# Can Greenbergian universals be induced from language networks?


## Abstract

Language networks have been proposed to be the underlying representation for syntactic knowledge (Roelofs, 1992; Pickering and Branigan, 1998). Such networks are known to explain various word order related priming effects in psycholinguistics. Under the assumption that word order information is encoded in these networks, we explore if Greenbergian word order universals (Greenberg, 1963) can be induced from such networks. Language networks for 34 languages were constructed from the universal dependencies treebank (Nivre et al., 2016) based on the assumptions in (Roelofs, 1992; Pickering and Branigan, 1998). We conducted a series of experiments to investigate if certain network parameters can be used to cluster various languages based on the word order typology proposed by Greenberg. Our results show that some network parameters robustly cluster the languages correctly, thereby validating language network as a valid representation for such linguistic generalizations.


## 1 Introduction

Establishing connections and relations between objects is an important way of representing knowledge (Siew et al., 2018). Such a way of representation lends itself to a succinct understanding of the structure and the complexity inherent in the knowledge representation. Such network representations are routinely used to understand complex systems such as social systems, biological systems, economic systems and so on (Pastor-Satorras and Vespignani, 2007; Caldarelli, 2007; Newman, 2010). Language seems well suited for this type of representation; after all, the knowledge of language and its use, is primarily about establishing relations between different kinds of linguistic objects (Borge-Holthoefer and Arenas, 2010; Solé et al., 2010). Indeed, the significance of such networks was appreciated quite early in the domain of meaning representation in terms of semantic relatedness (Collins and Loftus, 1975; Ober and Shenaut, 2006). Such networks encoding meaning have been shown to capture experimental results on lexical priming (McRae and Boisvert, 1998; McRae et al., 2005). Additionally, various resources (e.g., Wordnet) and as well as models (e.g., word2vec) have been proposed with the motivation of establishing relations between similar words (Miller, 1995; Mikolov et al., 2013). A network-based representation has also been proposed to subserve syntactic knowledge in the mind (Roelofs, 1992; Pickering and Branigan, 1998). Such a network has been claimed to correctly explain the syntactic priming effects in during language production/comprehension (Pickering and Ferreira, 2008; Tooley and Traxler, 2010). Indeed, recent computational work also employs network-based modelling to investigate human processing for various linguistic phenomenon (e.g., Futrell and Levy, 2019). Networks have also been used to quantify cognitive processes and representations related to various linguistic levels such as words, etc. (e.g Vitevitch, 2008; Allegrini et al., 2004; Chung and Pennebaker, 2007; Morrill, 2000; Vitevitch et al., 2011).

Network theory has been extensively used to understand (and visualize) such knowledge representations (Barabási, 2011). Network theory formalizes knowledge systems represented as networks, which contain nodes and edges describing the entities and the relations between them. Network theory enables us to extract specific information related to the connectedness and relationships between objects (Newman, 2010; Costa et al., 2011). The primary attraction of representing a complex system in the form of a network lies in its inherent property of visually presenting the relations present in the data, and the ability to abstract it to different levels. In effect, the network properties can be used to analyze a system effectively. The analysis can take place on various levels, ranging from a single node, to viewing the properties of the entire network as a whole (Albert et al., 2000).

The idea of language as a network has been gaining some traction in computational linguistics (e.g., Ferrer i Cancho et al., 2007; Lerner et al., 2009; Ke and Yao, 2008; Borge-Holthoefer and Arenas, 2010; Lerner et al., 2009; Choudhury et al., 2010; Ferrer I Cancho and Solé, 2001; Vitevitch et al., 2011; Ferrer i Cancho et al., 2004). One approach, that we explore here, is to construct language networks from annotated dependency treebank to encode syntactic relationship between lexical items. Previous works on such network representation of language have explored the properties of language networks formed through dependency treebanks (Ferrer i Cancho et al., 2004). Relatedly, Liu and Li (2010) used language network to successfully cluster languages into phylogenetic language groups using network parameters. As stated earlier, networks have also been hypothesized to be the representation that subserves syntactic knowledge in the mind (Roelofs, 1992; Pickering and Branigan, 1998). In particular, it has been used to explain syntactic priming with respect to various word order choices used during sentence comprehension and production (Pickering and Ferreira, 2008; Tooley and Traxler, 2010). This implies that networks can represent various syntactic rules (e.g., word order) in terms of nodes and their relationship with other nodes in the network. In other words, the network as a representation of language should contain the same generalisations as present in a language. Greenberg's universals (Greenberg, 1963) are a set of such generalisations that occur across languages. These universals and their status in language networks is the focus in this article.

In this work, we build a psycholinguistically motivated language network (Roelofs, 1992; Pickering and Branigan, 1998) for 34 languages to investigate if Greenberg's word order related language universals (GU) can be induced from the networks. To do this, we conduct two experiments. In the first experiment, we simply map the GUs onto a language network to see if a particular node property (percentage of outgoing arc) leads to the desired classification across languages. For example, for GU universal no. 3, we look at this parameter of the VSO nodes across albl language networks and see if the parameter values cluster the respective languages as prepositional or postpositional. In the second experiment, we automatically derive certain implicational universals stated by Greenberg. For example, we see which word order node (e.g., SVO, SOV, etc) best classifies the order of adposition and noun phrase. In effect, the first experiment is completely correlational and supervised – checking **if** a known node parameter leads to the correct language typology. The second experiment, is unsupervised – checking **which** node (and its parameter) leads to the correct language typology. Together, the two experiments shed light on whether language network can induce correct GU wrt word order and highlights the properties of the network where this information can be found.

