortativity coefficient a, percentage of nodes in the largest connected component $Conn.$, and mean shortest path length of the largest connected component $\langle d \rangle$. The quantities are measured for the layers of the MLN and for randomized equivalents with the same degree sequences (denoted with CM and grayed out). Error bars represent standard deviations and are reported in brackets behind the last significant digit.

MLN properties emerge from the assembly of independent measures rather than from a multiplex network design. Figures 1 (a) and (b) provide a visualisation of the links in each of the four MLN layers. Panel (a) treats each layer as separate whereas panel (b) describes the multiplex as an edge-coloured graph[9], in which different colours denote different relationships. As evident from (a) and (b), even if words might be disconnected on one layer, they could be connected on the MLN structure.

We report single-layer summary statistics and comparison to configuration models (i.e. random graphs preserving the degrees of an empirical network[38]) in Table 1. We focus on network features already analysed in previous works[15, 20, 39, 40]: degree, clustering, degree mixing, connectivity ratio and mean shortest path length. For instance, the degree of a node counts its connections[38] and on the phonological layer, it coincides with the size of its phonological neighbourhood[20, 23]. See Methods and SI Sect. S3 for precise definitions and interpretations.

Tab. 1 shows that the phonological layer has a smaller mean degree compared to other layers. Furthermore, apart from the association layer all other layers are highly fragmented. Their largest connected components include between 24.2% and 62.4% of the 529 words. These percentages quantify how many words could be reached from each other by navigating through links on individual layers. In contrast, the *multiplex network* is connected[41] – every pair of words is connected by a path that potentially exploits different relationship types. Compared to configuration models, except for the co-occurrence layer all other layers display higher clustering, non-zero assortativity and slightly larger path length in their largest connected component. In agreement with previous work[13, 40, 42, 43] these patterns suggest that the MLN layers display a core-periphery structure. From a cognitive perspective this feature might possibly facilitate navigation from word to word within cores (see SI Sect. S3 for further discussion). Layers also differ in their degree distributions, see SI Fig. S2: they are exponential-like for the synonym and phonological layers but much more heavy-tailed for the co-occurrence and association layers. Heavy tails indicate the presence of network hubs, i.e. words which are significantly more connected than average. These hub words play a central role for navigation through concept space[39, 40, 42] and could therefore be good candidates for earlier acquisition.

Is it appropriate to represent the ML in terms of four separate layers or can we get the same type of topological information with fewer layers? Structural reducibility analysis[44] can test if multiplex layers can be aggregated without losing information (see SI for details). Results reported in Fig. 1 (e) show that aggregation cannot be performed without information loss, demonstrating that the chosen 4-layer representation is in fact *irreducible*[44]. This is not to say that each layer plays a role in language acquisition, just that the information encoded in each layer is different from all other layers.

Before turning to modeling word acquisition, we investigate the MLN structure in more detail. Specifically, Fig. 1 (c) analyses the similarity between degrees of a word on different layers. We find that words in the feature and co-occurrence layers tend to have negative degree correlations (Kendall Tau $\kappa \approx -0.16, p < 0.0001$), indicating that hubs in one layer tend to have lower degrees in the other layers. This may suggest a type of semantic differentiation[15] in English in which words that have similar features tend to not be used in close proximity of each other, at least when speaking to young children. In contrast, the co-occurrence and phonological layers display positive degree correlations (Kendall Tau $\kappa \approx 0.27, p < 0.0001$): in the children's lexicon, words having many co-occurrences also have larger phonological neighbourhoods, further supporting the idea that phonological similarities and co-occurrences in child directed speech influence each other[26, 27].

Another question of interest is to what extent links overlap across layers. This can be quantified by edge overlap (or multiplexity[45]). Results are reported in Fig. 1 (d). As expected, we find some overlap between connections in the feature, co-occurrence and association layers; these layers capture semantic aspects of the ML. Beyond the overlap between these layers, links tend not to overlap significantly more than one would expect in configuration model across the other layers. This suggests that different MLN layers tend to capture different relational aspects of the ML, again emphasising that the MLN is a more complete model than single network representations. We now explore the importance of individual layers and of the MLN as a whole in explaining acquisition trajectories.

Multiplex Orderings: Results and Discussion

Here we explore the process of word learning through the MLN assuming a preferential acquisition scenario[16], where words are learned earlier if they are central in the language environment. We assume that relevance of the language environment can be summarised by connectivity in a language network[7,13,39,40] rather than making assumptions on the statistics of relationships to infer relevance, as is more standard in a Bayesian approach[37,46].

We generate a word acquisition ordering τ by ranking words according to measures computed on the MLN. We assume that if one ranking is predictive, it would indicate that the feature generating it is relevant to early language learning. We compare MLN-based orderings to those based on non-relational, word-specific information such as word frequency and word length. For consistency with the literature, we investigated features previously considered in linguistic networks such as word degree[15,17,43], closeness[39,40], betweenness[6,39,40] and PageRank[47]. For more details see Methods and SI Sect. S6.

