Can Greenbergian Universals be induced from language networks ?

Kartik Sharma, Kaivalya Swami, Aditya Shete, Samar Husain

Indian Institute of Technology, Delhi

August 29, 2019

Structure

Language as a network

Syntactic dependency networks

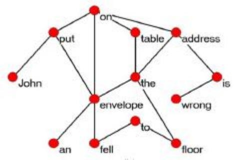


Figure: Cong and Liu, 2014

Co-occurrence networks

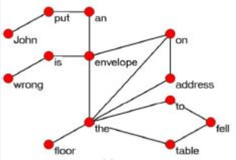


Figure: Cong and Liu, 2014

Syntactic Priming

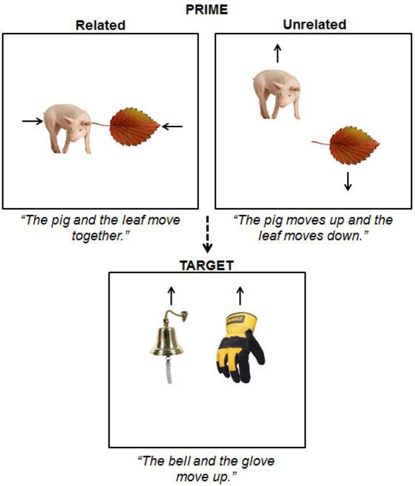


Figure: Hardy et al., 2018

Residual Activation Model

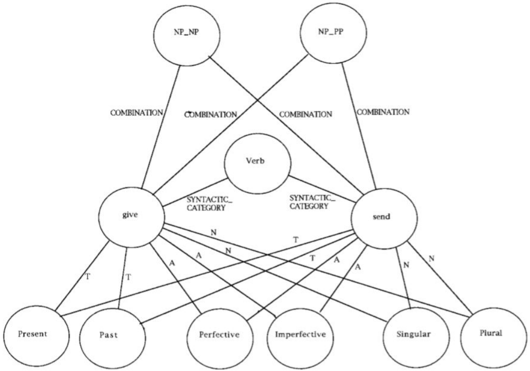


Figure: Representation of syntactic information *(Pickering and Branigan, 1998)*

Greenberg's universals

- The empirical generalizations shown by Greenberg (1963) highlights various universal word-order correlations.
- Most of these generalizations are **implicational** in nature.
- These universals have been tested on over 600 languages around the world (Dryer, 1992).
- Such results clearly mean that any model representing natural language grammar must be able to **induce** these universals/correlations.
- Residual Activation model is one such model and in this work, we investigate whether it is possible or not.

Brief overview and motivation

- We have used the ideas from residual activation model in order to transform a **syntactic dependency network** (as used by Liu and Li, 2010 etc.) into different layers (lemma (verbs), combinatorial) so that it **simulates the model** proposed by Roelofs (1992, 1993).

Brief overview and motivation

- We have used the ideas from residual activation model in order to transform a **syntactic dependency network** (as used by Liu and Li, 2010 etc.) into different layers (lemma (verbs), combinatorial) so that it **simulates the model** proposed by Roelofs (1992, 1993).
- Since most of the GUs have **word-order as its premise**, we chose the **word-order and the argument structure** as defining properties of the **combinatorial node**.

Brief overview and motivation

- We have used the ideas from residual activation model in order to transform a **syntactic dependency network** (as used by Liu and Li, 2010 etc.) into different layers (lemma (verbs), combinatorial) so that it **simulates the model** proposed by Roelofs (1992, 1993).
- Since most of the GUs have **word-order as its premise**, we chose the **word-order and the argument structure** as defining properties of the **combinatorial node**.
- In order to test whether the universals are induced, we checked whether the conclusion of the implication is **directly/indirectly being led to** by some graph-theoretic property of the combinatorial nodes.

