Evidence of community structure in phonological networks of various languages

CogSci 2023
29 July 2023
11:50 AM - 1:50 PM
P3-W-331-1447

Jazton Chern[1], Nichol Castro[2], & Cynthia S. Q. Siew[1]
[1]Department of Psychology, National University of Singapore
[2]Department of Communicative Disorders and Sciences, University at Buffalo

Contact:
Cynthia Siew, cynthia@nus.edu.sg
Jazton Chern, jazton@u.nus.edu
Nichol Castro, nicholca@buffalo.edu

Introduction

- Phonological networks across various languages share similar *macro-level* network properties
  - i.e., similar degree distribution and small-world structure (Arbesman, Strogatz & Vitevitch, 2018)
- Aim of current study: Zoom in to examine if *meso-level* properties of phonological networks are also similar across various languages.
  - This is based on Siew (2013), who found robust **community structure** for English
  - This study aims to expand on these findings via an analysis of community structure across other languages.
  - Data was derived from the CLEARPOND database (Marian et al., 2012) for the following four languages: Dutch, French, German, Spanish

What is community structure?

- Community structure refers to the presence of distinct groups or clusters of nodes within a network, characterised by dense internal connections and sparse connections between communities
- Modularity, *Q*, measures density of links within communities compared to links between communities (Newman, 2006)
- *Q* values closer to 1 indicate higher quality and distinctness of the community partitions
- Community detection was done using the Louvain method (Blondel et al., 2008)

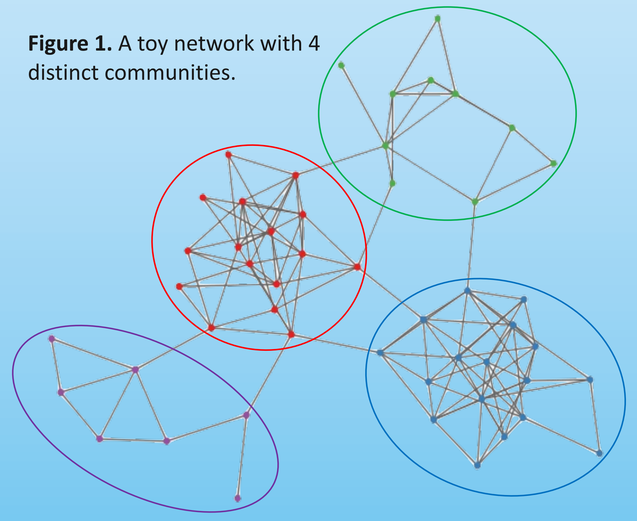

Figure 1. A toy network with 4 distinct communities.

Scan this QR code for the PDF version of this poster!

Method

Network Construction

- Each node is a real word that is represented by a phonological transcription, where every phoneme of a word is represented by a single character
- Edges placed between words where the Levenshtein distance between each of their phonological transcriptions is 1 (Luce & Pisoni, 1998)
  - Levenshtein distance refers to the number of single character edits (substitution, deletion, addition) required to transform one word to another
  - E.g., “at” /@t/, “bat” /b@t/, and “scat” /sk@t/ are phonological neighbors of “cat” /k@t/.
- Network edges are unweighted and undirected

Communities

- Networks of random communities constructed to serve as baseline comparison
  - Edges are placed randomly
  - All other variables kept constant
- Compare differences in lexical characteristics between communities in the network.
  - Mean no. of phonemes, word frequency, phonological neighbourhood density, and phonological neighbourhoood frequency were correlated to community size for each community.
- Compare biphone distributions between communities
  - Biphone: Phonological segment consisting of sequences of 2 consecutives phonemes occurring in words in the language

		% of Total	Real Network			Mean of 100 Random (ER) Networks	
Language	LCC Size	% of All Words	Q	Av. Degree	# of Comm [Min, Max]	Q	Av. Degree
Dutch	8527	0.307	0.743	7.23	44 [4, 1232]	0.332	7.23
French	15695	0.566	0.715	17.35	90 [3, 1727]	0.198	17.35
German	9986	0.360	0.734	7.28	54 [4, 864]	0.330	7.28
Spanish	8990	0.324	0.813	5.41	81 [6, 622]	0.414	5.41

Table 1. Modularity, Q, of Real vs Random Networks.

	Dutch		French		German		Spanish	
	Real	Random	Real	Random	Real	Random	Real	Random
Phonemes	-0.27 *	0.064	-0.52 ***	0.031	-0.33 *	0.063	-0.43 ***	0.020
Freq	0.45 **	0.057	0.24 *	-0.053	0.26 +	-0.021	0.096	-0.057
P Density	0.62 ***	0.030	0.79 ***	-0.050	0.77 ***	0.057	0.85 ***	-0.046
P NHF	0.23	0.14	0.57 ***	-0.068	0.60 ***	0.23 +	0.14	0.23 *

Note. Significance levels: ***$p < .001$, **$p < .01$, *$p < .05$, +$p < .10$.

Table 2. Correlation Coefficients of Lexical Characteristics of Real vs Random Communities to Community Size.

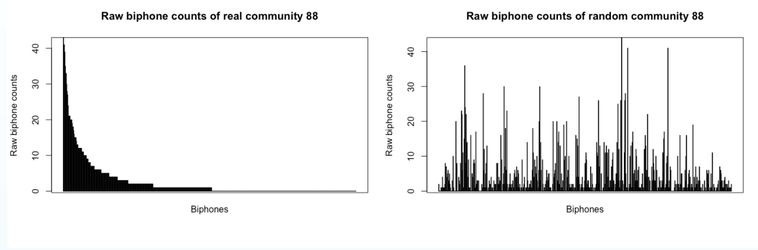

Figure 2. Raw Biphone Counts of a Real Community in French

Figure 3. Raw Biphone Counts of a Random Community in French

Discussion

- Robust community structure exists in the phonological networks of Dutch, French, German, and Spanish, and there are large overlaps with what has been previously established for English.
  - ✓ Each community is readily distinguished by the lexical properties of its constituent words
    - Larger communities tended to have shorter, highly frequent words with many phonological neighbors relative to smaller communities.
  - ✓ Biphone distributions within communities share a similar pattern to that of word frequencies in a language, i.e., a *power law* distribution (Zipf, 1935).
  - ✓ The most frequently occurring biphones in each real community can be combined to form longer phonological segments.
    - In Dutch community 43, the five most common biphones are “Nk”, “aN”, “Ik”, “Il”, and “IN” and can easily be combined by joining biphones via a shared phoneme, e.g., “aN” + “Nk” = “aNk”.
- Other attributes of community structure found by Siew (2013) for English may be extended to the 4 languages here, and potentially beyond as well.

References

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos, 20*(03), 679-685. DOI: 10.1142/S0218127410026078

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. DOI: 10.1088/1742-5468/2008/10/P10008

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing,19*(1), 1.

Marian, V., Bartolotti, J., Chabal, S., Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE 7*(8): e43230. DOI: 10.1371/journal.pone.0043230

Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences, 103*(23), 8577-8582. DOI: 10.1073/pnas.0601602103

Siew, C. S. (2013). Community structure in the phonological network. *Frontiers in Psychology, 4*, 553. DOI: 10.3389/fpsyg.2013.00553

Zipf, G. K. (1935). *The Psycho-Biology of Language.* Boston, MA: Houghton-Mifflin.