Linguistic Features Cluster in a Social Network\textsuperscript{1}

Peter W. Culicover, Andrzej Nowak and Wojtech Borkowski
The Ohio State University and University of Warsaw

1. Introduction

The close study of the grammars of natural languages that has been pursued over the past thirty-odd years has revealed that many if not most of the logically possible languages do not exist. Linguistic theories have sought to account for some of the gaps by appealing to the structure of the language faculty in the human mind. On such an approach, particular logically possible properties and correlations of properties are unattested because they are not possible configurations of the universal mental apparatus that embodies the capacity for language.\textsuperscript{2} The examination of which of the logical possibilities actually occur thus promises to illuminate the structure of the language faculty in a fundamental way.

We do not challenge the basic thrust of this approach. We will suggest here, however, that the absence of particular combinations of properties may in principle reveal not the just structure of the language faculty in human beings, but also the structure of social networks of human beings in which linguistic knowledge is communicated. The behavior of such social networks makes it virtually inevitable that some combinations of properties will not appear. The combinations that do and do not appear are not necessarily predictable in terms of the linguistic

\textsuperscript{1} For stimulating discussion and helpful comments regarding this work we are indebted to Sarah Thomason, Michael Broe, Brian Joseph, Greg Carlson, Elissa Newport, Jack Hawkins and audiences at UC San Diego, the University of Rochester, the University of Illinois, Michigan State University, the University of Stuttgart, and Keio University. Naturally, we are responsible for any errors.

\textsuperscript{2} Hence they are biologically impossible. Their non-existence is explained in much the same way as is the non-existence of flying pigs or talking dogs.
content of the properties in question, although computational complexity and generalization may play a role in some cases.

We explore here some of the ways in which the pattern of language distribution is a consequence of the properties of the social network in which language is embedded. In previous work, Nettle (1999) has shown through a computational simulation that the spatial distribution of languages can be attributed to interactions among speakers in the social network, assuming the Dynamic Theory of Social Impact (Nowak, et al. (1990)). We replicate his results here, and extend the approach to the clustering and correlation of linguistic features.

We report here on three main results. First, gaps in the set of actual languages (that is, languages that are possible but do not exist) may arise very naturally as a consequence of transmission of language properties in the social network. Second, the network produces weak correlations between properties that give the appearance of being tendencies, rather than universals. Third, contact between languages with different properties may produce languages with novel combinations of properties.

We apply these results to actual linguistic phenomena. Our argument is that part of the burden of linguistic explanation may be borne by social, as contrasted with cognitive mechanisms. We conclude by using our observations to sharpen the characterization of what aspects of linguistic change plausibly are the province of cognitive mechanisms.

2. The simulation model

Here we briefly summarize the technical details of the simulation model. The social network is represented by a square matrix of $k$ rows and $k$ columns (the default value is 50).
Linguistic Features Cluster in a Social Network

Each of $k^2$ locations can be occupied by an individual, so if all the locations are occupied there are $N = k^2$ individuals. Each individual is characterized in addition to the location in the matrix by three language parameters (or features) and additionally by the strength parameter. The strength parameter describes how influential each individual is in inducing change in other individuals.

Each language parameter can have one of a discrete set of values. In the simplest case the value is a binary variable with values of 0 and 1. The language parameters are independent. That is, there is no relationship between the values of different parameters.

The simulation proceeds according to the so called Monte Carlo dynamic. An individual (agent) is randomly selected from the matrix. For the chosen individual (agent) a specified number ($p$) of interaction partners are drawn randomly from a distance not greater than $d$ cells from the agent, where $d$ is chosen for the particular simulation. If the limiting distance $d$ is equal to one, then only cells adjacent to the cell occupied by the agent can be selected as the interaction partners. The selection of interaction partners is independent for each agent so it is possible that a specific partner can be selected more than once.

For each of the language parameters (independently of the two other parameters) the following procedure is repeated: for each value of the parameter the social pressure to adopt this parameter is computed by adding the strength of all the interaction partners (that were drawn in this round of simulation).

Another option of the simulation model is whether or not there is self influence. If self influence is chosen, the agent is treated as one of the interaction partners and the strength of the agent is added to the value of influences of the appropriate value of the parameter.
Linguistic Features Cluster in a Social Network

It is also possible to introduce noise into the simulation. If the value of noise is greater than 0 then the noise is added to the sum of the pressures from the interaction partners.\textsuperscript{3}

In sampling the values of its interaction partners, the agent adopts the value of the parameter that received the highest score. If two or more values have the same score then the value is chosen randomly between them. After the value of all the three parameters have been updated for a given agent (the interaction partners are the same for all the three parameters of an agent), a new agent is randomly drawn. A simulation step corresponds to choosing N agents, so approximately every individual is updated once in a simulation step. Since the random drawings are independent of each other, some individuals may not be chosen at a given step and others are not drawn at all.