We compare our orderings against an ensemble of empirical age of acquisition orderings, obtained from a probabilistic interpretation of the normative age of acquisition of words based on the MacArthur-Bates Communicative Development Inventory[26] (CDI) and over a population of roughly 1000 toddlers (see Methods). The comparison is performed by computing the average word overlap $O(\tau,t)$ of ordering τ with normative orderings, namely how many words τ correctly predicts as learned until the inventory includes t words. Word *gains* are then obtained by subtracting $O(\tau,t)$ from the expected number of correctly predicted words by random guessing (see Methods and SI Sect. S6). Word gains are normalised by the inventory size t and their Z-scores relative to statistics from random guesses are estimated. Positive *vocabulary normalised word gain* indicates better than random performance of ordering τ and negative values indicate performance inferior to random guessing.

Fig. 2 reports normalised word gains (a) and their Z-scores (b) for the best performing orderings from our analysis. No ordering always outperforms the others: we notice shifts in which word features are the most relevant for different periods of word learning. This agrees with the idea that the progression of acquisition changes over the course of development: children learn some initial words and then tend to generalise those words based on network connectivity later, i.e. preferential acquisition[16]. This cognitive interpretation of the observed shifts at different acquisition stages allow us to distinguish at least three distinct learning phases. In a first stage, comprising months 19 and 20, which we call the *very early learning stage* (VELS), we are trying to predict the first 40 learned words (gray overlays in Fig. 2). Direct sampling[48] for this period indicates that our results are below the 2.5% threshold for statistical significance – the network appears to contain little information relevant for predicting acquisition. This phase is followed by an *early learning stage* (ELS) covering an age range between 20 and 23 months. This phase is characterised by *closeness centrality outmatching* every other ordering (see also SI Sect. S6). Last we discriminate a *late learning stage* (LLS) comprising ages between 23 and 28 months. During this stage the dominant hub words have already been learned and more localized structure is likely affecting learning. LLS is characterised by degrees in the association layer and frequency performing equally to or slightly better than closeness and other word features (cf. SI Sect. S6).

Interestingly, while previous work has found frequency to be a good predictor for normative age of acquisition[49], results in Fig. 2 suggest that word length and some topological network information are more predictive of early acquisition ordering. Good performance is achieved by word length in ELS, possibly because of a *least effort effect*[1] that suggests that short words are easier to memorise and learn. This hypothesis is also supported by the observation that word length loses its predictive power at later learning stages (word gain Z-scores are compatible with random fluctuations after 300 words have been learned). We interpret this happens when more sophisticated learning strategies are being used. One of them might be mediated by degree orderings. Among the four layers, ordering words according to their degree in the association layer gives the best performance and it can be up to 100% more predictive than random guessing (cf. SI Fig. S7). This confirms previous results suggesting that association norms are generally good predictive models for early language learning[15,16,50]. We conjecture associations perform well in our case because they are strongly related with semantic memory[14,24,51].

Previous work has pointed out that in early learning, during VELS and ELS, it is very difficult to outperform random guessing because children may not have a clear strategy for learning new words early in language acquisition[50]. Instead here we find that better than random prediction is possible even in very early learning stages, at least for normative acquisition. Using the multiplex closeness of words it is possible to predict up to 160% words more than at random (cf. SI Fig. S7) and up to 25% more than by using degrees in the association layer (cf. SI Fig. S8). As reported in SI Sect. S6, we checked that multiplex closeness consistently outperforms every single-layer orderings during early learning stages. This is a promising result in that multiplex approaches may allow us to quantify learning strategies in toddlers. Also other multiplex measures such as versatile PageRank[52] perform up to 10% better than degrees in the association layer (cf. SI Fig. S8). The enhanced prediction ability of multiplex features indicates that the multiplex representation of the ML captures meaningful influences of multiple types of feature representations. Altogether, we can conclude that early learning is strongly related to the multiplex structure, specifically the closeness of words considering *all* layers that form the MLN.

Results for the LLS show a reversal of trends. Here we find that a single layer based on association degree ordering gives best predictability. Global and local multiplex measures, as well as word specific features such as word frequency, perform similarly or slightly worse until 400 words have been learned (see Fig 2 and SI Fig. S8). This indicates that preferential

acquisition is very important, but at this point in development the dominant relevant relationships are those captured by the association layer. This is compatible with the observed emergence of semantic learning in children[27]. However it has to be noticed that statistical significance deteriorates markedly during later stages of learning making comparisons of performance in the LLS difficult.
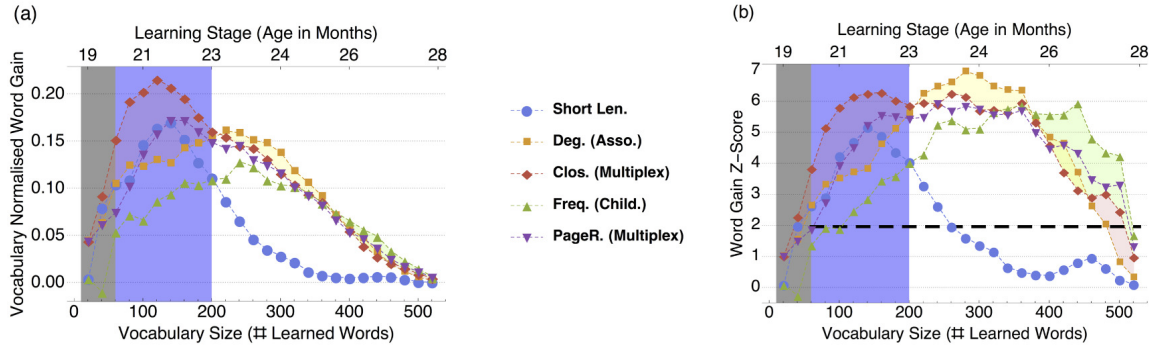