Data and tools used

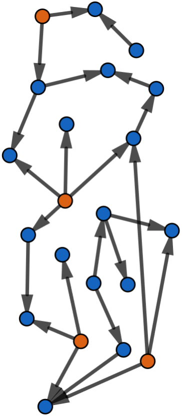
- For making language networks, we used the corpus of standard treebanks from **Universal Dependencies (UD)**.
- We are using 34 languages in this work, which are chosen according to the no. of sentences in the corpus.
- For analyzing the resulting complex networks, we used **Cytoscape** software after converting the treebank data (in **CoNLL-U** format) to node and edge list.
- We used **World Atlas of Language Structures (WALS)** in order to obtain structural properties of languages (for example, adposition order, noun-rel order etc.) required for clustering tasks in our experiments regarding GUs.

Brief outline

UD corpus

Brief outline

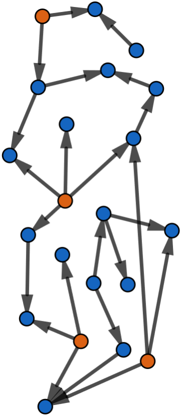
UD corpus



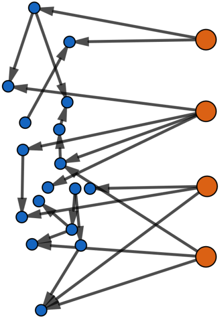
Base network

Brief outline

UD corpus



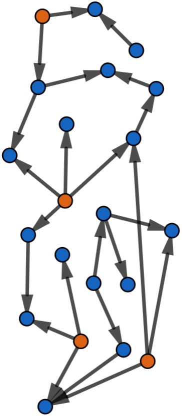
Base network



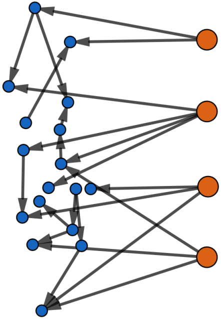
Layer 1 (orange)

Brief outline

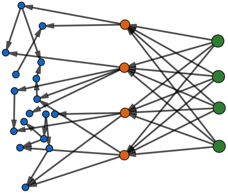
UD corpus



Base network



Layer 1 (orange)



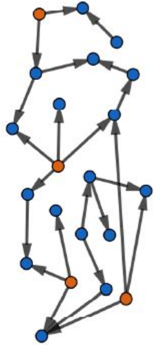
Layer 2 (green)

UD to base network

Node Data

- UD stores annotated sentences as a list of word lines.
- Each such word line is stored as characteristics of a node with an identifier which is just the pair of LEMMA and UPOS.
- Two word lines with same LEMMA and UPOS will be considered the same node.

UD corpus



Base network

UD to base network

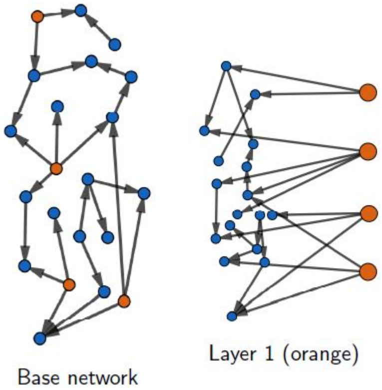
Edge Data

Considering one word line at a time, an edge stores:

- **Source/Target nodes**: The HEAD field of word line denotes the index of the head dependency of this word. We used node identifier (LEMMA:UPOS) of this word line as the target node and the identifier of the corresponding head as the source node. If the HEAD is 0, then there is no edge.
- **DEPREL**: DEPREL field of a word line denotes the relation class of this dependency (for example, "nsubj", "nobj" etc.)
- **Linear distance**: HEAD - ID of a word line gives the linear distance in the sentence for the corresponding dependency.