The paper is arranged as follows. We begin with a description of the data, tools and network formation in the Section 2. In section 3 we present the two experiments and discuss the results. Following this, in section 4 we conclude the paper and list out some future directions.

## 2 Methodology

### 2.1 Data and Tools

We use the **'Universal Dependencies'** (UD for short, henceforth) treebank (Nivre et al., 2016; Agić et al., 2015) to create the network. The UD has annotated data for 76 languages, of which we are utilizing 34[1]. Only those languages were selected that had a relatively large size (sentence count more than 2k) and that were present in the WALS (The World Atlas of Language Structure) database (Dryer and Haspelmath, 2013). WALS data in .csv format is directly available from the WALS online source. The UD CoNLL-U format was converted into a network (edges and nodes data) format in order to use the Cytoscape (Shannon et al., 2003) software. Cytoscape is an open-source network visualization and analysis software.

### 2.2 Language Network

The language network derived from the UD data is motivated by the syntactic representation proposed by Roelofs (1992) and adapted by Pickering and Branigan (1998). This model has been used to explain syntactic priming during comprehension Pickering and Ferreira (2008) and production Tooley and Traxler (2010). The model consists of layers of linguistic elements connected to each other. Nodes representing word tokens are connected to 'lemma' nodes. The 'lemma' nodes are associated with syntactic information such as category, morphological information, etc. The 'lemma' nodes associated with the verbs are connected to the 'combinatorial' nodes representing their syntactic subcategorization information, in other words, the

[1]Ancient Greek, Arabic, Basque, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, French, German, Hebrew, Hindi, Indonesian, Italian, Japanese, Latvian, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Slovak, Slovene, Spanish, Swedish, Turkish, Ukrainian, Urdu

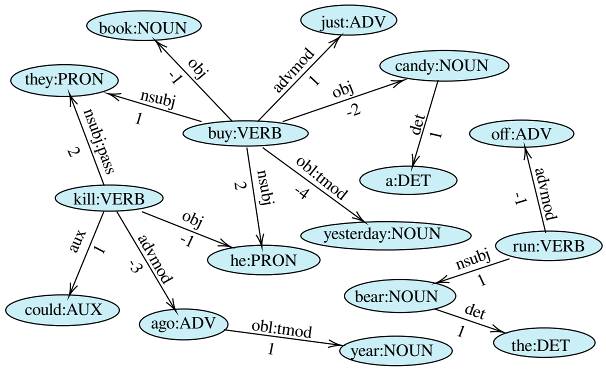



Figure 1: A sample base-network (see point 1 in section 2.2) derived from 4 sentences. These sentence are *'They could kill him years ago', 'He just bought a candy yesterday' 'The bear ran off', 'They buy books'*

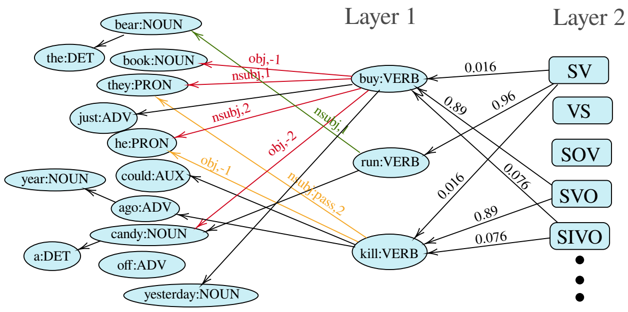



Figure 2: The final network derived from the base network above (Figure 1) for the same sentences. All reported results are on this network.

typological word order information. When a verb is required in speech, an activation of concept results in the activation of the highest activated 'combinatorial' node which in turn activates the relevant 'lemma' node. Interestingly, activation of this 'lemma' node results in the activation of syntactically similar verbs. This is because verb lemma that have similar syntactic properties are linked to the same combinatorial node. In this work, we prepare a similar network. This results in a layered network in which the last layer explicitly contains word-order properties such as 'SVO', 'SOV', 'VSO', etc. The 'combinatorial' node described in the network discussed in Pickering and Branigan (1998), is modelled here as a node which encapsulates the argument structure of the verb nodes connected to it.

Creation of the network is done in multiple steps, which we describe below. We illustrate this through figures 1 and 2 above.

1. Universal Dependencies data was converted to a node and edge data. The nodes are defined as the LEMMA of a word tagged with its part of speech (UPOS), which we will call LEMMA:UPOS. The other properties (e.g., FEATS) of each node given in the CoNLL-U format are also associated with each node. The edges between the nodes are directed and represent dependency links from HEAD of a word/node to the dependent node. In addition, the edges have certain attributes such as (a) linear distance: distance between the connected nodes based on the linear position of the nodes in the corresponding sentence (calculated as HEAD - INDEX from the CoNLL-U format), (b) dependency relation (DEPREL) : dependency relation between the nodes (provided as DEPREL in the CoNLL-U format. The resulting network at this stage is shown in Figure 1.

2. Next, we select only those verb lemma nodes that are finite[2] (obtained from VerbForm attribute in the FEATS column in the CoNLL-U data). This is done in order to have a more robust generalization regarding the argument structure of individual verbs as non-finite instances of verbs can drop their arguments. This leads to the formation of Layer 1 shown in Figure 2.

[2]The finiteness information is determined using both the FEATS of both verb lemma as well as its auxiliary.

3. We then create layer 2 (see figure 2) which has nodes corresponding to various word order possibilities of verb arguments, e.g., SOV, SVO, VSO, etc. These layer 2 'combinatorial' nodes are connected to the layer 1 lemma nodes. The connection between the lemma and the combinatorial node represents the probability of a verb appearing with a particular argument structure and its word order. We considered all the combinations (without replacement) of 'S' (denoting **subject**), 'V' (denoting **verb**), 'O' (denoting **object**), 'I' (denoting **indirect object**) containing at least one 'V'. Some of these nodes are : **SV, VS, SOV, SVO, SIOV** etc. Layer 2 thus consists of 48 pre-defined nodes[3] similar to combinatorial nodes in Pickering and Branigan (1998).