Figure 2. Left: Vocabulary normalised word gains for different orderings: shorter words first (Short Len.), words with higher degree in the association layer first (Deg. Asso.), closer words on the whole MLN first (Clos. Multiplex), more frequent words first (Freq. Child.) and higher multiplex Page Rank first (PageR. Multiplex). A normalised word gain of 0.20 means that 20% of words in the vocabulary have been predicted by the ordering strategy on top of expectations from random guessing. Higher word gains indicate better predictability. Right: Statistical significance indicated by word gain Z-score of the respective orderings. The distribution of the random overlaps can be approximated by a Gaussian after 60 words have been acquired (see SI Sect. S5). In this range, represented by the black dashed line, a score $Z \geq 1.96$ denotes word acquisition patterns that are different from random fluctuations with a 2.5% significance level. In both plots, error bars are the same size as the dots. Best performing orderings are highlighted with different overlays, which identify different learning stages: VELS (in gray), ELS (in light purple) and LLS (in white). Direct sampling was used in VELS for testing the statistical significance of results.

So far, we evaluated orderings based on single network layers and the combined multiplex, showing that prediction performance can benefit from taking account of information from multiple layers. In this analysis we assumed that each layer plays an equal role. In the next section we evaluate the changing influence of layers during the evolution of the lexicon.

Optimisation of layer influences

To explore the influence of different layers on word acquisition, we consider linear combinations of word features on each layer in order to obtain weighted multiplex measures (cf. Methods section). For a given network metric, we optimise the coefficients of the linear combinations to maximise vocabulary normalised word gains. The resulting layer weights, that maximise predictability, indicate the influence of the respective layers on word acquisition. We explored several types of optimisation, based on degree, closeness, betweenness and local clustering (cf. SI Sect. S7 and SI Tab. S3). We focus on the two best performing quantities from the ordering experiments, namely degree and closeness.

In order to avoid over-fitting, we perform a Monte Carlo robustness analysis: we optimise over subsets of word trajectories sampled uniformly at random, consisting of only 80% of the original words. Averages are computed over these trajectories and over different realisations of the age of acquisition trajectories from the normative age of acquisition ordering. We also confirmed that the improvement in performance of optimised multiplex parameters is strongly dependent on the structure and overlap of layers, as experiments on randomised multiplex networks result in far smaller word gains (see SI Sect. S8).

Optimisation results (cf. Fig. 3) indicated negligible contribution of the phonological layer, with weights ($\approx 10^{-3}$). Degree and closeness orderings based only on the phonological layer perform poorly as well, yielding word gains compatible with random fluctuations at a 5% significance level (see SI Sect. S6.1 and SI Fig. S5). We retrieved the same optimisation results for the other layers when the phonological layer was excluded. In addition, we considered phonological word scores coming from an "extended" phonological layer for adults, already analysed by Stella and Brede[21] and including 30000 words (see also SI Sect. S6.1.1). Assessing the influence of adults' phonology over word acquisition in toddlers matches previous approaches in the relevant literature[53] and we present results in our MLN framework in Fig. 3 but also in SI Sect. S7.