This simulation model is applicable to a wide range of socially transmitted attributions. In the computational simulation that we report on here, we interpret the simulation as modelling the distribution of logically independent linguistic properties in a social network over time. The network consists of interactions and social relationships between individuals. There are linguistic communications across this network that pass between individuals. A language learner acquires knowledge of language by interacting with (a subportion of) this network, and by acquiring the linguistic features of the individuals with which it interacts, with some suitable weighting based on proximity, frequency of contact, social mobility, and so on.

Crucially, the Social Impact Theory that underlies this model presumes that the content of the parameters has nothing to do with the particular configuration of features displayed by a given agent or group of agents at any point in time. Agents change their

\textsuperscript{3} The noise is calculated as random number drawn from a flat distribution (0,1) multiplied by the maximum strength of the individuals (to normalize it to the strength of individuals in the simulation).
feature values strictly as a function of the feature values of the agents that they interact with. While this presumption may be controversial when the feature values are social attitudes, it is fully consistent with prevailing views on how knowledge of language is transmitted from a community to its children. In brief, the children acquire the prevailing linguistic features of the community. The simulation model derives its interest in the case of language change from the fact that the interactions between the learners and the language community are complex, and the outcomes by no means self-evident. We take the view, in fact, that the historical development of languages, and \textit{a fortiori} much of their synchronic character, can only be understood in terms of the social interactions.\footnote{We thus agree with the position taken by Thomason & Kaufman (1988:36) citing Heath (1978:71) citing Coteanu (1957:147): “Selon nous, cette question ne dépend pas du caractère de la structure grammaticale des langues en contact, mais d’une série de facteurs de nature sociale.” and “...the history of a language is a function of the history of its speakers, and not an independent phenomenon that can be thoroughly studied without reference to the social context in which it is embedded.”}

3. Gaps

3.1. How gaps arise

We will suppose for the sake of the simulation that the class of possible grammars of natural languages can be characterized entirely in terms of values of features.\footnote{In fact this must be true in a trivial sense; see Culicover (1999) for discussion.} A prevalent view in current linguistic theory is that most if not all of the most theoretically interesting aspects of language variation, language change and language acquisition can be accounted for in terms of a small set of binary features, called ‘parameters’. For our purposes, however, it is sufficient to assume that whatever the features are, however many there are, and whatever values they have, learners are influenced to adopt the values of the community through social interaction. We are interested here in understanding what aspects of linguistic
change, variation and acquisition can be attributed entirely to the social interaction. Hence, crucially, the features in our simulation is entirely without linguistic content. In practice, in order to simplify computations and presentations, we will assume that features are two-valued.

Our simulation supposes that there are three two-valued features, which define eight distinct languages. Gaps occur when certain feature combinations are not attested. Our simulation shows that gaps may arise over the course of time, as the values of two of the features become strongly correlated. To take a simple example, suppose that there are two two-valued features that define four languages:

\[ [+F1,+F2] \quad [-F1,+F2] \quad [+F1,-F2] \quad [-F1,-F2] \]

If the geographical distribution of \(-F2\) becomes sufficiently restricted, it may fail to overlap with \([+F1]\). In such case, one of the languages, namely \([+F1,-F2]\), will cease to exist. Such a situation may occur simply as a consequence of the social structure, and in itself tells us nothing interesting about the relationship between \([+F1]\) and \([-F2]\).

Figure 1 shows the random distribution of feature values for three features in a population of 2500 (=50x50). The upper lefthand image shows the distinct languages as differences on the gray scale. The other images show the distribution of + and – values for the three features FIRSTs, SECONDS and THIRDs.
The population of each of the eight languages is shown in the histogram in Figure 2. As can be seen, the languages are distributed more or less evenly over the entire population, as would be expected from a randomized assignment of feature values.
Figure 2. Population of the eight languages

We have omitted intermediate steps in the simulation for reasons of space.

After 69 steps the distribution of languages and features is as in Figure 3.
Figure 3. Distribution of languages and features after 150 steps

The histogram in Figure 4 shows the population levels of the eight languages.
The loss of languages illustrated in this particular instance of the simulation is not unique. It is a consequence of the particular assumptions made in the simulation about how individuals interact in the network. Running the same simulation under the same parameters yields a
Linguistic Features Cluster in a Social Network

different pattern of features and languages each time, but the results are the same. We this
simulation 100 times. The following chart shows the number of times a given number of
languages remained in the simulation after 200 steps.

As can be seen, in 50 of the 100 runs of the simulation there were eight languages after
200 steps. But in 32 runs there were 7 languages, in 10 runs there were 6 languages, and so on.
So while the precise number of languages that will remain after a certain number of steps is not

![Number of languages at step 1000](chart)

**Figure 6. Distribution of languages after 1000 steps**

predictable, it is clear that gaps in the set of languages can and will arise over the course of time
as a consequence of the interaction in the network. The chart in Figure 6 shows that over a longer
time span the number of languages for the same simulation tends to decline.