    - These combinatorial nodes are obtained by computing two layer 1 properties. These are average sentential distance of the core arguments (subject, object and indirect object) and their proportions. Average sentential distance is obtained by grouping all the nodes with argument relation edges and computing their average linear distance from the verb. This is done for each core argument. In addition we also compute the proportion of each core argument in a group relative to total no of core arguments for a verb in layer 1.
    - In order to connect the verb lemmas in layer 1 with the nodes in layer 2, we computed the probabilities with which these verbs appear in a specific argument structure configuration in the treebank. We assume that the word order of a certain verb remains same and it is just the argument structure that can show variations.[4] The average distance of the verb relative to the argument can be formalized as a tuple (*subj-dist*; *obj-dist*, *inobj-dist*), where, *subj-dist* is the average distance between the verb and the subject group, etc. For example, if the distances are (1; -1; 0), then the word order is SVO. If the word order is SVO, the concerned verb can connect to any of the following - SV, SVO, SVIO, SIVO, SVOI, ISVO.
    - In order to identify which one of the above possibilities the verb must have, we devised probabilities for each possible node. Here, we used the proportionate size of each group - 'subject', 'object', 'indirect object', as a parameter to find the probability. For example, if the proportion is given as (0.5, 0.5, 0) then it is expected that the verb is transitive. On the other hand, a proportion of (0.75, 0.25, 0) does not clearly identify a certain group and thus we need a method to associate a verb with more than one group.[5]

4. We then formed layer 2 (as shown in the Figure 2) – connecting the verbs with the edges that have the probabilities as their weights. As discussed, layer 2 of the language network comprises of the 'combinatorial' nodes which are connected to the verb lemma nodes from layer 1 of the network. The combinatorial nodes store the argument structure as well as word order property of its connected nodes. The probabilities on the edges connecting these nodes to the lemmas denotes the weights of these connections. Considering Figure 2, the probabilities of connections of "buy:VERB" (in Layer 1) with "SV" (0.016), "SVO" (0.89) and "SIVO" (0.076), shows that "buy" predominately follows "SVO". The sample network shown in figure 2 shows that the language that this network represents is a "SVO" language.

## 3 Experiments

The experiments discussed in this section assume that a large probability is related with a strong connection and more likelihood that the connected nodes show the 'combinatorial' property encapsulated by the concerned Layer 2 node. Further, in order to do the network analysis, we'd used the sentential distance only as a weight to the edges. For the connections between Layer 1 and Layer 2, we used the inverse of the

[3]Specifically, the 48 nodes are SV, VS, OV, VO, IV, VI, SVO, SOV, VSO, VOS, OVS, OSV, ISV, IVS, VSI, VIS, SIV, SVI, IOV, IVO, VOI, VIO, OVI, OIV, SIOV, SIVO, SOIV, SOVI, SVOI, SVIO, IOVS, IOSV, IVSO, IVOS, ISVO, ISOV, VOSI, VOIS, VISO, VIOS, VSIO, VSOI, OVIS, OVSI, OSIV, OSVI, OIVS, OISV

[4]In a way, capturing the dominant word order pattern of a verb which we are really interested in.

[5]We considered the proportions of subject, object, indirect object as a vector in 3D space. We have a pre-defined set of proportions (or classes) which correspond to the layer 2 nodes – (1,0,0): SV/VS, (0,1,0): VO/OV, (0.5, 0.5, 0): transitive of any order and so on. Since these nodes or target vectors are not distributed uniformly in terms of distance, we used the angular distance of the corresponding unit vectors as a measure to calculate probabilities (after proper normalization). This method allowed us to remove any bias for an input proportionate vector vis-à-vis a particular layer 2 node.

probabilities as the edge weights so that the range is from $[1, \infty]$. All network analysis was performed using Cytoscape Shannon et al. (2003). In particular, Cytoscape provides a tool named Network Analyzer which was used to analyse the network with various parameters[6]. All analysis reported below has been done on the network parameters corresponding to the nodes in layer 2.

In the first experiment, we simply map the GUs onto a language network to see if a particular node property (percentage of outgoing arc) leads to the desired classification across languages. In the second experiment, we automatically derive certain implicational universal stated by Greenberg (1963). For example, we see which word order node (e.g., SVO, SOV, etc) best classifies the order of adposition and noun phrase.

### 3.1 Experiment 1

In order to map the Greenbergian universals wrt certain linguistic orders onto the network, we reduced the problem to only probing the node parameters of the layer 2 'combinatorial' nodes. This was done because we are interested in word order generalizations correlated to the order related to the verb. In particular, we looked at each of the word-order based Greenbergian universal and translated them to a particular network parameter of various combinatorial nodes in layer 2. The orders SOV, SVO, VSO etc. are believed to be encoded in the parameter 'Outperc' of the Layer 2 nodes. Outperc is defined as the out-degree of the concerned node divided by the total no. of nodes in layer 2. A language is deemed to be SOV if the SOV node's 'Outperc' is high relative to other nodes in layer 2. We investigate if the distribution of 'Outperc' across all language networks leads to the correct language typology clusters. This experiment is intended as a supervised way of identifying language typology clusters based on Greenberg's word order universals. The data available in WALS (Dryer and Haspelmath, 2013) was used to get the word order patterns related to the Greenbergian universals for various language.