Optimal layer influences are displayed as a function of age for the MLN layers in Fig. 3 (c) and (d). Similar patterns of layer influence are found for rankings based on both degree and closeness. We see a clear distinction between different learning stages. When degree is considered, cf. Fig. 3 (c), in VELS the main layer contributions stem from the association and the feature norm layers. Instead, when optimising closeness, cf. Fig. 3 (d), in VELS all layers are found to contribute equally. In
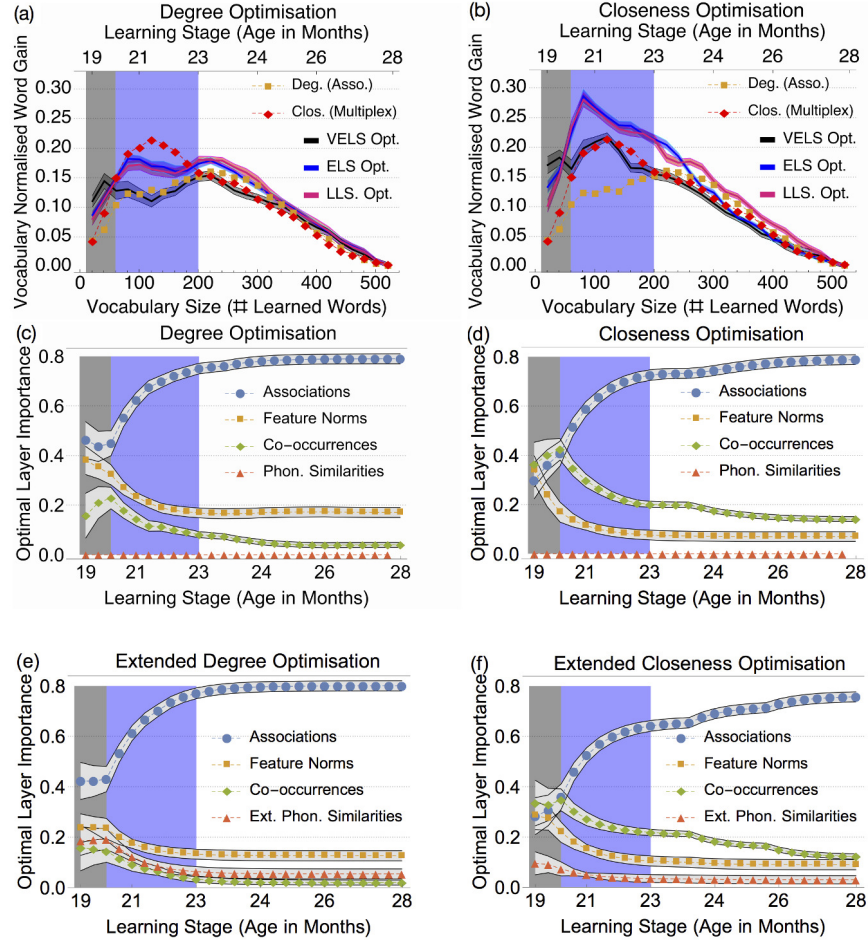
Figure 3. Top (a-d): Optimisation results for the original MLN. (a,b): Vocabulary normalised word gains corresponding to the optimal layer influences at the end of VELS (black), middle of ELS (blue), and middle of LLS (pink) for degree optimisation (a) and closeness optimisation (b). Error margins represent standard deviations over randomised ensembles. In (b) no optimisation of degree outperforms multiplex closeness during ELS. (c,d): Average optimal layer weights indicating the influence or importance of layers over different learning stages obtained from Monte Carlo experiments with degree (c) and closeness (d) optimisation. The phonological layer resulted in influences around 10^{-3}. The same optimisation trends were obtained when the phonological layer was excluded from the optimisation. Bottom (e,f): For comparison, optimal layer weights for the MLN where rather than the phonological layer with 529 words an extended one with 30,000 words is used, instead. Considering the extended phonological layer leads to word gains at least 6% better than (a) and (b) (cf. SI Table 3).

both cases the ELS, the next learning stage, marks a transition region between the VELS (before) and the LLS (after) in which only the association layer dominates. This transition is what motivated our choice of distinguishing between VELS and ELS. For degree-based optimisations, contributions from the co-occurrence layer become very small in the LLS, while associations contribute 80% of weight. An analogous transition is observed in the ELS phase when closeness is optimised and after that associations make up 80% of the total layer influence. The dominant influence of associations in later stages of learning is indicative of the emergence of preferential acquisition[16] that we already observed in ordering experiments.

When extended phonology is considered, in Fig. 3 (e) and (f), the influences of semantic layers remain consistent, however, phonology has a higher impact when phonological neighbourhood size/degree is considered rather than closeness. In VELS the extended phonology has importances similar to co-occurrences but the influence of both approaches 0 later in development.

In Fig. 3 (a) and (b) we compare vocabulary normalised word gains of trajectories based on optimal layer influences against our ordering experiments (cf. SI Fig. S9 for the respective Z-scores and previous section). On the one hand, when predicting only words learned in VELS, the optimal influence orderings outmatch both degree in the association layer and multiplex closeness during VELS but perform worse at later stages. Direct sampling indicates that these results are significant with a

p-value of 0.01. Optimal influences up to the middle of ELS and LLS provide similar word gains, which are both up to 25% higher than the optimal word gains from VELS. This suggests that the *same learning strategies* may be adopted *after* VELS, starting in ELS and lasting throughout LLS. On the other hand, results in Fig. 2 (a) highlight the importance of closeness centrality for predicting word acquisition at early stages. In fact, in ELS even the optimal linear combination of degrees results in *smaller word gains* compared to the unweighted multiplex closeness centrality. Optimal linear combinations of closeness perform even better than the unweighted multiplex closeness, see Fig. 2 (b). Notice that closeness is a *global* network feature, accounting for the position of a node relative to all other nodes. This is distinct from *local* features, such as degree, which only measure a node's relationship to its first neighbours. Our findings give support to the hypothesis that, particularly in the early learning stages, word learning is influenced by the global structure of the multiplex lexical network. This global multiplex structure is able to capture some important word patterns that influence early lexicon development.

