Cross-linguistically Small World Networks are Ubiquitous in Child-directed Speech

Steven Moran, Danica Pajović, Sabine Stoll
Department of Comparative Linguistics, University of Zurich
Plattenstrasse 54, CH-8032 Zurich, Switzerland
{steven.moran, danica.pajovic, sabine.stoll}@uzh.ch

Abstract

In this paper we use network theory to model graphs of child-directed speech from caregivers of children from nine typologically and morphologically diverse languages. With the resulting lexical adjacency graphs, we calculate the network statistics $\{N, E, <k>, L, C\}$ and compare them against the standard baseline of the same parameters from randomly generated networks of the same size. We show that typologically and morphologically diverse languages all share small world properties in their child-directed speech. Our results add to the repertoire of universal distributional patterns found in the input to children cross-linguistically. We discuss briefly some implications for language acquisition research.

Keywords: network theory, linguistics, corpus linguistics, child language acquisition

1. Overview

Despite the remarkable diversity of linguistic structures in the world's 7000 or so languages, children can acquire any language. This fact presents many questions, including importantly: what are the underlying cognitive mechanisms that enable children to acquire language? And are there universal patterns in the linguistic input to children that potentially bootstrap these mechanisms?

Consider one salient difference among the world's languages (especially the under-studied ones): how words are constructed and how they relate to syntax. When analyzed in detail, it is rather difficult to define what a word is cross-linguistically (Hall et al., 2008). In some languages words represent what English speakers consider full phrases; in other languages the word and morpheme (smallest function bearing linguistic unit) are synonymous. Contrast two utterances from Indonesian (Gil and Tadmor, 2007) and Cree (Brittain, 2015):

(1) O, Ei lagi minum susu.
oh Ei more drink milk
'Oh, Ei is drinking more milk.' (Indonesian)

(2) Chi-wâp-iht-â-n â kâ-pushch-ishk-iw-â-t.
2-light-by.head-TR.INAN.NON3-2SG>0 Q PVB.CONJ-put.on-by.foot-STEM-TR.ANIM-3SG>4SG
'You see? She was putting it on.' (Cree)

Indonesian is an example of a language with a fairly low degree of synthesis, whereas Cree belongs to one of the most genuinely polysynthetic language families of the world (and features both noun incorporation and polypartite stems).[1] Clearly the frequency in which children hear a particular form is a function of synthesis combinatorics (Stoll et al., 2017). That is, in languages where morphology is in a closer one-to-one relationship between word and grammatical function, these forms will occur more frequently in the input. There will be greater transition probabilities in languages with more tokens than in morphologically-rich languages which have more types. Nevertheless, regardless of morphology, children from all languages learn to identify words and to produce them.

For a long time, Universal Grammar (UG) was the answer to such problems in language acquisition. In UG, language is the product of innate functions (Chomsky, 1957), where rules and parameters are hard-wired and the acquisition process involves language-specific tuning of linguistic structures (Chomsky, 2000). Because the language acquisition device is posited as innate, models of UG are not necessarily data-driven, but instead theoretical and mainly focused on 'Language' as an abstract system – centered historically on the syntactic structure of English and a few other major languages.

No matter what theoretical approach researchers adopt, they must explain how children identify patterns in their linguistic input and make use of these in productive generalizations as observed in their linguistic output. Usage-based or constructivist approaches are functionalist in that they take into account the way that language is used and contexts in which linguistic elements appear. Increased access to richly annotated linguistic data and computing power, coupled with approaches particularly in corpus linguistics, have shown that there are discernible distributional and predictable patterns in the input to children. For example, grammatical knowledge can be learned from patterns in CDS (Gegov et al., 2011; Freudenthal et al., 2007; Redington et al., 1998; Cartwright and Brent, 1997). Distributional patterns are also predictors of different grammatical categories to varying degrees, depending on the grammatical properties of the language (Mintz, 2003; Stoll et al., 2009; Stumper et al., 2011; Moran et al., In press).