#### 3.1.1 Results

Two network parameters, namely 'Outperc' and 'Outdegree' were used for analysis. As stated above, 'Outperc' is the fraction of verbs connected to a particular combinatorial node. 'Outdegree' is the number of verbs connected to a particular class. The results for various universals are given below

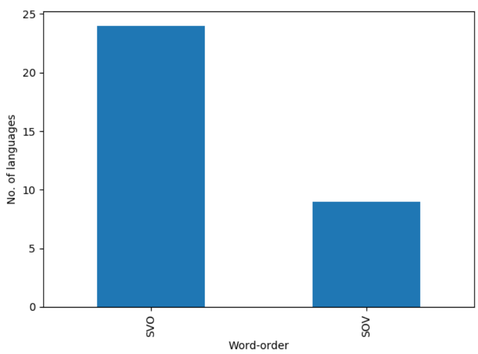


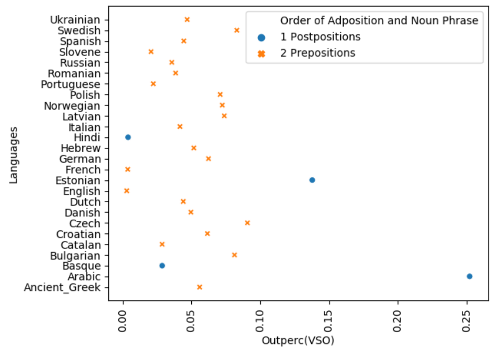

Figure 3: Dominant subject and object order across all language networks.

1. Universal 1 - *"In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object."*

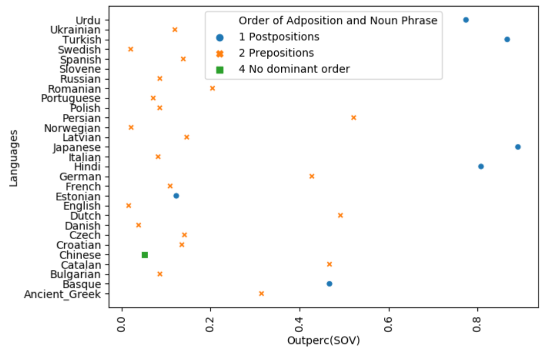
   Since we used only finite verb forms in the second layer, the properties shown in the third layer are expected to be of a general rule for declarative sentences. Figure 3 shows the histogram of the maximum 'Outperc' over all 34 languages.

   Results show that for all the languages, 'Outperc' is maximum for either SOV or SVO nodes. Thus, verifying the universal using the networks used in this analysis.

[6] These were, *In-degree, Out-degree, Outperc, Edge Count, Average shorted path length, Betweenness centrality, Closeness centrality, Closeness centrality, Clustering coefficient, Neighborhood connectivity, Eccentricity*. For details on these parameters, see Newman (2010). Also see: https://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/index.html#complex

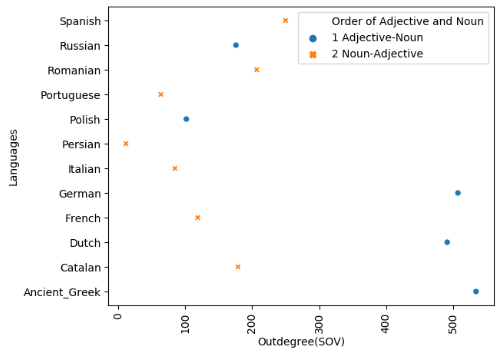



(a)

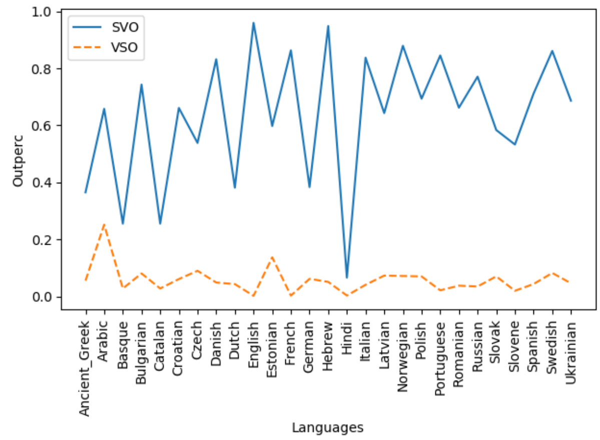



(b)

Figure 4: (a) Outperc for VSO node across all languages and corresponding typology clusters based on postpositional vs prepositional languages. (b) Outperc for SOV node across all languages and corresponding typology clusters based on the order of adposition and noun phrase.




(a)




(b)

Figure 5: (a) Outdegree for SOV node across all languages (with genitive following nouns) and corresponding typology clusters based on order of adjective and noun. (b) Outperc values across various languages for VSO and SVO nodes.