Conclusions and future work

Introducing the framework of *multiplex lexical networks* (MLN), we have described and analysed normative word acquisition patterns of children between the ages of 19 and 28 months. Our approach considers semantic, phonological and syntactic contexts and captures more linguistic information than separate analysis of any single layer network previously could[14, 16, 17, 50]. We go beyond a purely topological description by exploiting network properties to predict the order of normative word learning in children. Interestingly, we find that the best topological feature in predicting word learning changes through the course of development. This allows us to use network information to distinguish three stages of learning: (i) a very early learning stage where all but the phonological layer contribute substantially to prediction, (ii) an early learning stage which marks a transition period, and (iii) a late learning stage in which contribution from word associations dominates word learning. We believe the last point is due to free associations serving as a much better proxy for detecting superordinate words compared to other network features (cf. SI Sect. 6.3 for a quantitative discussion). Further assessment of the extent to which the taxonomic organization serves as a mediator variable influencing word prediction is an interesting direction for future work.

Comparing the predictive power of various MLN features over time confirms: (i) the superiority of some multiplex network characteristics relative to single layer networks and word specific measures such as frequency or word length, and (ii) the special role closeness centrality might play in early word learning. This is emphasised in two ways: (i) strong performance of multiplex closeness, outperforming all combinations of local network features in the early learning phase, and (ii) the best overall predictions of the order words are learned is related to a weighted multiplex version of single-layer closeness centralities. We thus find strong indications that, particularly at early stages of learning, word acquisition in children is driven by minimising relational distances of learned words relative to other words in both the semantic and syntactic space of the mental lexicon.

With the MLN, we explored questions about normative lexical acquisition, uncovering developmental learning stages and quantifying the influence of certain types of linguistic information on early acquisition. Nevertheless, our MLN is limited in important ways. It is important to bear in mind that the MLN representation is only a projection of an individual child's full ML. Additionally, the results presented in this paper consider normative lexical acquisition. That is to say, a specific child may not learn words in the same order as the normative orderings. Instead normative order is obtained by averaging over roughly 1000 productive vocabulary reports. We attempt to address this shortcoming by sampling multiple orders probabilistically from these norms. Extending the model and testing the universality of the learning strategies in VELS and ELS on longitudinal data for individual children would be an interesting future research direction[2].

Another limitation is the use of a relatively small vocabulary. While the CDI is a commonly used check list of words a child produces, there are many words that a child may know which are not on the CDI. By expanding our model to the full vocabulary of a child, and for longer periods of development, we can further increase our understanding of word learning.

While the present study focuses on word learning in English, results should be corroborated by considering word learning in different languages. Future work can extend our analysis by considering correlations in word learning across other cognitive and linguistic domains to build an even richer picture of language organisation and learning than presented here. A stream of work in psycholinguistics has explored detailed mechanisms of word learning[37]. Such work typically explores how words are memorised when clues are presented in different order and focuses on the underlying mechanisms of inferring meaning-object mappings, pointing out that learning is strongly dependent on context. Integrating context-dependent information in our MLN framework would surely be an exciting future research direction for investigating individual children's learning trajectories.

The marginal influence of the original phonological layer found in this study can be reconciled with previous findings. Dautriche *et al.*[8] showed that 18 months old toddlers failed to learn an object label when it was a phonological neighbour of a noun they already knew. In our dataset 80% of words learned by toddlers within the first two months are nouns. This noun-richness is in opposition with orderings extracted from phonological degree and closeness, which identify mainly non-noun words, verbs in particular, as learned initially. This trend is not surprising, as having learned a verb earlier makes it easier for toddlers to learn its phonological neighbours[8]. However, this contrast in noun-richness at early development stages reflects in the poor performance of the original phonological layer in predicting normative acquisition ordering. Nonetheless, we show

that phonology influences word acquisition by considering a phonological network extended to the adults' lexicon, like in[53]. This extended layer identifies nouns as early candidates for learning and provides significant prediction results. This interplay between phonology and acquisition is in agreement with previous findings in the relevant literature[23, 43, 53, 54].

We have shown strong evidence that multiplex lexical networks capture a richer picture of the mental lexicon than previous works using only single-layer networks, in spite of the above limitations. More importantly, this work provides novel formalised methods for exploring and explaining the influence of linguistic features on early acquisition.

Methods

Age of acquisition dataset

Age of acquisition orderings are constructed based on the MacArthur-Bates Communicative Development Inventory (CDI) norms[36]. The CDI, based on parent report of the productive vocabulary of children aged between 16 and 30 months, has been shown to be related to future language ability in young children[36, 55]. The norms are averaged over more than 1000 vocabulary reports and indicate the percentage of children, at a given age, that reportedly produce (and understand) a specific word. From these norms, we sample orderings of word learning. For example, "mommy" and "ball" are reported as produced by 93% and 64% of children by 16 months, respectively; these words are (usually) learned earlier than "chair" which is produced by only 14% of 16 month olds. We assume a word is known once 50% of children in a given month produce that word. We then sample an order of words learned within a month assuming that higher rates of production probabilistically indicate earlier learning. Starting from position one of the ranking, we: (i) sample a learned word with probability proportional to the percentage of children which are able to produce it, conditioned on the overall production rate being over 50%, (ii) erase it from the list of candidate words to be learned at later positions and (iii) proceed one position down in the ranking and start from (i). The sampling starts from month 16 and it stops when every word has been sampled. Note that a non-probabilistic version of this ordering has been used in other developmental modelling contexts[16, 50].