3.2. **Simulation parameters**

The preceding simulation makes the following specific assumptions regarding the context
of interaction among individuals:
Linguistic Features Cluster in a Social Network

(i) individuals interact only with immediate neighbors; that is, “interaction distance” is 1;

(ii) individuals are influenced by only two interaction partners at any point in time;

(iii) individuals are highly resistant to change, in that their current state is part of the input to their next state;

(iv) there is no noise, that is, no uncategorized influence on each agent’s state.

We can compute the tradeoffs in the simulation model between interaction distance and interaction partners. The number of languages that persist as individuals interact declines dramatically as the interaction distance increases, particularly from 1 to 2. When the interaction distance is small, the number of languages declines slightly as the number of interaction partners increases, but when the interaction distance is larger, the number of languages grows as the number of partners increases. The following chart illustrates the interactions graphically, with the number of languages remaining after 250 steps of the simulation mapped against the interaction distance and the number of partners.
Another measure that corresponds to the degree of organization in the system is that stress, which measures the amount of clustering. (The lower the stress the greater the amount of clustering.) The dependence of stress on the interaction distance and number of interaction partners is shown in Table 1 and in Figure 8.

<table>
<thead>
<tr>
<th>Interaction Distance</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.098383</td>
<td>0.082917</td>
<td>0.081817</td>
<td>0.073533</td>
<td>0.071072</td>
<td>0.074978</td>
<td>0.075783</td>
</tr>
<tr>
<td>2</td>
<td>0.050017</td>
<td>0.038994</td>
<td>0.033922</td>
<td>0.031906</td>
<td>0.029450</td>
<td>0.028461</td>
<td>0.021694</td>
</tr>
<tr>
<td>3</td>
<td>0.022939</td>
<td>0.027572</td>
<td>0.014550</td>
<td>0.024633</td>
<td>0.015189</td>
<td>0.018594</td>
<td>0.022361</td>
</tr>
<tr>
<td>4</td>
<td>0.017228</td>
<td>0.007783</td>
<td>0.008450</td>
<td>0.007178</td>
<td>0.006272</td>
<td>0.012672</td>
<td>0.013750</td>
</tr>
</tbody>
</table>

Table 1. Mean level of stress after 200 steps
As can be seen, stress tends to go down as the interaction distance increases. It is relatively insensitive to the number of interaction partners, although we do see a slight increase in stress when the interaction distance is high and there are many partners.

### 3.3. A case study: West Germanic verb clusters

The simulation suggests that the actual possibilities realized in linguistic variation will in general display a subset of the logical possibilities, independent of grammatical considerations. We have seen that feature values that are in a minority relative to other feature values tend to die out in non-equilibrium situations. The reason for this is that the size of the population that supports a particular feature value causes a learner to be exposed to a greater quantity of evidence for that feature value. Hence the learner tends to adopt the majority position, other things being equal.
The question of why it is that a particular feature value is in a minority is an independent question, one that certainly can have a linguistic as well as a sociological basis. Nettle (1999) shows that biases against certain linguistic properties can have such an effect when coupled with the differential effects of social status. Nettle characterizes inherent bias as follows (112): “they are hard to acquire, which causes languages to change away from them.” In the context of current linguistic theorizing, this idea is a powerful one. It suggests that non-existing linguistic properties are a consequence first of the social structure, and second of learnability, in some sense related to complexity of computation. In this section we explore the application of this basic perspective to the details of a particular linguistic phenomenon, and show how the distribution of possibilities can be plausibly attributed to the operation of the social network on the relative computational complexity of the various possibilities.

The linguistically interesting question here is what constitutes computational complexity in the domain of grammar and language acquisition. We hypothesize that to the extent that there is internal structure within a given grammatical domain, those logical possibilities whose structure is more transparent, in the sense of correspondences with conceptual structure (CS), will be more highly favored.\(^\text{6}\) Those possibilities that are minimally transparent will be rare if they exist at all, simply for this reason. When they exist they will have the character of one-of-a-kind constructions, i.e. idioms.\(^\text{7}\) But given this, there will also be missing possibilities due to interactions within the social network. The role of abstract grammatical interactions independent of the complexity of the syntax-CS correspondences may be minimized or eliminated entirely.

As a concrete example we consider the order of verbs in Continental West Germanic verb clusters, as described by Zwart (1994). In cases where there is an auxiliary verb and an infinitival verb, there are two logical possibilities, \(V_1 - V_2\) and \(V_2 - V_1\), which we notate 1-2 and 2-1, respectively, following Zwart. It is customary and in fact quite natural from the perspective of

\(^{6}\) For discussion, see Culicover (1999).
\(^{7}\) A typical example of an opaque idiom is \textit{by and large}. Some constructions have a productive and an idiomatic component. For discussion, see Culicover & Jackendoff (1999)
linguistic theory to take the ordering of verbs to be a consequence of the ordering of the auxiliary and the VP whose head is \( V_2 \), and not a primitive element of the grammar.