Gegov et al. (2011) call these *invisible patterns*, which they aim to discover using network theory to model language acquisition data. This line of inquiry is summarized by Vitevitch (2008), Beckage et al. (2011) and Gegov et al. (2011). Networks have many properties that allow us to model,

[1] Another example is verbal inflection: English typically has four forms, e.g. kick, kicks, kicked, kicking. But compare Chintang, a language spoken in rural Nepal. It has more than 4000 inflectional forms per verb (Stoll et al., 2017).

compare and visualize data from vastly different input sources (Mihalcea and Radev, 2011). Network analysis has been applied to various issues in language acquisition (Ke and Yao, 2008; Vitevitch, 2008; Solé et al., 2010; Beckage et al., 2011; Gegov et al., 2011). One area of research into the input that children receive has been to model language-specific child-directed speech as lexical adjacency network graphs (Adamo and Boylan, 2008; Ke and Yao, 2008).

Small-world patterns (low average path length (L) and high clustering coefficient (C), see definitions below) and scale-free structures (power-law degree distribution) purportedly may be the product of language evolution towards an optimal cost-path navigation in the mental lexicon for speech production (Barabási and Albert, 1999; Ke, 2007).[2] This is perhaps not surprising given universal mechanisms of network formation, which as Ke (2007) notes, are "common pattern of life systems at all levels, ranging from food webs studied by ecologists, to the neural systems in the brain studied by neuroscientists which have been applied in computer sciences as artificial neural networks".

Ke (2007) advocates for networks as a means to model and investigate global structures in CDS and notes that convergent features in global networks appear whether or not those networks are encoded with semantic or grammatical relationships. These global structural characteristics reflect principles of self-organization of the lexicon and purportedly facilitate cognitive processing (see Discussion). Until now, only English and Chinese CDS have been investigated in detail with lexical adjacency networks. A cross-linguistic analysis has been presumably absent due to a lack of accessible and interoperable typologically-rich cross-linguistic data from longitudinal child language acquisition corpora.

2. Data and language sample

The ACQDIV corpus consists of ten longitudinal language acquisition corpora from nine languages, which are listed in Table 1.[3] The ACQDIV corpus focuses on the acquisition period from ages two-to-three.

ISO	Language	Speakers	Classification
ctn	Chintang	6K	Sino-Tibetan
cre	Cree	87.2K	Algic
ind	Indonesian	23.2M	Austronesian
jpn	Japanese	128.1M	Japanese
ike	Inuktitut	34.5K	Eskimo-Aleut
rus	Russian	166.2M	Indo-European
sot	Sesotho	5.6M	Niger-Congo
tur	Turkish	70.9M	Altaic
yua	Yucatec	766K	Mayan

Table 1: Language sample

These languages were selected from five clusters calculated via maximum diversity sampling (Stoll and Bickel, 2013) from the AUTOTYP database (Bickel et al., 2017) and from the World Atlas of Language Structures (Dryer and Haspelmath, 2013). The clustering algorithm identifies maximal diversity with respect to several widely-studied typological parameters, including presence and nature of agreement and case marking; word order; degree of synthesis; polyexponence and inflectional compactness of categories; syncretism; and inflectional classes.

In most corpora in ACQDIV, the recording sessions for each target child took place every other week (two corpora are much denser). Session lengths vary both within and across corpora and range from half an hour to four hours. All recording sessions were transcribed and morphologically glossed.[4] The size of the corpora also vary considerably, as shown in Table 2.

Corpus	Utterances	Sessions
Chintang	393030	477
Cree	20648	25
Indonesian	915759	997
Inuktitut	46683	77
Japanese	437348	362
Russian	827589	450
Sesotho	69575	115
Turkish	401262	373
Yucatec	93185	125

Table 2: Corpus size

Each corpus was developed and coded independently. Six corpora are encoded in CHILDES/CHAT or TalkBank XML (MacWhinney, 2000). Three are encoded in SIL's Toolbox in project-specific schemas. In recent work we describe how we transformed these different data formats into a single uniform and normalized database (Moran et al., 2016), which we query, analyze and visualize with various tools including SQLite,[5] R (R Core Team, 2013) and Python (using Networkx (Hagberg et al., 2008), NLTK (Bird et al., 2009), Numpy (Walt et al., 2011), Pandas (McKinney and others, 2010), SciPy (Jones et al., 2001))[6] and Gephi (Bastian et al., 2009).