Base network to Layer 1

1. Verbs are filtered from the nodes of this base network.
2. Next, we pruned out all the **non-finite instances** of dependency edges from each verb.
3. For each verb, we considered the dependencies with **subject, direct object and indirect object** only and found two measures for each argument class :
   - Average Frequency
   - Average Distance



Base network to Layer 1

Average frequency

- DEPREL is checked for particular argument class (for example, for subject, we look for "nsubj", "csubj", "nsubj:pass" in DEPREL field of each edge in each verb)
- Average frequency is then calculated as:

$$\text{Average frequency} = \frac{\text{no. of edges with a particular core argument class}}{\text{Total no. of edges with core arguments}}$$

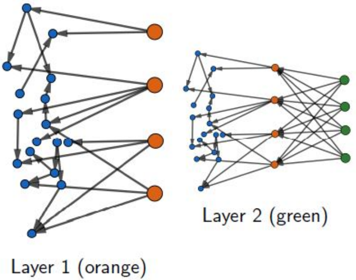
Average distance

- DEPREL is checked for a particular argument class
- Average distance is then calculated as:

$$\text{Average distance} = \frac{\text{Sum of linear distance of dependencies with a particular arg class}}{\text{No. of edges with a particular arg class}}$$

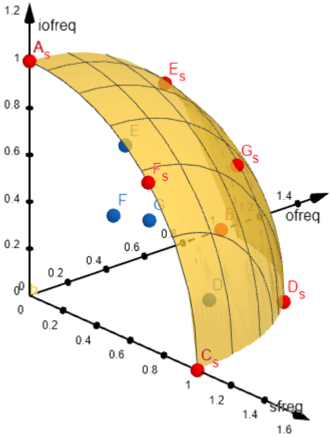
Layer 1 to Layer 2

- For each verb, we considered the **frequencies and distances** for each argument class in order to generate **probabilities** of the verb having certain classes.
- These classes are defined by both - word order information and arguments information. For example, "SV", "SVO" etc.
- These classes constitute the Layer 2 and the edges from layer 1 to layer 2 store the corresponding probabilities found for each such class.

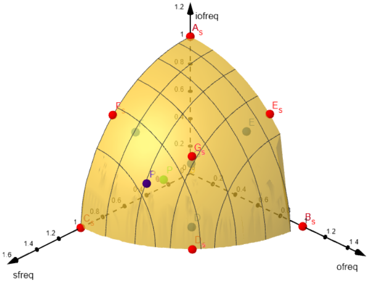


Finding the probabilities for verb classes

- The distances (d_s, d_o, d_i) are used to find the ordering of the arguments relative to the verb and thus the word-order.
- In order to find the probability of a verb possessing a particular argument structure, we use co-ordinate system on (f_s, f_o, f_i) as shown in adjoining figure. Here, $A, B, \cdots$ are corresponding to the possible argument structures and $A_s, B_s, \cdots$ denote their spherical projections.



Finding the probabilities for verb classes



Suppose that a verb corresponds to a point P.

1. The point closest to P among the blue points is found (let it be C^*).
2. P is projected on the sphere giving P_s.
3. Probability is assigned to the red points by considering the spherical distance between P_s and the red points $(D(P_s, C_i))$.
4. Probability is given according to : $P(P_s, C_i) = N(D(P_s, C_i) - D(P_s, C^*), \sigma^2)$, $N(\mu, \sigma^2)$ denotes random variable following normal distribution, σ^2 is arbitrarily chosen.

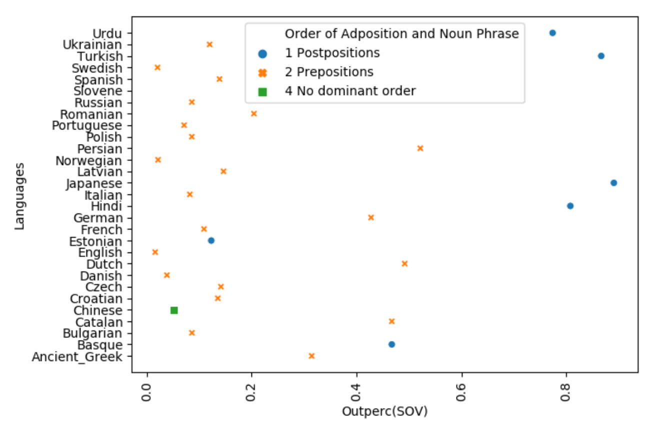
Such a spherical representation is chosen because each of the red points are equidistant in such a space.