2. Universal 3 - *"Languages with dominant VSO order are always prepositional."*

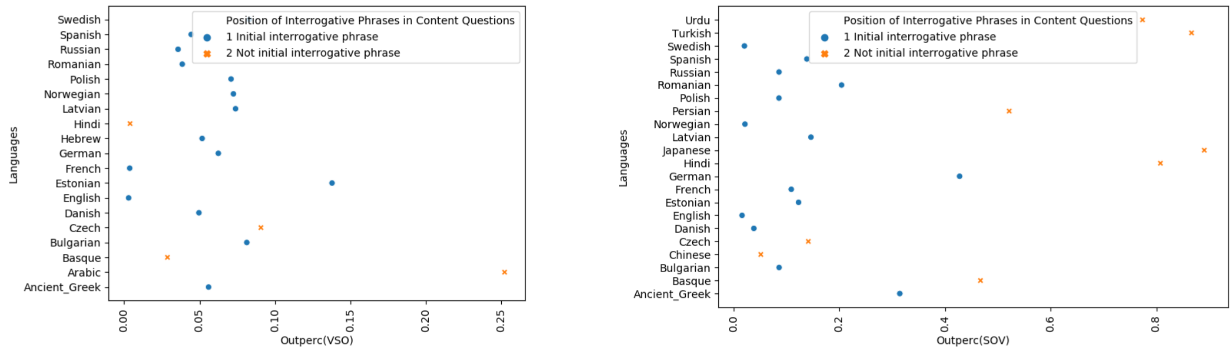
   The feature "85A Order of Adposition and Noun Phrase" in WALS was used to get the information on languages with prepositional vs post-positional. As mentioned above, none of the language networks have a dominant VSO order. Nevertheless, we went ahead to form the clusters using the 'Outperc' of the VSO nodes. The clustering is shown in figure 4a.

   Results show that a higher "VSO outperc" corresponds to post-positional feature. We conclude that our network is not able to induce this universal in its strong form. One reason for this could be that the none of the treebank data for the languages used (including Arabic) had a dominant VSO order for finite verbs.

3. Universal 4 - *"With overwhelmingly greater than chance frequency, languages with normal SOV order are post-positional."*

   Similar to the previous approach, we looked at the 'Outperc' of the SOV nodes in various language networks and looked at the resultant clustering. Figure 4b shows the clusters.

   Resuls show a a clear classification of languages with postpositions vs prepositions. We see that higher values of 'Outperc' for SOV nodes correspond to postpositional languages.




(a) (b)

Figure 6: Outperc for VSO node across all languages and corresponding typology clusters based on the order of interrogative phrases.

4. Universal 5 - *"If a language has dominant SOV order and the genitive follows the governing noun, then the adjective likewise follows the noun."*

   We looked at the languages where genitive follows the noun using the WALS data and then, made the clusters of SOV node's distribution based on the adjective-noun order. This is shown in figure 5a.

   Results show that while 'Outperc' for the SOV nodes was not able to cluster the languages, a related node parameter "Outdegree" performed better. Thus providing some support for the universal from the networks.

5. Universal 6 - *"All languages with dominant VSO order have SVO as an alternative or as the only alternative basic order."*

   Since we didn't have any language with dominant VSO order, we show a comparative plot of 'Outperc' of SVO and VSO across languages in figure 5b.

   A correlation analysis suggests that, other than certain languages, over all, the $R^2$ came out to be just 0.07, suggesting that the networks are unable to capture this generalization.

6. Universal 12 - *"If a language has dominant order VSO in declarative sentences, it always puts interrogative words or phrases first in interrogative word questions; if it has dominant order SOV in declarative sentences, there is never such an invariant rule."*

   We used the relevant feature in WALS data to plot the 'Outperc' of VSO and SOV for the languages obtained from WALS. This is shown in figure 6

   Results show that increase in the VSO 'Outperc' does not lead to the right typology cluster. Interestingly, the 'Outperc' for SOV nodes for different languages gave better results. Thus providing partial support for the universal from the networks.

To summarize, the result show that the language typology related to (a) order of subject-object across languages, (b) presence of prepositions in SOV languages, (c) order of adjectives and noun vis-à-vis order of gentive and noun in SOV languages, and (d) position of interrogative word in VSO/SOV language, can be derived from the 'Outperc'/'Outdegree' parameter of the layer 2 nodes in various language networks.

### 3.2 Experiment 2

Experiment 1 targetted a specific universal and mapped it on to the network using a prespecified node property (Outperc/Outdgree of SOV, VSO, SVO layer 2 nodes). In experiment 2, we asked a more general question – which node parameter in different language networks leads to the best language typology classification based on Greenberg's universals? The linguistic orders that we looked at were taken from WALS Dryer and Haspelmath (2013)]; these were, (a) Order of subject, verb and object, (b) Order of Adposition

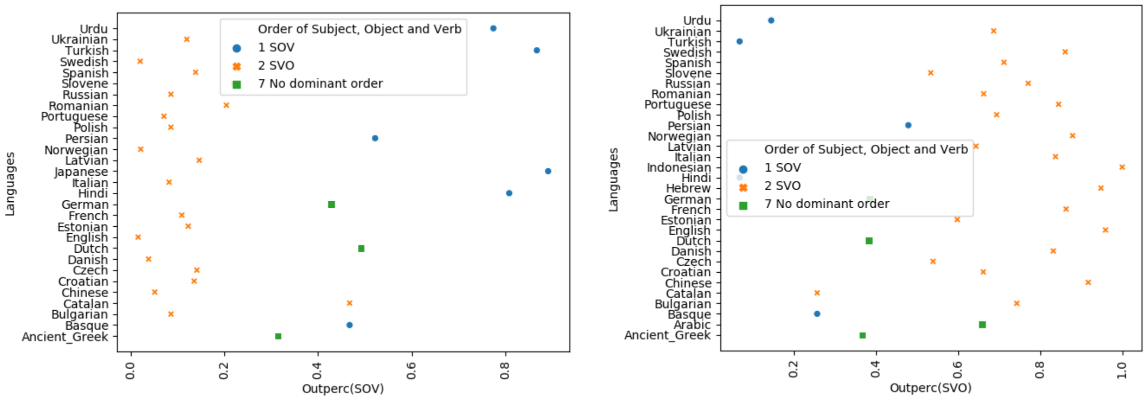



Figure 7: Top two language clusters wrt the order of subject, object and verb. The Outperc parameter for SOV nodes across all languages lead to the best distinction between SOV vs SVO languages.

