Ultrametric Distance in Syntax

Mark D. Roberts

Charles University in Prague, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

Abstract

Phrase structure trees have a hierarchical structure. In many subjects, most notably in **taxonomy** such tree structures have been studied using ultrametrics. Here syntactical hierarchical phrase trees are subject to a similar analysis, which is much simpler as the branching structure is more readily discernible and switched. The ambiguity of which branching height to choose, is resolved by postulating that branching occurs at the lowest height available. An ultrametric produces a measure of the complexity of sentences: presumably the complexity of sentences increases as a language is acquired so that this can be tested. All ultrametric triangles are equilateral or isosceles. Here it is shown that $\overline{X}$ structure implies that there are no equilateral triangles. Restricting attention to simple syntax a minimum ultrametric distance between lexical categories is calculated. A matrix constructed from this ultrametric distance is shown to be different than the matrix obtained from features. It is shown that the definition of c-command can be replaced by an equivalent ultrametric definition. The new definition invokes a minimum distance between nodes and this is more aesthetically satisfying than previous varieties of definitions. From the new definition of c-command follows a new definition of of the central notion in syntax namely government.

1. Introduction

1.1. Ultrametric Literature

Ultrametrics are used to model any system that can be represented by a bifurcating hierarchical tree. To list briefly some areas where ultrametrics have been applied. Perhaps the most important application is to taxonomy, Jardine and Sibson (1971, Ch.7), and Sneath and Sokal (1973). Here the end of a branch of the tree represents a species and the ultrametric distance between them show how closely the species

are related. Hierarchical cluster methods classify species and also shows how closely species are related. This technique has also been used in semantics, Shepard and Arabie (1979). The technique can become quite complicated because it involves statistical analysis with continuous variates. Ultrametrics have been applied frequently in the theory of spin glass, Weissman (1993). Ultrametrics have been used for description of slowly driven dissipative systems, which exhibit avalanche-like behaviour, these include earthquakes, extinction events in biological evolution, and landscape formation, Boettcher and Paczuski (1997); also ultrametrics can describe systems with fast relaxation, Vlad (1994). Ultrametrics are used in the theory of neural nets, Parga and Virasoro (1986). The dynamics of random walks on ultrametric spaces have been studied, Ogielchi and Stein (1985). Ultrametrics have been applied to the thermodynamics of macromolecules such as RNA, Higgs (1996), the directed polymer problem Perlman and Schwarz (1992), and sociology Schweinberger and Snijders (2003). Bounds on the size of ultrametric structure have been discussed by Baldi and Baum (1986). From a more theoretical angle, a category theory approach has been elucidated by Rutten (1996), a model theoretic approach to ultrametrics is given by Delon (1984), and ultrametric might be related to T-theory, Andreas Dress and Terhalle (1996). The relationship between ultrametric distance and hierarchy is further discussed in Guénoche (1997). Construction of optimal ultrametric trees is discussed by Young and DeSarbo (1995). Ultrametrics are related to p-adelic quantities, Karwowski and Mendes (1994), Murtagh (2004) and B. Dragovich and Volovich (2009). P-adelic quantities are used in string theory: the way that ultrametrics enters here is explained in §10&§13.4 of Brekke and Freund (1993). There does not seem to be any straightforward connection of any of the above to the optimization techniques of Prince and Smolensky (1997). As well as ultrametric trees, there are also *decision trees* Hammer (1998), and the connection between them is still not known. Some of the above ultrametric applications have been reviewed by R. Rammal and Virasoro (1986). Chomskian syntatic trees have been axiomatized by Backofen and Vijay-Shankar (1995).

1.2. Ultrametric Inequalities

There is the following relationship between trees and ultrametrics. An N-leaf edge(node)-weighted tree corresponds to an N × N square matrix M in which M_{ij} = the sum of the weights of the edges (nodes) in the shortest path between i and j. When the weights are non-zero and non-negative, M is a distance in the usual sense.

$$\forall x, y, z \quad M_{xy} = 0 \quad \text{if } x = y \tag{1}$$
$$M_{xy} > 0 \quad \text{for} \quad x \neq y \tag{2}$$
$$M_{xy} = M_{yx} \tag{3}$$
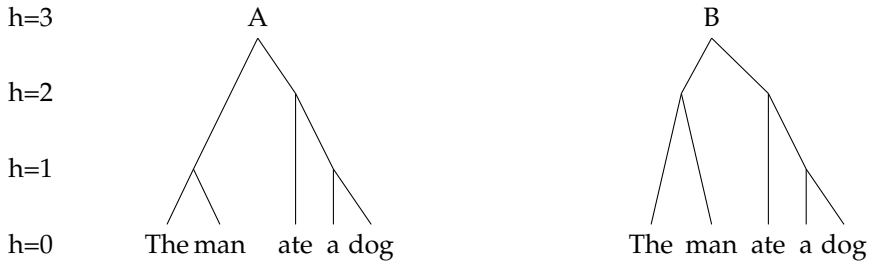$$M_{xy} \leq M_{xz} + M_{zy}. \tag{4}$$