3. Method

A network is a graph data structure that consists of vertices and edges (nodes and links), or formally: G = (V, E) where V is a collection of vertices, $V = \{V_i, i = 1, n\}$, and E is a collection of edges over V: $V, E_{ij} = \{(V_i, V_j), V_i \in V, V_j \in V\}$. For a thorough description of graphs and graph types in regard to network analysis with child language acquisition data, see Pajovic (2016).

[2] Recent work by Broido and Clauset (2018) shows that scale-free networks are actually rare across scientific domains. Whether the scale-free property exists in lexical adjacency networks of child-directed speech should be investigated.

[3] Three letter language name identifiers are ISO 639-3 codes. Population figures are from the Ethnologue (Lewis et al., 2009).

[4] Additional annotation layers of interest such as utterance-level translations, time stamps for the beginning and end of utterances, coding for addressees, morpheme segmentation, and part-of-speech tags are available for all corpora to various degrees.

[5] `https://www.sqlite.org/`

[6] `https://www.python.org/`

We create each lexical co-occurrence graph by splitting utterances on white space characters to delimit unique word forms as they are transcribed by experts for each language in our sample. Unique word forms represent the nodes in a network. A link is placed between two nodes if they directly co-occur within an utterance. An example is given in Figure 1. The size of the nodes in this example is determined by the degree of the node, i.e. the more links a node has, the bigger it is drawn. This network was produced from these two example utterances from Russian, below:[7]

(3) a. *xxx* *idi* *mjachik* *narisuju.*
M.SG.NOM.AN go.IPFV.IMP.2SG ball.M.SG.ACC.INAN draw.PFV.NPST.1SG
xxx come I'll draw you a little ball.

b. *Idi* *mjachik* *narisuju.*
go.IPFV.IMP.2SG ball.M.SG.ACC.INAN draw.PFV.NPST.1SG
Come I'll draw you a little ball.

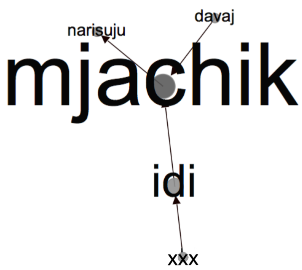

Figure 1: Lexical adjacency network for two Russian utterances

This example illustrates that nodes appearing multiple times in the data set will only appear as a single node in the graph network. If an utterance only consists of a single word (and if this word never appears in any multi-word utterance), this word will be placed as a 'lonely' node in the network. This procedure actually creates a so-called *multidigraph*, because it allows multiple links from a source to a target node. We use the graph library NETWORKX (Hagberg et al., 2005) to convert multidigraphs into *weighted digraphs* in which the weight of edges correspond to the frequency of the edge from target to source node.

For each language, we create a single network graph that models pooled child-directed speech from all caregivers to children between the ages of two and two-and-a-half. For example, our Russian CDS network contains all the pooled utterances from 24 adults. We think this concatenation is necessary given that the actual amount of speech children are exposed to is much greater than the small fraction of the actual input each child receives during the timespan under observation. Further we are not concerned here with the output of the children, but instead what input a hypothetical learner will encounter during the language acquisition process. Therefore we also include child-surrounding speech when it is available.

Research applying network theory to questions in child language acquisition has typically focused on three types of networks: co-occurrence, syntactic and semantic, and network parameters with high coverage (for a summary, see Gegov et al. (2011) and Pajovic (2016)).