Construction of the multiplex network

The four layers of the MLN were selected because of their use in previous single-layer network studies investigating word learning of young children[6, 13, 16, 17]. The four layers we consider here are (i) a semantic layer based on the Florida Free Association Norms[24] where a link exists from word A to word B if word B is a free associate of the cue word A; (ii) a layer capturing feature similarity based on the McRae Feature Norms[25] where words A and B are connected if they share at least one semantic feature; (iii) a layer based on word co-occurrences (measured in child-directed speech[26]) where words A and B are connected if they co-occur more than 45 times, where the threshold 45 was chosen approximately to match the link density of the other semantic layers; and a (iv) layer capturing phonological similarities (based on IPA transcription from WordNet 3.0[56]) where words A and B are connected if they have IPA transcriptions with edit distance one. We treat the association layer as undirected and edge weights in layers (i) to (iii) are ignored. The resulting MLN is composed of 529 words which represent the intersection between the CDI data and the set of words which have at least one connection on any layer (see SI).

Cognitive interpretation of network metrics

We investigate the same local and global features of words/nodes that were investigated in previous works about network representations of language: degree, closeness, betweenness and PageRank (see SI). In a given network, the degree of a node is the number of its links[38]. In our MLN, a word is characterised by one degree measure per layer. For instance, on the association layer a word with degree $k^{(Asso.)}$ has $k^{(Asso.)}$ associates. On the phonological layer the degree of a word coincides with its phonological neighbourhood size[43]: a word with degree $k^{(Phon.)}$ has $k^{(Phon.)}$ similar sounding words. Higher degree words in semantic networks were found to have lower age of acquisition norms[15], while degree on the phonological layer correlated positively with word confusability[23]. On the whole MLN, we can associate each word to its multidegree, namely the sum of its degrees on all the layers[11]. Fundamentally, the usefulness of measures such as degree would suggest that words with multiple associates, multiple phonological neighbors and multiple shared features are those words that are learned earliest. The mechanisms behind this phenomenon could be that words with many associates and shared features are those words that are important in the language environment because they play a central role or because they are important to the parent or child. We remain agnostic as to the exact reason behind the increased 'importance' of a word, but suggest that it can be summarized by the degree. Closeness relates to how fast information spreads from a node to others[38]. It is interesting for cognitive science since previous work showed that words closer on semantic topologies tend to be processed together in shorter time[4, 5, 39]. Betweenness captures the extent to which a node falls on the shortest path between pairs of words[38]. It captures centrality of words through a hypothetical navigation on the ML[6, 39, 40] (cf. SI Sect. S6). PageRank identifies the likelihood of reaching a given node by a random walk in a network. High PageRank words on semantic networks were found easier to retrieve in fluency tasks[47]. Closeness, betweenness and PageRank on multiplex networks exploit jumps through layers (cf. SI Sect. S6).

Overlap measures and word ranking

A word trajectory is an ordered list $\tau = (w_1, w_2, ..., w_N)$ indicating the exact order in which words are learned. We define τ_{aoa} to be the ensemble of rankings derived from the CDI norms. We sample each specific ordering τ_{aoa} based on the probability of production as reported in the age of acquisition norms.

Predicted word orderings are generated as follows. First, each word is ranked according to a word score, s_i, which is derived from network specific or extrinsic word features (e.g. frequency or word length). Words are then ordered according to word score, starting with the largest score. Words receive a position in the predicted trajectory according to their position in the ordering of word scores. Ties in word orderings s_i are taken into account by averaging over all resolutions. For example, one might wish to consider orderings based on degree in the association layer and would obtain the following word scores $s_{food} = 62$, $s_{water} = 45$, $s_{eat} = 20$. The corresponding predicted acquisition trajectory would be $\tau = (food, water, eat)$. Apart from the features mention in the main text of the paper, we also considered many others (cf. SI Sect. 6).

To evaluate predictive performance of a word score, we first measure the overlap $O(\tau, t)$ between the predicted learning trajectory τ and the empirically known trajectory τ_{aoa} at time t by counting the number of words that co-occur in τ and τ_{aoa} up to time t and averaging over different τ_{aoa}. We define the word gain $g(\tau, t)$ as the overlap minus random fluctuations, i.e.

$$g(\tau, t) = O(\tau, t) - \langle O(\tau_{ran}, t) \rangle. \tag{1}$$

The vocabulary size normalised word gains are computed as $G(\tau, t) = g(\tau, t)/t$ while the word gain Z-Scores consider deviations of overlaps from random guessing in terms of standard deviations σ, as $Z = g(\tau, t)/\sigma(O(\tau_{ran}, t))$.