Matters become somewhat more complex when there are three verbs involved. Suppose for the sake of argument that there is a parameter that specifies the relative order of a verbal head and its verbal complement in a language or dialect. This parameter would have two values, \( V_h - V_c \) and \( V_c - V_h \). Even if there are three or more verbs, we would expect there to be only two possible orderings, one in which each head precedes its complement, and one in which each head follows its complement. However, there are significantly more than two possible orderings when there are more than two verbs in a sequence. Zwart shows that in Standard Dutch both 1-2 and 2-1 are possible orders. The dialects tend to choose one of the two orderings.

For three-verb clusters the general order is 1-2-3 in Standard Dutch. In contrast, it is 3-2-1 in High German and Frisian. But, notes Zwart, “many exceptions exist.” Standard Dutch has the order 1-2-3, 1-3-2 and (marginally) 3-2-1 when \( V_3 \) is a past participle. Frisian allows 3-2-1 and High German 3-2-1 and 1-2-3. In the case of the *infinitivus pro participio* (IPP), where the participial \( V_2 \) is replaced by an infinitival form, the orders 1-2-3 and 1-3-2 are found in Dutch and High German. Dutch also shows 3-1-2. West Flemish apparently has 2-3-1. What apparently never occurs is *2-1-3*, Zwart notes. Finally, when there are four-verb clusters, 4-3-2-1 is possible in Dutch, and the order 4-1-2-3 is found with IPP on \( V_2 \) and \( V_3 \), as is 1-4-2-3 marginally. The following table summarizes the possibilities.

<table>
<thead>
<tr>
<th>Order</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>Standard Dutch; dialects</td>
</tr>
<tr>
<td>2-1</td>
<td>Standard Dutch; dialects</td>
</tr>
<tr>
<td>1-2-3</td>
<td>Standard Dutch; High German</td>
</tr>
<tr>
<td>1-3-2</td>
<td>Standard Dutch; High German</td>
</tr>
<tr>
<td>*2-1-3</td>
<td></td>
</tr>
</tbody>
</table>
Linguistic Features Cluster in a Social Network

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3-1</td>
<td>West Flemish⁸</td>
</tr>
<tr>
<td>3-1-2</td>
<td>Dutch</td>
</tr>
<tr>
<td>3-2-1</td>
<td>Frisian; High German</td>
</tr>
<tr>
<td>4-3-2-1</td>
<td>Dutch</td>
</tr>
<tr>
<td>4-1-2-3</td>
<td>Dutch</td>
</tr>
<tr>
<td>1-4-2-3</td>
<td>Dutch</td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. West Germanic verb clusters

This set of possibilities clearly cannot be subsumed under a single ordering parameter. Of course, it is possible that there is more than one ordering parameter. There could be a parameter for the tensed auxiliary verb and its complement and another parameter for an infinitival auxiliary verb and its complement.⁹ If this were the case, there would be four logical possibilities for the three-verb clusters. But then we would not expect more than four possibilities in cases of four and more verb clusters, unless embedded auxiliary verbs could not be distinguished on the basis of how deeply embedded they are. At some point, however, the notion of “parameter” loses any explanatory force and becomes simply a notational device for keeping track of variation.¹⁰ The situation is made additionally complex if we allow for the possibility that verbs can “reorder” within a canonical ordering, as also suggested by Zwart (1994) (see also Zwart (1995:216)).

There are, in effect, two puzzles here. One is that we do not find just two possible orders, 1-2-3-… and …-3-2-1, or even just one order, as we might expect if there was a single ordering parameter. The other is that we do not find all possible orders, as we might expect if there were distinct ordering parameters for various levels of embedding.

⁸ This order occurs only when V1 is an auxiliary verb (Zwart (1995:220)). Zwart attributes the order to topicalization of the VP of which V2 is the head and V3 the head of the complement.
⁹ Actually, the situation is more complicated than this. Zwart (1995:217, n.7) notes that the 2-1 order is preferred in Dutch when 1 is an auxiliary and 1-2 is preferred when 1 is a modal.
Zwart’s solution is to assume that there is basically a single order, 1-2-3-…, and that there are movements of V or VP to designated positions. These movements are constrained by grammatical principles, so that not all possible reorderings are found. There are two types of movement, on his analysis. First, infinitival verbs adjoin, to the right or the left depending on the language/dialect, to a higher modal verb. Second, a participle moves to the specifier of an auxiliary verb.