Experiment 1: Description

- The whole network is analyzed using a network analyzing software, in particular, Cytoscape.
- A parameter **'Outperc'** is devised for word-order related nodes (SVO, SOV, VSO etc.) , which is defined as outdegree of the node divided by the sum of outdegree of all such nodes.
- Here, we considered only the word-order based Greenbergian universals, in particular, GU 1, 3, 4, 5, 6, 12.
- The statements given by *Greenberg, 1963* were directly related with "Outperc" or "Outdegree" of the layer 2 nodes considering that these parameters directly correlate with the **likelihood** of the concerned language to have a particular **word-order**. We used WALS for obtaining required typological information for each language.

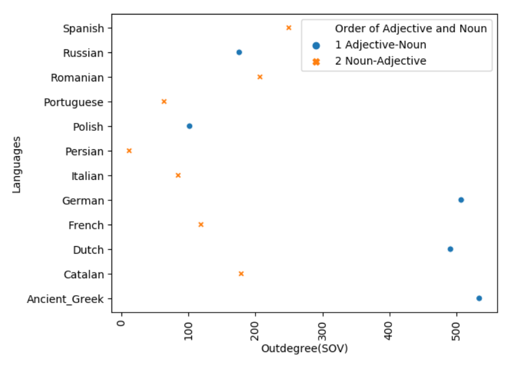
Results

GU 4: *With overwhelmingly greater than chance frequency, languages with normal SOV order are post-positional.*



Results

GU 5: *If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun.*



Experiment 2: Description

- In the previous experiment, we **restricted our search** for the structural properties to only "Outperc" of a specific node.

Experiment 2: Description

- In the previous experiment, we **restricted our search** for the structural properties to only "Outperc" of a specific node.
- But since the network is interconnected, one should also look into the other graph-theoretic properties like *Degree, Eccentricity, Neighborhood Connectivity* etc.

Experiment 2: Description

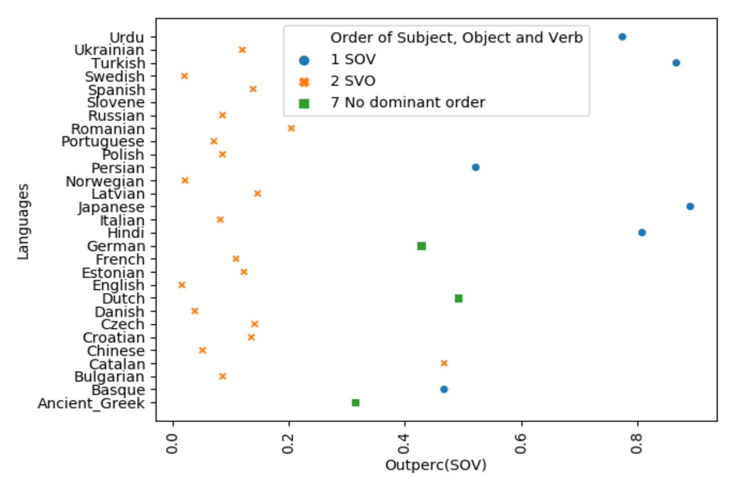
- In the previous experiment, we **restricted our search** for the structural properties to only "Outperc" of a specific node.
- But since the network is interconnected, one should also look into the other graph-theoretic properties like *Degree, Eccentricity, Neighborhood Connectivity* etc.
- This motivated us to do an **unsupervised search** over all the node parameters over all layer 2 nodes, in order to find the one which is distributed according to some structural property of a language.