For each language in our sample, we create a lexical adjacency network and load it into R. We use the IGRAPH libraries (Csardi and Nepusz, 2006) to calculate:

- N: the total number of nodes
- E: the total number of edges
- <k>: the key parameter, i.e. the average number of links adjacent to a node
- L: the average length of the shortest path between all pairs of nodes (average geodesic length)
- C: the clustering coefficient, i.e. the likelihood of neighbors of a node being connected, averaged across all nodes

The first three statistics are straightforward. The fourth consists of the short average path length (L) from one node to every other node. It is the most characteristic feature of small-world networks. Here 'small' refers to the fact that any two nodes in such a network can be reached through a few intermediate nodes (Watts and Strogatz, 1998; Ke, 2007; Milgram, 1967).

The fifth network statistic that we measure is the clustering coefficient of a node V_i. It is defined as the number of all edges between all of the nodes in a connected neighborhood divided by the total number of possible edges in the entire neighborhood of V_i. A clustering coefficient of 0 suggests that no neighbor of a node is connected to the other neighbors of that node; 1 means that all neighbors of a node are connected to each other. Values between 0 and 1 imply that there is a number of neighbors of a node which are also neighbors of each other (Watts and Strogatz, 1998; Mihalcea and Radev, 2011). Compared to random networks of equal size, small-world networks have a much higher clustering coefficient (Ke, 2007).

To evaluate each statistic, we use the standard approach of constructing random graphs, then we calculate their statistics for the five parameters above, and then we compare the two sets. To generate random networks, we use the Erdős-Rényi *G(n,p)* model, where *n* is the number of nodes in the network we want to compare (e.g. the Russian CDS lexical adjacency graph), and *p* is the probability of edge creation (also calculated from each lexical adjacency graph). Our random networks are directed graphs and we calculate *p* as $2m/(n(n-1))$ where *m* is the number of edges in our input.

4. Results

Our results are given in Table 3. The networks constructed from CDS all show small-world characteristics: their number of edges are greater than in the randomly-generated graphs; degree is higher in the random graphs; the principally short average path lengths are similar as in the random

[7]The child's name has been anonymized in this example.

graphs; and the clustering coefficient is much higher in the CDS networks than in the random graphs.[8]

When we compare the networks across languages, Inuktitut has the smallest number of edges and also the smallest node degree, but the highest average path length. This is in-line with linguistic expectations given Inuktitut's regular agglutinative morphology; there are few combinations of bigrams delimited by white space. On the other hand compare Indonesian, which has a higher key parameter (the number of connections a node has). This finding is also in-line with linguistic expectations. As illustrated above, Indonesian's morphology is isolating and words are combined much more frequently than in the morphologically more complex languages in our sample. Overall, we see interesting differences between the network parameters in Table 3 that reflect differences in the typological structures of languages in our sample, which we plan to explore in detail in future work.

5. Discussion

Although the *ability* to learn language is held to be innate, non-nativist and input-based approaches to language acquisition theorize that children are not born with grammatical categories or rules, but acquire them by generalizing from the CDS that they hear. Hence grammatical categories may be so-called emergent, that is, they emerge during the language acquisition process (e.g. Tomasello (2009), Cohn (2011), and Theakston and Lieven (2017)) and are not hardwired into our genetics.

Therefore one area of important research is to examine CDS from typologically maximally diverse languages and to identify distributional patterns in the input that appear cross-linguistically. Network theory is one tool for modeling CDS and for mining patterns in it.[9]

In this paper we show that typologically diverse and morphologically very different languages all exhibit small-world network properties when we model CDS as lexical adjacency graphs. Our finding is in line with child language acquisition models that have defined network links in terms of semantic or grammatical relationships, both of which exhibit convergent features in their global structures (Ke, 2007), but of course more work is needed, cf. Telesford et al. (2011).

What we have not shown and cannot answer at this point is whether distributional patterns facilitate cognitive processing. This is of course a key question in cognitive science and beyond the scope of this paper. Regarding small world characteristics, it is not difficult to imagine how their characteristic properties, including efficient information transfer and properties of regional specialization, could account for universal properties like fast retrieval from the mental lexicon. However, more substantive work is needed to show for example that small world properties constrain memory models to facilitate retrieval, e.g. Reitter and Lebiere (2012). Nevertheless, to answer whether general-purpose mechanisms are involved in language learning, we need to also know what distributional regularities exist in languages cross-linguistically, so we can determine on which mechanisms they might operate.