Taking the 1-2-3-… order to be basic, it is clear that for a language that is uniformly …-3-2-1, everything must move to the left (on the assumption that each verb is the head of a projection). The theoretical commitments necessary to permit this type of derivation are unattractive from our perspective; see Rochemont and Culicover (1997) for a detailed discussion. Moreover, and more importantly, for this ordering and for the intermediate cases, we find that the more complex structures in which some but not all of the possible reorderings take place must be annotated in a way that is equivalent to an explicit enumeration of what the possible orders are. For example, consider the Standard Dutch orderings 1-2-3, 1-3-2, 3-2-1, and 3-1-2 for IPP. The first is basic, by assumption. The second is derived by movement of 3 to a position to the left of 2. The third is derived by movement of 3-2 to the left of 1. The fourth appears to be a topicalization of 3, the IPP infinitive, a plausible analysis in view of the fact that 4-1-2-3 and 1-4-2-3 are also possibilities with this construction.

Contemporary movement theory (see Chomsky 1995) posits a “feature-checking” mechanism that licenses movement of a constituent to a given location in a structure. Ignoring for the moment IPP, let us call the feature in question here [F] and suppose that the constituent that moves adjoins to the left of [F]. The abstract structure of 1-2-3 will lack [F], that of 1-3-2- will be 1-[F]-2-3, and that of 3-2-1 will be [F]-1-[F]-2-3. Similar remarks hold for clusters of four verbs, where 4-3-2-1 will be derived from [F]-1-[F]-2-[F]-3-4.
In effect, whether two verbs that participate in a head-complement relation will appear in the order $V_h - V_c$ or the opposite is notated by whether or not $[F]$ appears to the left of $V_h$.

There cannot be a uniform parametric account of the word order possibilities in this case, because the individual representations involving $[F]$ are independent of one another in a given language. A uniform parametric account would be one in which either $[F]$ does not appear in a language, or it appears freely in a language. But what we actually find is that the language and dialects under consideration here differ in terms of which verbs appear with $[F]$, not whether or not $[F]$ appears in the language. Even if $[F]$-1-2 ($\Rightarrow$ 2-1) occurs in a language (such as High German), this does not entail that 1-[F]-2-3 ($\Rightarrow$ 1-3-2) will occur. Hence it is not sufficient to learn the parameters, one must learn the cases, or at least, general properties of the linear sequences. In fact, even if the correct characterization of the knowledge does make use of the feature $[F]$, the relevant linear sequences must be correctly identified by a learner in order for the learner to know where the $[F]$ goes in the abstract representation.

Now, if one must learn the cases individually then it is clear that it does not matter for descriptive purposes whether the cases are simply represented in terms of the possible orderings within clusters, or whether they are represented in terms of structures, along the lines that Zwart proposes. A structural account would be relevant if the structure constrained possible clusters that occur. But this does not appear to be the case. For example, the clusters 3-1-2 and 4-1-2-3 clearly involve movement of the lowest verb to the left, a relation that can be stated without reference to the structure.

Let us consider the alternative perspective, which is that the correct representation of the knowledge of the possible clusters can be expressed in terms of linear properties of the sequence of verbs. The key question that must be resolved in fleshing out such a proposal are the parameters of variation. Let us suppose that the parameters are simply the relative orderings of pairs of verbs in the sequence. In a three-verb sequence, the parameters are 1-2, 2-3, and 1-3
Linguistic Features Cluster in a Social Network

where one does not follow from the other two. These parameters will produce eight possible combinations.

<table>
<thead>
<tr>
<th>Language name</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>1-2</td>
<td>2-3</td>
<td>(1-3)</td>
<td>1-2-3</td>
</tr>
<tr>
<td>G</td>
<td>2-1</td>
<td>2-3</td>
<td>1-3</td>
<td>*2-1-3</td>
</tr>
<tr>
<td>A</td>
<td>1-2</td>
<td>3-2</td>
<td>1-3</td>
<td>1-3-2</td>
</tr>
<tr>
<td>E</td>
<td>2-1</td>
<td>3-2</td>
<td>(1-3)</td>
<td>3-2-1</td>
</tr>
<tr>
<td>F</td>
<td>1-2</td>
<td>2-3</td>
<td>(3-1)</td>
<td>1-2-3</td>
</tr>
<tr>
<td>B</td>
<td>2-1</td>
<td>2-3</td>
<td>3-1</td>
<td>2-3-1</td>
</tr>
<tr>
<td>C</td>
<td>1-2</td>
<td>3-2</td>
<td>3-1</td>
<td>3-1-2</td>
</tr>
<tr>
<td>H</td>
<td>2-1</td>
<td>3-2</td>
<td>(3-1)</td>
<td>3-2-1</td>
</tr>
</tbody>
</table>

Table 3. Verb clusters generated by simulation

It can be seen from the table that all possible orders are produced by these three parameters, including the unattested *2-1-3. For languages A, F, E and H the value of P3 is

\[ P3 = \text{<topicalized, untopicalized>} \]

11 We recognize that it is possible to postulate linguistically more sophisticated parameters that will produce the distribution of possible sequences. One property is the absolute position of the tensed verb, \( P1 = \text{<tensed initial, tensed final>} \). One property is the position of the head with respect to the complement, \( P2 = \text{<head first, complement first>} \). And the third property is the position of the IPP infinitive, \( P3 = \text{<topicalized, untopicalized>} \). The following table aligns the parameter values with the possible orders.