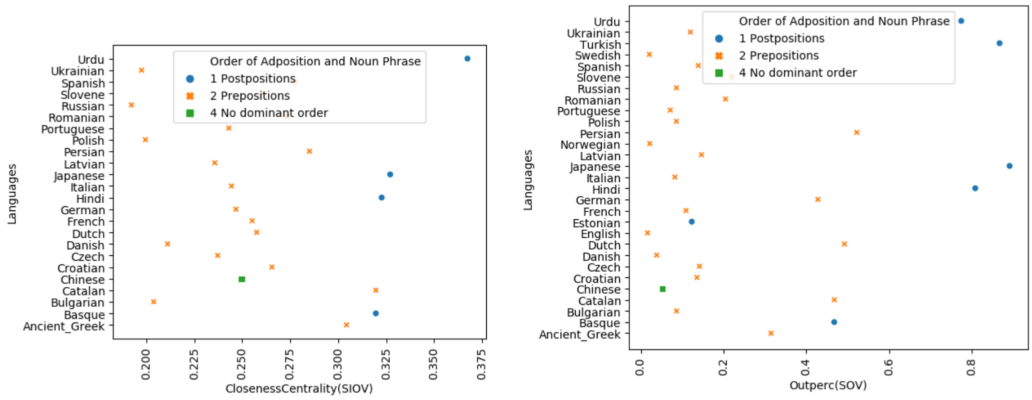



Figure 8: Manually identified language clusters wrt the order of adposition and noun phrases. The Outperc parameter for SOV nodes across all languages lead to a good distinction between language where the adposition follows the noun phrase vs those where it precedes the noun phrase.

and Noun Phrase, (c) Order of Adjective and Noun, and (d) Position of Interrogative Phrase and Content Questions.

We investigate the various parameters[7] for each node in layer 2 to see which node-parameter combinations across all the languages lead to the best language classification for a particular word order. For example, consider "Order of Adposition and Noun Phrase". In order to find which parameter of which layer 2 node can lead to the best classification of languages based on this order, we get a particular node-parameter values from all language networks, and check if this distribution leads to the correct classification of languages as given in the WALS data. The correlation between the node-parameter values and the correct language cluster (which is already known) is quantified by silhouette value Rousseeuw (1987). This silhouette value is obtained for all the (nodes $\times$ parameters) node-parameter combinations and the highest score gives us the node-parameter that classifies the languages best based on the word order under consideration. A greater silhouette value corresponds to better clustering. Intuitively, the silhouette value captures the cohesiveness of the data point with its cluster.

To summarize, experiment 2 discusses a method to induce the linguistic orders by probing all possible parameters for each verb-order nodes that are contained in layer 2.

[7]These were, *In-degree, Out-degree, Outperc, Edge Count, Average shorted path length, Betweenness centrality, Closeness centrality, Closeness centrality, Clustering coefficient, Neighborhood connectivity, Eccentricity*. For details on these parameters, see Newman (2010). Also see: https://med.bioinf.mpi-inf.mpg.de/netanalyzer/help/2.7/index.html#complex

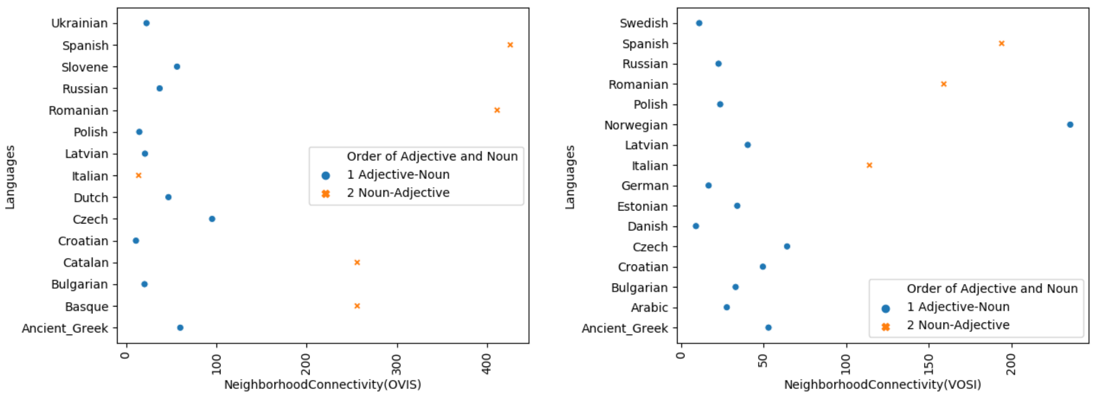



Figure 9: Top two language clusters wrt the order of adjective and noun.

### 3.2.1 Results

Below we discuss the results for the various word order patterns. The top two clusters based on silhouette values are shown for each pattern.

1. Order of Subject, verb and object:

   Results show that 'Outperc' of the SOV node clusters the languages much better than 'Outperc' of the VSO node (see figure 7). Results also suggest that 'Outperc' outperforms all other node parameter. Recall that 'Outperc' is the percentage of outgoing edges from a node. This means that, as far as the current set of languages is considered, the 'Outperc' property of the 'SOV' node can alone be effectively used to decide the word order of the language. This suggests that while there is a lot of variability wrt SVO order in various languages compared to SOV order.

2. Order of Adposition and Noun Phrase :

   The top silhouette scores for various parameter-node pairs did not lead to a good cluster of languages based on this feature. This is not to say that the appropriate clustering cannot be derived from the cluster. Indeed, a manual analysis of the various clusters shows that the 'ClosenessCentrality' parameter of SIOV nodes across all the languages does lead to good language clusters for this feature. In addition, 'Outperc' of the SOV nodes lead to good clusters (see figure 8). 'ClosenessCentrality' gives us a measure of how close the node in question is to the other nodes in the network. Given this definition, it is difficult to see why such a parameter should lead to the correct clustering. Interpreting the results on the other parameter, namely, 'Outperc' for SOV nodes is easier. It shows that the order of subject, object and verb can predict the order to adpositition and noun phrase as was hypothesized by Greenberg.