6. Acknowledgements

The research leading to these results has received funding from the European Unions Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 615988 (PI Sabine Stoll). We gratefully acknowledge Shanley Allen, Aylin Küntay, and Barbara Pfeiler, who provided the Inuktitut, Turkish, and Yucatec data for our analysis, respectively. We also thank three anonymous reviewers for their feedback.

7. Author contributions

SM, DP, SST designed the research. DP, SM performed the research and analyzed the results. SST provided data. SM wrote the paper.

8. Bibliographical References

Adamo, M. and Boylan, S. (2008). A network approach to lexical growth and syntactic evolution in child language acquistion. *Unpublished manuscript*.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *International AAAI Conference on Weblogs and Social Media*.

Beckage, N., Smith, L., and Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS one*, 6(5):e19348.

Bickel, B., Nichols, J., Zakharko, T., Witzlack-Makarevich, A., Hildebrandt, K., Rießler, M., Bierkandt, L., Zúñiga, F., and Lowe, J. B. (2017). The AUTOTYP typological databases. version 0.1.0. Online: `https://github.com/autotyp/`.

Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

Brittain, J. (2015). Corpus of the Chisasibi Child Language Acquisition Study (CCLAS). Source: 19-A1-2006-08-16ms.u289. Online: `http://phonbank.talkbank.org/access/Other/Cree/CCLAS.html`.

Broido, A. D. and Clauset, A. (2018). Scale-free networks are rare. *arXiv preprint 1801.03400v1*.

Cartwright, T. A. and Brent, M. R. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition*, 63(2):121–170.

[8]Interestingly, these properties are found in the networks created from the utterances produced by each child. When compared to the adults' graphs, the children's graphs show much lower average degrees, but this is to be expected as the vocabulary size of the children is much smaller. We will follow this line of research elsewhere.

[9]Critics note that graphs and matrices (or data tables) are interchangeable data structures and that networks provide no additional benefits than matrices. They are wrong, however, because networks can also be visualized and therefore provide researchers with additional tools and techniques for exploratory analysis, e.g. network growth models in time-series analysis.

CDS	N	E	E_random	<k>	<k>_random	L	L_random	C	C_random
Chintang	60535	173350	347894	5.727265	11.49398	3.861909	6.497079	0.01949274	0.0001666573
Cree	4577	9945	19943	4.345641	8.714442	3.925917	5.89046	0.0354484	0.002096497
Indonesian	26264	257285	516430	19.59222	39.32607	3.174789	3.73446	0.07556786	0.001496866
Inuktitut	11705	7851	15703	1.341478	2.683127	6.714106	23.04485	0.008594746	0.000283802
Japanese	24051	151235	303140	12.57619	25.2081	3.156601	4.261239	0.03455292	0.001044477
Russian	47895	290832	581179	12.14457	24.26888	3.382483	4.599158	0.03299296	0.0004945935
Sesotho	5519	22409	44735	8.120674	16.21127	3.265634	4.363033	0.05016898	0.002802469
Turkish	61537	324164	648276	10.53558	21.06947	3.621385	4.933678	0.03675761	0.0003510219
Yucatec	15967	38638	77134	4.839732	9.661677	4.020815	6.317499	0.02289779	0.0006281727

Table 3: Network parameters comparison

Chomsky, N. (1957). *Syntactic Structures*. Mouton and Co.

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge University Press.

Cohn, A. (2011). Features, segments, and the sources of phonological primitives. In G Nick Clements et al., editors, *Where do phonological features come from? Cognitive, physical and developmental bases of distinctive speech categories*, pages 15–42. John Benjamins Publishing Company.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5):1–9.

Matthew S. Dryer et al., editors. (2013). *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., and Gobet, F. (2007). Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science*, 31(2):311–341.

Gegov, E., Gobet, F., Atherton, M., Freudenthal, D., and Pine, J. (2011). Modelling language acquisition in children using network theory. In B. Kokinov, et al., editors, *European Perspectives on Cognitive Science*. New Bulgarian University Press.