<table>
<thead>
<tr>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>2-verb sequence</th>
<th>3-verb sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tensed init</td>
<td>Head 1st</td>
<td>+Topic</td>
<td>1-2</td>
<td>3-1-2</td>
</tr>
<tr>
<td>Tensed final</td>
<td>Head 1st</td>
<td>+Topic</td>
<td>2-1</td>
<td>3-2-1</td>
</tr>
<tr>
<td>Tensed init</td>
<td>Head 1st</td>
<td>-Topic</td>
<td></td>
<td>1-2-3</td>
</tr>
<tr>
<td>Tensed final</td>
<td>Head 1st</td>
<td>-Topic</td>
<td>2-3-1</td>
<td></td>
</tr>
<tr>
<td>Tensed init</td>
<td>Comp 1st</td>
<td>+Topic</td>
<td></td>
<td>3-1-2</td>
</tr>
<tr>
<td>Tensed final</td>
<td>Comp 1st</td>
<td>+Topic</td>
<td>3-2-1</td>
<td></td>
</tr>
<tr>
<td>Tensed init</td>
<td>Comp 1st</td>
<td>-Topic</td>
<td></td>
<td>1-3-2</td>
</tr>
<tr>
<td>Tensed final</td>
<td>Comp 1st</td>
<td>-Topic</td>
<td></td>
<td>3-2-1</td>
</tr>
</tbody>
</table>
irrelevant since 1-2 and 2-3 imply 1-3, and 3-2 and 2-1 imply 3-1. For the sake of the illustration we may take the feature P3 to be completely determined by the other two features, recognizing that this is an unprincipled assumption.  

Let us now consider the simulation. Suppose that there are in fact these three parameters, and that the unattested *2-1-3 actually exists at some point in a dialect of the language, but is relatively weakly attested. Figure 9 displays the development of a language group after 775 steps.

Figure 9. Histogram, step 775

---

12 This assumption allows us to keep the number of distinct languages at 8, a number that is an artifact of the assumption made in the simulation that there are three two-valued features at play.
We can see in Figure 9 that there are four more or less robust languages at this point, three non-existent languages and one that is weak. Let the languages in the histogram be as defined above in Table 3. On this view, the non-existent *2-1-3 is not ruled out in principle, it simply does not exist as a consequence of the interactions that take place in the network.

It is apparent that there is no linguistic content in the simulation, and we have chosen to name the languages in such a way that the simulation would illustrate the point. We do not claim to have explained the West Germanic data. The point is simply to show that the non-existence of a particular combination of properties, in this case the absence of *2-1-3, is not necessarily linguistically interesting. Even if there are linguistic factors at play, a point to which we will turn momentarily, it would be a mistake to conclude that the non-existence of a particular combination
of properties signifies that this combination is impossible for some principled reason. It may simply be, as suggested earlier, that this combination of features is possible but dispreferred, and hence dies out under pressure from more favored alternatives.

So we must consider the question, is there any way to predict which of the languages will in fact go extinct? In the simulation model as it stands there is no way to do these, because the three features and their values are equipotent. But, suppose that we add the further condition that a language property that is computationally more complex or “marked” than its competitors in the set will be dispreferred. In the case of word order in West Germanic verb clusters, we may reasonably hypothesize that the processing of a head and its complement is facilitated by adjacency and canonical order. In the case of the order *2-1-3, the head of 2-3 is 2. However, 2 is not adjacent to 3, which would constitute a processing burden. Added to this is the fact that the head 1 intervenes between 2 and 3, which constitutes a computational burden with respect to the identification of 1 as the head of *2-3. Our suggestion is that while this particular state of affairs is not grammatically impossible, and could be processed by a native speaker if it was sufficiently well-attested in a language, it is highly marked and will be dispreferred by the language learner. It remains, of course, to provide an explicit characterization of relative computational complexity, a challenge that would take us beyond the bounds of the present study. See Culicover and Nowak (in preparation a).

Note that the order 3-1-2 also exists. Whatever story we construct regarding relative computational complexity should not predict that 3-1-2 cannot occur, although we may claim that it is dispreferred and therefore less likely to occur; perhaps it occurs in a minority of cases, for a principled reason. It is plausible, although purely a speculation at this point, that a

---

13 For discussion of the factors that might enter into such complexity, see Hawkins (1994) and Kirby (1997).
14 We leave open the question of whether or not there is a universal canonical order. For some important ideas about the relative ease of processing uniform branching structure, see Hawkins (1994).
leftward displacement is computationally less costly than a rightward displacement, other things being equal, a view that has its roots in traditional thinking about movement in generative grammar (see for example Kayne (1994)).