Figure 1: Different syntactic descriptions of "the man ate a dog"

However if the triangle inequality (4) is replaced by

$$M_{xy} \leq \max\{M_{xz}, M_{zy}\}. \tag{5}$$

then M is an ultrametric, equation (5) implies (4).

1.3. Syntactic Phrase Trees

For the analysis of syntactic phrase trees the necessary technique is quite simple and is illustrated by the examples in Section 2. Psychological analysis of phrase trees has been carried out by Johnson (1965) and Levelt (1970). The phrase tree approach has critics Botha (1965); also Evans and Levison (2009) question the existence of language universals. The examples here mainly follow the examples in Lockward (1972), Kayne (1981), McCloskey (1988), and especially Haegeman (1994). There are at least five reasons for introducing an ultrametric description of syntax.

The first is to completely specify tree (also called *dendrogram*) structure. Consider the following example illustrated by Figure 1.

For current syntactic models sometimes nodes are taken to occur at the highest level, and sometimes the two trees are equivalent see McCloskey (1988, footnote 6); however consider the ultrametric distance between 'the' and 'man',

$$A(\text{the}, \text{man}) = 1, \qquad B(\text{the}, \text{man}) = 2, \tag{6}$$

where the numbers are the height of the lowest common node above the two lexical items. This ambiguity does not occur in current syntactic models, and a purpose of an ultrametric model is to disambiguate the difference in height, because this might have consequence in the complexity of the encoded model, see the next point.

The second is it gives a measure of the complexity of a sentence: the greater the ultrametric distance required the more complex a sentence is. The above can also be viewed in terms of 'closeness'. The example Figure 1 illustrates that current syntactic

models give no notion of how 'close' determiners and nouns are. However ultrametrics do give an indication of closeness and this can be compared: *firstly* to the closeness indicated by features, *secondly* to the idea that if elements of a sentence are not sufficiently close then there is a BARRIER Chomsky (1986b) to movement, roughly speaking barriers impede the movement of phrases to different places in a sentence. Only the closeness as indicates by features is looked at here. In traditional syntax phrases can be iteratively embedded to give sentences of unbounded length and complexity. A degree of sentence complexity perhaps corresponds to the height of the tree representing the sentence. As people can only process a finite amount of information this height must be finite. In the traditional theoretical framework there is no finite bound on sentence length. An upper bound could perhaps be found by experiment. Inspection of phrase trees suggests a first guess of $h = 12$.

The third is that it means that syntax is described in the same formalism as that used in a lot of other sciences, for example those topics described in the first paragraph of Section 1.1, so that there is the possibility of techniques being used in one area being deployed in another.

The fourth is that an ultrametric formulation might allow a generalization so that ideas in syntax can be applied to other cognitive processes.

The fifth, see the next section 1.4, and perhaps the most important, is that it might be possible to use some sort of minimum distance principle in syntax: it could be this minimum description which would have application in other cognitive processes. In other words that ultrametric trees should be simple rather than complicated and that the sort of mechanism use to encode simple tress might be used elsewhere.

1.4. Ockham's Razor

Minimum description in science goes back several hundred years to "*Ockham's razor*" or perhaps further, see for example Sorton (1947, page 552). The principle of least action (see for example Bjorken and Drell (1965, §11.2)) in physics is that minimal variation of a given action gives field equations which describe the dynamics of a system. For example, Maxwell's equations can be derived from a simple action by varying it. In the present context one would hope that syntax allows for a minimum encoding of semantic information, the minimum encoding being given by some ultrametric measure. A different approach along these lines is that of Rissanen (1983) and Zadrozny (1999). Briefly they assign a length of 1 to each symbol in a sentence, then the MINIMUM DESCRIPTION LENGTH states that the best theory to explain a set of data is the one which minimizes both the sum of: i) the length, in bits, of the description of the theory, and ii) the length, in bits, of data when encoded with the help of the theory. Christiansen (2001) discusses how constraint handling rules (CHR) can be applied to grammars. This can be thought of as a minimizing procedure.
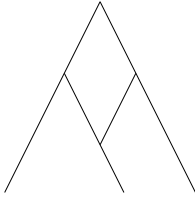
Figure 2: A ʀᴇᴛɪᴄᴜʟᴀᴛᴇ tree
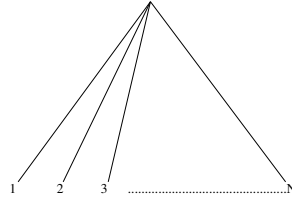
Figure 3: N-ᴀʀʏ branching