Gil, D. and Tadmor, U. (2007). The MPI-EVA Jakarta Child Language Database. A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology and the Center for Language and Culture Studies, Atma Jaya Catholic University. Source: HIZ-1999-05-20.0556. Online: `https://jakarta.shh.mpg.de/acquisition.php`.

Hagberg, A., Schult, D., and Swart, P. (2005). NetworkX: Python software for the analysis of networks. Technical report, Technical report, Mathematical Modeling and Analysis, Los Alamos National Laboratory, 2005. http://networkx. lanl. gov.

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August.

Hall, T. A., Hildebrandt, K. A., and Bickel, B. (2008). Introduction theory and typology of the word. *Linguistics*, 46(2):183–192.

Jones, E., Oliphant, T., Peterson, P., et al. (2001). SciPy: Open source scientific tools for Python. Online: `http://www.scipy.org/`.

Ke, J. and Yao, Y. (2008). Analysing language development from a network approach. *Journal of Quantitative Linguistics*, 15(1):70–99.

Ke, J. (2007). Complex networks and human language. *arXiv preprint cs/0701135*.

Lewis, M. P., Simons, G. F., and Fennig, C. D. (2009). *Ethnologue: Languages of the world*, volume 16. SIL international Dallas, TX.

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

McKinney, W. et al. (2010). Data structures for statistical computing in Python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. SciPy Austin, TX.

Mihalcea, R. and Radev, D. (2011). *Graph-based natural language processing and information retrieval*. Cambridge University Press.

Milgram, S. (1967). The small world problem. *Psychology Today*, 2(1):60–67.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1):91–117.

Moran, S., Schikowski, R., Pajović, D., Hysi, C., and Stoll, S. (2016). The ACQDIV database: Min(d)ing the ambient language. In Nicoletta Calzolari (Conference Chair), et al., editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Moran, S., Blasi, D. E., Schikowski, R., Küntay, A. C., Pfeiler, B., Allen, S., and Stoll, S. (In press). A universal cue for grammatical categories in the input to children: frequent frames. *Cognition*.

Pajovic, D. (2016). Acqdiviz: Visualising development in longitudinal first language acquisition data. Master's thesis, University of Zurich.

R Core Team, (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Redington, M., Chater, N., and Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22(4):425–469.

Reitter, D. and Lebiere, C. (2012). Social cognition: Memory decay and adaptive information filtering for robust information maintenance. In *Proceedings of the Twenty-*

Sixth Conference on Artificial Intelligence (AAAI-12), Toronto, Canada.

Solé, R. V., Corominas-Murtra, B., Valverde, S., and Steels, L. (2010). Language networks: their structure, function, and evolution. *Complexity*, 15(6):20–26.

Stoll, S. and Bickel, B. (2013). Capturing diversity in language acquisition research. *Language Typology and Historical Contingency: In Honor of Johanna Nichols. Amsterdam: John Benjamins*, pages 195–216.

Stoll, S., Abbot-Smith, K., and Lieven, E. (2009). Lexically restricted utterances in russian, german and english child-directed speech. *Cognitive Science*, 33:75–103.

Stoll, S., Mazara, J., and Bickel, B. (2017). The acquisition of polysythetic verb forms in Chintang. In Michael Fortescue, et al., editors, *Handbook of Polysynthesis*. Oxford University Press.

Stumper, B., Bannard, C., Lieven, E., and Tomasello, M. (2011). Frequent frames in German child-directed speech: A limited cue to grammatical categories. *Cognitive Science*, 35(6):1190–1205.

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., and Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain Connectivity*, 1(5):367–375.

Theakston, A. and Lieven, E. (2017). Multiunit sequences in first language acquisition. *Topics in Cognitive Science*.

Tomasello, M. (2009). *Constructing a language*. Harvard University Press.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*, 51(2):408–422.

Walt, S. v. d., Colbert, S. C., and Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22–30.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684):440–442.