4. Weak correlations

The appearance of gaps in the set of languages is an extreme case of a more general property of the simulation model, which is that feature values correlate. In other words, the model evolves over time so that there is a greater than chance tendency for one feature value to appear when another appears. However, the correlation is typically not 100%, so that the opposite pattern may also occur, although less frequently.

In a random initial state the level of correlation between feature values is zero. Since the value of a feature for an individual is determined by the interactions between that individual and its interaction partners, over time all of the feature values of an individual tend to become more highly correlated with those of its interaction partners. An illustration of a typical history of correlation is given in Figure 11.
Figure 11. History of correlations over 100 steps

As can be seen, features F2 and F3 correlate increasingly closely over the course of the simulation, while features F1-F3 and F1-F2 do not.

It is instructive to observe the kind of pattern that is associated with such correlation. In Figure 12 we see a local region of the space for F2 and F3 after 100 steps.
The visual evidence reveals the tendency for F2 and F3 to have the same value, even while there are exceptions. The correlation data confirms this observation. We suggest that such tendencies have a linguistic interpretation along the lines of Greenbergian universals (Greenberg (1963)). Consider one chosen more or less at random, for example, Universal 4: “With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.”\(^{15}\) Let F2 be [+/-postpositional], and let F3 be [+/-OV]. Figure 12 can be

---

\(^{15}\) Similar universals are:

**Universal 9:** “With well more than chance frequency, when question particles or affixes are specified in position by reference to the sentence as a whole, if initial, such elements are found in prepositional languages and, if final, in postpositional.”

**Universal 17:** “With overwhelmingly more than chance frequency, languages with dominant VSO order have the adjective after the noun.”
understood as an illustration of the situation in which most SOV languages are postpositional, although some are not; we illustrate in Figure 13.

Figure 13. Hypothetical correlation between SOV and pre/post-position

While it is almost certainly true that there are other sorts of explanations for some implicational universals of this type, the possibility cannot be ruled out that at least some are due simply to the correlation of features through the interaction of the network.

It might perhaps seem counterintuitive that correlations should arise in the absence of some logical connection between the features. The idea that unrelated attributes can become
Linguistic Features Cluster in a Social Network

correlated in the process of social communication has been suggested in the context of attitudes by Abelson (1979). Assuming the representation of society as a social network, Abelson suggested that since all the attitudes are transmitted by the same connections among individuals in the network, the individuals located close to each other in the network tend to have similar attitudes on all the issues what will cause independent attitudes to become correlated. In a cellular automata framework, the mathematical model for this phenomenon was developed by Lewenstein, et al. (1992) and quantitatively demonstrated by Latane’ (1996).

In the context of language change there are two main reasons for the emergence of correlations. The first mechanism is entirely due to clustering, the second is due to the fact that individuals are equally strong when transmitting all the feature values. Consider a situation in which two linguistics features are randomly distributed in a population of N individuals. Due to chance those features are likely to be slightly correlated at the initial state. The size of this correlation may be quite high for small N, and it will approach 0 for larger N.\(^{16}\) As clustering develops, individuals are no longer independent. In fact, every individual in a cluster has the same value on one or more parameters. The independent unit of observation is now effectively a single cluster, not a single individual. Since the effective number of observations is now determined by the magnitude of the number of clusters and not by the number of single individuals, a much larger correlation may be expected by chance.

This can be shown by an imaginary example. Assume that the simulation area forms a circle. Let us also assume that the in the process of simulation each of the two valued features formed two perfect semi circles.

\(^{16}\) Standard statistical tables provide the size of correlation coefficient that may be expected due to chance.
It is easy to see that the only situation when the two features will be uncorrelated is when the lines dividing the two clusters for each parameter are perpendicular, which is a very unlikely event. Another equally unlikely event is that the two lines are parallel, which will result in a perfect correlation. Every other arrangement will cause the specific value of one feature to be more frequently paired with the specific value of the other feature, and thus a correlation will appear.
Thus we see that correlations emerge as a necessary property of the social network, and do not necessarily reflect any interaction between the parameters, biases in favor of one value or another, or underlying cognitive or computational mechanisms.

5. Contact

5.1. Change through contact

We can also use simulations to model the effects of contact between languages. In such a simulation, the space is divided into two or more distinct regions, each of which contains a single language at the outset. We consider the simplest case, where there are only two languages in contact. All of the speakers in one region have one set of feature values, and all of the speakers in the other region have another set. Change can be seen at the border between the two groups, as speakers influence one another. Since speakers influence one another with respect to all features at the same time, we will expect to find an increase in correlation over time as the border languages become more widespread.