Calculating optimal combinations of layers

To allow for varying influences of layers within the multiplex we construct word scores as convex combinations of individual layer influence, i.e. a word score s_w for word w is obtained as:

$$s_w = \alpha s_w^{(Asso)} + \beta s_w^{(Feat)} + \gamma s_w^{(Co-occ)} + (1 - \alpha - \beta - \gamma)s_w^{(Phon)}, \tag{2}$$

where $s_w^{(Asso)}$, $s_w^{(Feat)}$, $s_w^{(Co-occ)}$, and $s_w^{(Phon)}$ are word scores obtained from the respective single layer metrics and the coefficients α, β, γ give the influence of each layer on the overall word score s_w. This linear combination of single-layer features is similar in spirit to previous decompositions of multiplex centrality, see[57] and discussions in[44]. The optimisation finds the influences α, β, γ that maximise vocabulary normalised word gain (either over the entire time period or over specific learning phases) after removing 20% of the words at random. 20% was chosen in order not to remove too many words particularly in the very early word learning stages. Performance is evaluated on all words. Optimisation was performed using a differential evolution method[58]. Averages are calculated over 50 configurations of left out words for 30 normative age of acquisition orderings.

References

1. i Cancho, R. F. & Solé, R. V. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences* **100**, 788–791 (2003).

2. Aitchison, J. *Words in the mind: An introduction to the mental lexicon* (John Wiley & Sons, 2012).

3. Quillian, M. R. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral science* **12**, 410–430 (1967).

4. Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior* **8**, 240–247 (1969).

5. Collins, A. M. & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological review* **82**, 407 (1975).

6. Borge-Holthoefer, J. & Arenas, A. Semantic networks: structure and dynamics. *Entropy* **12**, 1264–1302 (2010).

7. Baronchelli, A., Ferrer-i Cancho, R., Pastor-Satorras, R., Chater, N. & Christiansen, M. H. Networks in cognitive science. *Trends in cognitive sciences* **17**, 348–360 (2013).

8. Dautriche, I., Swingley, D. & Christophe, A. Learning novel phonological neighbors: Syntactic category matters. *Cognition* **143**, 77–86 (2015).

9. Wasserman, S. & Faust, K. *Social network analysis: Methods and applications*, vol. 8 (Cambridge university press, 1994).

10. Kivelä, M. *et al.* Multilayer networks. *Journal of Complex Networks* **2**, 203–271 (2014).

11. De Domenico, M. *et al.* Mathematical formulation of multilayer networks. *Physical Review X* **3**, 041022 (2013).

12. Boccaletti, S. *et al.* The structure and dynamics of multilayer networks. *Physics Reports* **544**, 1–122 (2014).

13. Beckage, N. M. & Colunga, E. Language networks as models of cognition: Understanding cognition through language. In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, 3–30 (Springer, 2015).

14. De Deyne, S. & Storms, G. Word associations: Network and semantic properties. *Behavior Research Methods* **40**, 213–231 (2008).

15. Steyvers, M. & Tenenbaum, J. B. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science* **29**, 41–78 (2005).

16. Hills, T. T., Maouene, M., Maouene, J., Sheya, A. & Smith, L. Longitudinal analysis of early semantic networks preferential attachment or preferential acquisition? *Psychological Science* **20**, 729–739 (2009).

17. Beckage, N., Smith, L. & Hills, T. Small worlds and semantic network growth in typical and late talkers. *PloS one* **6**, e19348 (2011).

18. Kenett, Y. N., Anaki, D. & Faust, M. Investigating the structure of semantic networks in low and high creative persons. *The Cognitive, Emotional and Neural Correlates of Creativity* 6 (2015).

19. Goñi, J. *et al.* The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory. *Cognitive processing* **12**, 183–196 (2011).

20. Vitevitch, M. S. What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research* **51**, 408–422 (2008).

21. Stella, M. & Brede, M. Patterns in the English language: Phonological networks, percolation and assembly models. *Journal of Statistical Mechanics* **2015**, P05006 (2015).

22. Stella, M. & Brede, M. Investigating the phonetic organisation of the english language via phonological networks, percolation and Markov models. In *Proceedings of ECCS 2014*, 219–229 (Springer International Publishing, 2016).

23. Vitevitch, M. S., Chan, K. Y. & Roodenrys, S. Complex network structure influences processing in long-term and short-term memory. *Journal of memory and language* **67**, 30–44 (2012).

24. Nelson, D. L., McEvoy, C. L. & Schreiber, T. A. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers* **36**, 402–407 (2004).

25. McRae, K., Cree, G. S., Seidenberg, M. S. & McNorgan, C. Semantic feature production norms for a large set of living and nonliving things. *Behavior research methods* **37**, 547–559 (2005).

26. MacWhinney, B. *The CHILDES project: The database*, vol. 2 (Psychology Press, 2000).

27. Wiethan, F. M., Nóro, L. A. & Mota, H. B. Early lexical and phonological acquisition and its relationships. In *CoDAS*, vol. 26, 260–264 (SciELO Brasil, 2014).

28. De Domenico, M., Granell, C., Porter, M. A. & Arenas, A. The physics of spreading processed in multilayer networks. *Nature Physics* **12**, 901–906 (2016).