3. Order of Adjective and Noun:

   The two clusters based on silhouette scores show that 'Neighborhood Connectivity' of OVIS and VOSI nodes for various languages were able to cluster the languages really well (see figure 9). 'Neighborhood Connectivity' corresponds to the average connectivity of its neighbours. While the result does give us the desired clusters, it is difficult to interpret the linguistic validity of the parameter.

4. Position of Interrogative Phrase in Content Questions :

   Finally, for the cluster based on position of interrogative phrase the silhouette scores for the cluster based on "Outperc" parameter for the SOV nodes gave one of the best results (see figure 10).

## 4 Discussion and Conclusion

Our work shows that that word order generalisations are encoded in a network and can be automatically derived from it. In particular, the results from experiment 1 showed that when the subject-object-verb orders

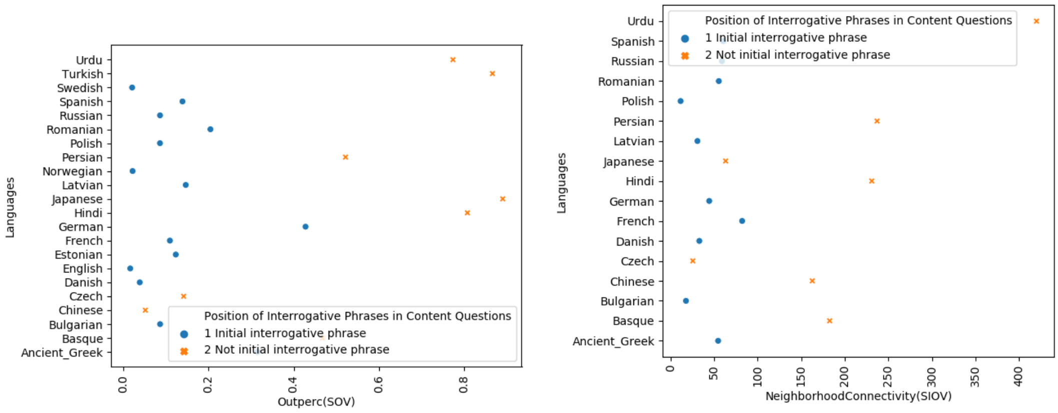



Figure 10: Top two language clusters wrt the position of interrogative phrase in content questions.

| WALS | Network Parameter 1 | | | Network Parameter 2 | | | Network Parameter 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| feature | Node | Parameter | Silhouette | Parameter | Parameter | Silhouette | Node | Parameter | Silhouette |
| 81A | SOV | Outperc | 0.53 | SVO | Outperc | 0.304 | OSV | Outperc | 0.3 |
| 85A | *SIOV* | *Closeness C* | *-0.25* | *SOV* | *Outperc* | *-0.25* | - | - | - |
| 87A | OVIS | Neighborhood C | 0.63 | VOSI | Neighborhood C | 0.58 | OISV | Eccentricty | 0.34 |
| 93A | SOV | Outperc | 0.48 | VOIS | Neighborhood C | 0.604 | SIOV | Neighborhood C | 0.466 |

Table 1: Top three silhouette score for the clusters related to the 4 word order patterns. 81A: Order of subject, verb and object; 85A: Order of adposition and noun phrase; 87A: Order of adjective and noun; 93A: position of interrogative phrases in content question. Note that the results for 85A are based on manual evaluation as the top silhouette scores failed to give the correct clusters. Closeness C = Closeness Centrality; Neighbourhood C = Neighbourhood connectivity.

found in the Greenbergian universals are probed through the combinatorial nodes, the correct word order typologies could be found. In addition, experiment 2 showed that similar (combinatorial) node-parameters lead to the right language clusters. We found that simply by inducing verb order and using the appropriate parameters, we can derive other linguistic order which share implicational relations with the verb order. These results are in accordance with the claim that networks are a meaningful representation of a linguistic knowledge. The nodes which led to the best classification based on a particular feature were major word orders, e.g., SOV, SVO, SVIO, etc. It is interesting to notice that including the order of indirect object induced certain linguistic features in Layer 2.

Our analysis was affected by multiple factors such as the treebank size, alignment of languages in UD and WALS, etc. For example, silhouette score is higher when clusters are dense and well-separated. Since the cluster sizes are non-uniform, so is the density of clusters which is a function of the number of points in a cluster. The number of points in a cluster follows a power law, which is the primary reason for the non-uniformity in the cluster sizes. The analysis in experiment 1 failed to induce any VSO-order based universal since no language considered has a dominant VSO order in the respective treebank. Similarly, while the 'Outperc' parameter that encodes the combinatorial property of the nodes in layer 2 was quite effective in classifying languages, in some cases where there is no dominant verb order pattern, 'Out-Degree' helps. While both 'Outperc' and 'Out-Degree' are very easy to interpret, other parameters such as 'Eccentricity', 'NeighborhoodConnectivity', that also lead to good clusters, are less transparent in their interpretability vis-á-vis linguistic generalizations.

Indeed, the fact that the language network in this work lends itself to interpretability is a very attractive feature of this approach. Since, the network's properties and the representation is tractable, we can investigate the linguistic validity of various parameters. While the current work has shown some promise wrt capturing simple word order generalizations, it remains to be seen if such a representation can capture other complex linguistic constraints such as agreement, filler-gap dependencies, islands, etc. (cf. Linzen et al., 2016; Futrell and Levy, 2019).