In Figure 16 we illustrate this contact situation. We assume for the sake of the illustration that each feature is four-valued, so that in principle there are sixteen possible languages. However, since each feature has only two of the four values to begin with, these are the only values that can be shared with the other speakers. Hence there can be only eight languages. We assume that the interaction distance is 2, and that the number of interaction partners is 4. As can be seen, at the initial stage all of the speakers in the lefthand field have one set of feature values (gray, gray, black) and all of the speakers in the righthand field have another set (black, white, white).
We are able to introduce some diversity in the set of languages that emerge in a contact situation by allowing for noise. When there is noise, a speaker will not necessarily adopt exactly the feature values of its neighbors. As noted earlier, noise incorporates all factors that are not otherwise controlled for in the simulation that can affect the state of an individual, such as imperfect communication, movement in space, and so on. Another variable is whether individuals take their own current state as input into their next state, which we term “self-support”. By removing self-support, we treat all relatively weak individuals as highly susceptible to outside influence, for whatever political, social or economic reason. We illustrate the results of a single run of the model with 80% noise, which produces rather dramatic (in fact, exaggerated) results. In this simulation the speakers
on the righthand side of the border are powerful individuals and those on the left are relatively weak.

Figure 17. Map of languages in contact after 100 steps

There are eight distinct languages in this space. Notice that in spite of the high level of noise, few of the speakers of the socially dominant language in the righthand portion of the space have changed their language at all, while a number of those in the lefthand side have changed completely.
5.2. A case study

As with gaps, it is possible to find real cases of language change that are mimicked by our simulation model. Thomason & Kaufman (1988) discuss the case of Asia Minor Greek in contact with Turkish. They argue that the appearance of Turkish features in just the portion of the Greek population that is in contact with Turkish provides strong evidence that these speakers of Greek are adopting the features of Turkish and not producing spontaneous change. At the same time, the Turkish spoken in this region shows little if any effect of contact with the Greek. Our simulation is a caricature of the real situation, of course, since real languages are defined by many thousands, if not tens or hundreds of thousands, of distinct features: phonological, morphological and syntactic rules and regularities, as well as an enormous number of distinct words and idioms. Nevertheless, our model suggests that under the right circumstances, such a unidirectional influence is possible and plausible. The crucial factor in creating such a situation is that the speakers of one language are politically and economically more powerful than the speakers of the other language. Such is the situation in the Greek-Turkish contact, where the Turks are the employers and the Greeks the workers who must become bilingual in order to function in the workplace.

6. Further perspectives

We have argued in this paper that it is possible in principle to attribute to the social network such phenomena as gaps in the set of languages, tendencies for certain properties to co-occur in languages, and the emergence of new languages when language come into contact. It is almost certainly the case that the intrinsic content of the features also plays an explanatory role as well in each of these categories of phenomena. Our point is simply that there are at least two
potential sources of explanation. This observation at first glance appears to complicate the task of the theorist, who is no longer able to posit that the distributional patterns of linguistic phenomena reveal something interesting about the language capacity.

However, upon closer examination, it appears that there is always a potential explanatory role for linguistic theory (as contrasted with the social network theory). It is necessary to explain why it is that certain feature values or feature combinations are less preferred than others, perhaps non-existent. In a relatively small homogeneous population, it may not be possible to distinguish between accidental gaps and systematic gaps. It is for this reason that detailed studies of variation among closely related and neighboring dialects are unlikely to provide convincing evidence that will allow us to distinguish between linguistically interesting gaps and those that are the accidental product of the social interaction. But if the same gaps occur in regions that do not communicate with one another as a consequence of distance or natural barriers, then it is reasonable to surmise that the gaps and tendencies reflect some intrinsic linguistic properties, and are not due simply to the accidental growth in the population of one feature and the decline of another. Similarly, correlations that go in the same direction in non-communicating regions are more likely to be due to intrinsic properties of the features.

We speculate that one measure of complexity that underlies the relative preference of one set of feature values over another concerns the mapping between syntactic and semantic representations, along the lines of Culicover (1999). On such a view, the syntactic representation reflects more or less faithfully the (universal) conceptual structure organization. The less faithful the representation, the more complex the mapping. Whether there are higher level computational measures, where groups of mappings are preferred over others according to the extent to which they express the correspondences in the same way, is a question that we must leave open here.

It is also apparent that the simulation we have discussed here is limited in its scope. It is possible to extend this approach to explore productively the effects of migration, aging and death,
Linguistic Features Cluster in a Social Network

the relative computational complexity of the features, individual bias, and agents’ capacity for multilingualism. Moreover, the number of features would have to be expanded considerably before the simulation could be regarded as even remotely realistic. Nevertheless, we believe that even this simple simulation demonstrates several points of some theoretical importance.

References


Linguistic Features Cluster in a Social Network