29. Szell, M., Lambiotte, R. & Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences* **107**, 13636–13641 (2010).

30. Gómez-Gardenes, J., Reinares, I., Arenas, A. & Floría, L. M. Evolution of cooperation in multiplex networks. *Scientific reports* **2** (2012).

31. Battiston, F., Cairoli, A., Nicosia, V., Baule, A. & Latora, V. Interplay between consensus and coherence in a model of interacting opinions. *Physica D: Nonlinear Phenomena* (2015).

32. Pilosof, S., Porter, M. A. & Kéfi, S. Ecological multilayer networks: A new frontier for network ecology. *arXiv preprint arXiv:1511.04453* (2015).

33. Stella, M., Andreazzi, C. S., Selakovic, S., Goudarzi, A. & Antonioni, A. Parasite spreading in spatial ecological multiplex networks. *arXiv preprint arXiv:1602.06785* (2016).

34. Liu, H. & Cong, J. Empirical characterization of modern chinese as a multi-level system from the complex network approach. *Journal of Chinese Linguistics* **42**, 1–38 (2014).

35. Martinčić-Ipšić, S., Margan, D. & Meštrović, A. Multilayer network of language: A unified framework for structural analysis of linguistic subsystems. *Physica A: Statistical Mechanics and its Applications* **457**, 117–128 (2016).

36. Dale, P. S. & Fenson, L. Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers* **28**, 125–127 (1996).

37. Xu, F. & Tenenbaum, J. B. Word learning as bayesian inference. *Psychological review* **114**, 245 (2007).

38. Newman, M. *Networks: an introduction* (Oxford University Press, 2010).

39. Motter, A. E., de Moura, A. P., Lai, Y.-C. & Dasgupta, P. Topology of the conceptual network of language. *Physical Review E* **65**, 065102 (2002).

40. Sigman, M. & Cecchi, G. A. Global organization of the WordNet lexicon. *Proceedings of the National Academy of Sciences* **99**, 1742–1747 (2002).

41. De Domenico, M., Solé-Ribalta, A., Gómez, S. & Arenas, A. Navigability of interconnected networks under random failures. *Proceedings of the National Academy of Sciences* **111**, 8351–8356 (2014).

42. Gravino, P., Servedio, V. D., Barrat, A. & Loreto, V. Complex structures and semantics in free word association. *Advances in Complex Systems* **15**, 1250054 (2012).

43. Vitevitch, M. S., Chan, K. Y. & Goldstein, R. Insights into failed lexical retrieval from network science. *Cognitive psychology* **68**, 1–32 (2014).

44. De Domenico, M., Nicosia, V., Arenas, A. & Latora, V. Structural reducibility of multilayer networks. *Nature communications* **6** (2015).

45. Gemmetto, V. & Garlaschelli, D. Multiplexity versus correlation: the role of local constraints in real multiplexes. *Scientific reports* **5** (2015).

46. Frank, M. C., Goodman, N. D. & Tenenbaum, J. B. Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science* **20**, 578–585 (2009).

47. Griffiths, T. L., Steyvers, M. & Firl, A. Google and the mind predicting fluency with pagerank. *Psychological Science* **18**, 1069–1076 (2007).

48. Grimmett, G. & Stirzaker, D. Probability and random processes (2001).

49. Kuperman, V., Stadthagen-Gonzalez, H. & Brysbaert, M. Age-of-acquisition ratings for 30,000 english words. *Behavior Research Methods* **44**, 978–990 (2012).

50. Beckage, N. M., Aguilar, A. & Colunga, E. Modeling lexical acquisition through networks. *Proc. of the 37th Conf. of the Cog. Sci. Society* (2015).

51. Zock, M. Words in books, computers and the human mind. *Journal of Cognitive Science* **16**, 355–378 (2015).

52. De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S. & Arenas, A. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature communications* **6** (2015).

53. Carlson, M. T., Sonderegger, M. & Bane, M. How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of memory and language* **75**, 159–180 (2014).

54. Stamer, M. K. & Vitevitch, M. S. Phonological similarity influences word learning in adults learning spanish as a foreign language. *Bilingualism: Language and Cognition* **15**, 490–502 (2012).

55. Fenson, L. *et al.* Variability in early communicative development. *Monographs of the society for research in child development* **59**, 1–185 (1994).

56. Miller, G. A. WordNet: a lexical database for english. *Communications of the ACM* **38**, 39–41 (1995).

57. Solá, L. *et al.* Eigenvector centrality of nodes in multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **23**, 033131 (2013).

58. Price, K., Storn, R. M. & Lampinen, J. A. *Differential evolution: a practical approach to global optimization* (Springer Science & Business Media, 2006).

59. De Domenico, M., Porter, M. A. & Arenas, A. Muxviz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks* **3**, 159–176 (2014).

Author contributions statement

M.S., N.B. and M.B. designed the original study, M.S., N.B. and M.B. conceived the experiments, M.S. conducted the experiments, M.S., N.B. and M.B. analysed the results. All authors reviewed the manuscript.

Additional information