## References

Željko Agić, Maria Jesus Aranzabe, Aitziber Atutxa, Cristina Bosco, Jinho Choi, Marie-Catherine de Marneffe, Timothy Dozat, Richárd Farkas, Jennifer Foster, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Yoav Goldberg, Jan Hajič, Anders Trærup Johannsen, Jenna Kanerva, Juha Kuokkala, Veronika Laippala, Alessandro Lenci, Krister Lindén, Nikola Ljubešić, Teresa Lynn, Christopher Manning, Héctor Alonso Martínez, Ryan McDonald, Anna Missilä, Simonetta Montemagni, Joakim Nivre, Hanna Nurmi, Petya Osenova, Slav Petrov, Jussi Piitulainen, Barbara Plank, Prokopis Prokopidis, Sampo Pyysalo, Wolfgang Seeker, Mojgan Seraji, Natalia Silveira, Maria Simi, Kiril Simov, Aaron Smith, Reut Tsarfaty, Veronika Vincze, and Daniel Zeman. 2015. http://hdl.handle.net/11234/LRT-1478 Universal dependencies 1.1. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague.

Réka Albert, Hawoong Jeong, and Albert-László Barabási. 2000. Error and attack tolerance of complex networks. *nature*, 406(6794):378.

Paolo Allegrini, Paolo Grigolini, and Luigi Palatella. 2004. Intermittency and scale-free networks: a dynamical model for human language complexity. *Chaos, Solitons & Fractals*, 20(1):95–105.

Albert-László Barabási. 2011. The network takeover. *Nature Physics*, 8(1):14.

Javier Borge-Holthoefer and Alex Arenas. 2010. Semantic networks: Structure and dynamics. *Entropy*, 12(5):1264–1302.

Guido Caldarelli. 2007. *Scale-free networks: complex webs in nature and technology*. Oxford University Press.

Ramon Ferrer i Cancho, Andrea Capocci, and Guido Caldarelli. 2007. Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, 17(07):2453–2463.

Ramon Ferrer i Cancho, Ricard V. Solé, and Reinhard Köhler. 2004. Patterns in syntactic dependency networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 69 5 Pt 1:051915.

Monojit Choudhury, Diptesh Chatterjee, and Animesh Mukherjee. 2010. Global topology of word co-occurrence networks: Beyond the two-regime power-law. In *Proceedings of the 23rd international conference on computational linguistics: Posters*, pages 162–170. Association for Computational Linguistics.

Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social communication*, 1:343–359.

Allan M. Collins and Elizabeth Loftus. 1975. https://doi.org/10.1037//0033-295X.82.6.407 A spreading activation theory of semantic processing. *Psychological Review*, 82:407–428.

Luciano da Fontoura Costa, Osvaldo N Oliveira Jr, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha. 2011. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412.

Matthew S. Dryer and Martin Haspelmath, editors. 2013. https://wals.info/ *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Ramon Ferrer I Cancho and Richard V Solé. 2001. The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482):2261–2265.

Richard Futrell and Roger P. Levy. 2019. https://doi.org/10.7275/jb34-9986 Do RNNs learn human-like abstract word order preferences? In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, pages 50–59.

Joseph H Greenberg. 1963. Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2:73–113.

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

Alan J Lerner, Paula K Ogrocki, and Peter J Thomas. 2009. Network graph analysis of category fluency testing. *Cognitive and Behavioral Neurology*, 22(1):45–52.

Tal Linzen, Emmanuel Dupoux, and Yoav Goldberg. 2016. Assessing the ability of lstms to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, 4:521–535.

Haitao Liu and Wenwen Li. 2010. Language clusters based on linguistic complex networks. *Chinese Science Bulletin*, 55(30):3458–3465.

Ken McRae and Stephen Boisvert. 1998. https://doi.org/10.1037/0278-7393.24.3.558 Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24:558–572.

Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris Mcnorgan. 2005. https://doi.org/10.3758/BF03192726 Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4):547–559.

Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013. http://arxiv.org/abs/1301.3781 Efficient estimation of word representations in vector space.

George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41.

Glyn Morrill. 2000. Incremental processing and acceptability. *Computational linguistics*, 26(3):319–338.

Mark Newman. 2010. *Networks: an introduction*. Oxford university press.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan T McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *LREC*.

Beth A. Ober and Greg Shenaut. 2006. https://doi.org/10.1016/B978-012369374-7/50012-2 Semantic memory. *Handbook of Psycholinguistics*, pages 403–453.

Romualdo Pastor-Satorras and Alessandro Vespignani. 2007. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press.

M. J. Pickering and V. S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin*, 134(3):427–459.

Martin J. Pickering and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39:633–651.

Ardi Roelofs. 1992. A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42(1-3):107–142.

Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65.

P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, Schwikowski B., and T. Ideker. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504.

Cynthia S Q Siew, Dirk U Wulff, Nicole Beckage, and Yoed Kenett. 2018. https://doi.org/10.31234/osf.io/eu9tr Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics.

Ricard V Solé, Bernat Corominas-Murtra, Sergi Valverde, and Luc Steels. 2010. Language networks: Their structure, function, and evolution. *Complexity*, 15(6):20–26.

K. M. Tooley and M. J. Traxler. 2010. Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, 4(10):925–937.

Michael S Vitevitch. 2008. What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research*.

Michael S Vitevitch, Gunes Ercal, and Bhargav Adagarla. 2011. Simulating retrieval from a highly clustered network: Implications for spoken word recognition. *Frontiers in psychology*, 2:369.