Experiment 2: Description

- In the previous experiment, we **restricted our search** for the structural properties to only "Outperc" of a specific node.
- But since the network is interconnected, one should also look into the other graph-theoretic properties like *Degree, Eccentricity, Neighborhood Connectivity* etc.
- This motivated us to do an **unsupervised search** over all the node parameters over all layer 2 nodes, in order to find the one which is distributed according to some structural property of a language.
- In order to assess these **clusters** identified by WALS data, we make use of the **silhouette score and visual evaluation**.

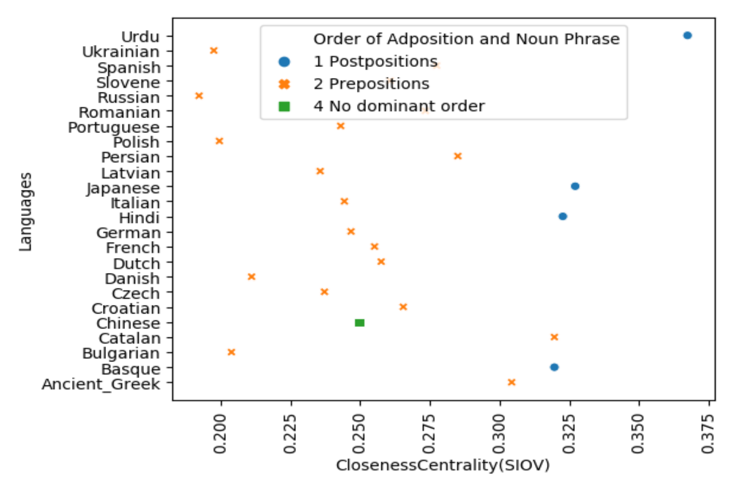
Results

Order of Subject, Verb, Object



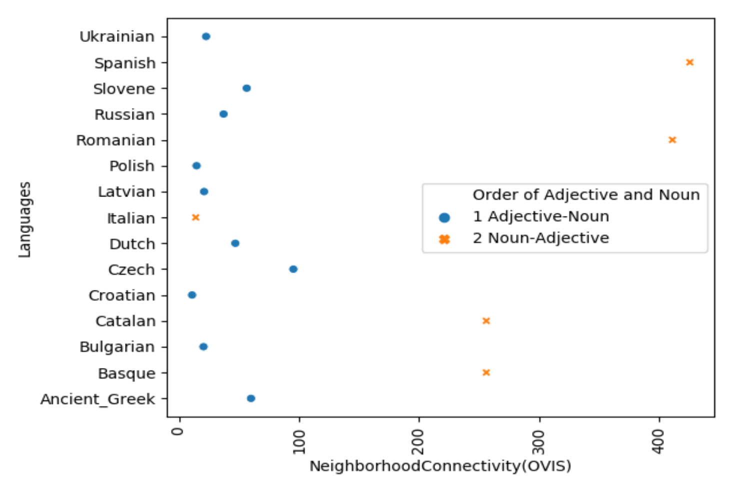
Results

Order of Adposition and Noun Phrase



Results

Order of Adjective and Noun Phrase



What have we shown ?

- Our work gives further evidence about the possibility of network being the **structure of the underlying network**.
- Our work provides some support that word-order generalizations can be **automatically derived** from a network if **conceptualized** in a meaningful way.
- We showed in Experiment 2 that some syntactic information can be derived from the distribution of various parameters of **just a few nodes** without considering the rest of the network.
- We understand that some results are not very strong due to multiple factors like *treebank size, alignment of languages in UD and WALS, inconclusiveness of a proper clustering score* etc. But they definitely are stepping stones into potential further studies in this aspect.

What can be done ?

- Rather than abstracting verbal information, one can also use **other syntactic categories** to form the layers.
- One can try to induce the **universals** or structural properties that are not considered here.
- Other **psycholinguistically inspired models**, for example, Long-term implicit learning (Bock and Griffin (2000)) can be modeled and investigated in a similar way.

Thank You

Presented by:
Kartik Sharma
cs1170342@cse.iitd.ac.in

Code available at:
https://github.com/Ksartik/SyntaxFest2019_paper18