1.5. Reticulate & N-ary Trees

A ʀᴇᴛɪᴄᴜʟᴀᴛᴇ tree is a tree in which there are one or more sets of reconvergent branches, illustrated by Figure 2, a ɴᴏɴ-ʀᴇᴛɪᴄᴜʟᴀᴛᴇ tree is a tree in which the branches do not reconverge. N-ᴀʀʏ branching is illustrated by Figure 3. Bɪɴᴀʀʏ branching is N-ᴀʀʏ branching with N = 2. N-ᴀʀʏ branching can be replaced by binary branching if additional layers are used. A sᴡɪᴛᴄʜᴇᴅ tree is a tree in which all the branches are binary. Syntactic phrase trees are ɴᴏɴ-ʀᴇᴛɪᴄᴜʟᴀᴛᴇ and sᴡɪᴛᴄʜᴇᴅ. In most linguistic theories all syntactic phrase trees have $\overline{X}$ structure, Jackendoff (1977). $\overline{X}$ structure implies binary branching, see Section 2.3 and Figure 8. Here attention is restricted to theory which has $\overline{X}$ structure.

1.6. Sectional Contents

In Section 2 it is shown how to represent trees by matrices and triangles. All $\overline{X}$ triangles are isosceles but not equilateral. In Section 3 the matrix **U** for the minimum ultrametric distance for lexical categories is given. For simplicity discussion is limited to active voice sentences with only determiners, nouns, transitive verbs, adjectives, and prepositions. Inclusion of case theory, COMP, INFL,.. might be of interest but would complicate matters. In Section 4 the singular matrix **F** for features is given. **F** is not an ultrametric matrix and there appears to be no relation to **U**. In Section 5 it is shown that the notion of ᴄ-ᴄᴏᴍᴍᴀɴᴅ is equivalent to an ultrametric minimum distance. This allows a new definition of government to be given. In appendix Section 7 other linguistic hierarchies are discussed; in particular there appears to be at least two separate occurrences of culturally determined partial ordered hierarchies - the *accessibility hierarchy* for relative clauses and the *universal colour ordering*. For completion in appendix Section 7 there is a very brief account of what these hierarchies are, a comparison and contrasting of them, and the speculation that they are specific examples of a *grand cultural hierarchy*. The question arises of why such hierarchies should exist, and it might be because they reduce the amount of memory needed to process information by clumping information together in the style of Miller (1956); for a more
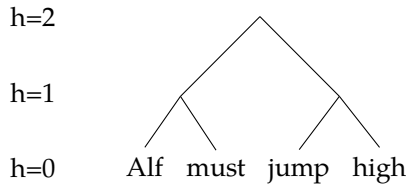
Figure 4: The simplest BINARY tree for "Alf must jump high"

recent reference see Cowan (2001). A *hierarchy* is an example of a representation as discussed by Roberts (2005).

2. $\overline{\text{X}}$ Structure Implies No Equilateral Triangles

2.1. Binary and N-ary Branching for simple sentences

$\overline{\text{X}}$ structure implies BINARY BRANCHING Haegeman (1994, p.139), and the Figure 8. To see what this implies for ultrametric distances consider all five species of BINARY BRANCHED tree. The fixed word order in 'Alf must jump high' reduces the total number of possible matrices form 15 to 8. The *first* has diagram Figure 4 (compare Haegeman (1994, p.141 diagram 84a)) and corresponding matrix:

$$
\text{First} = \begin{matrix}
\bullet & A & M & J & H \\
A & 0 & 1 & 2 & 2 \\
M & . & 0 & 2 & 2 \\
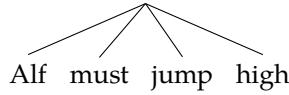J & . & . & 0 & 1 \\
H & . & . & . & 0
\end{matrix} \tag{7}
$$

respectively, where A, M, . . . are short for 'Alf', 'must'. The matrices corresponding to the other *four* BINARY BRANCHED trees are (compare Haegeman (1994, p.141&142 diagrams 84b,c,d,e):

$$
\text{Second} = \begin{matrix}
\bullet & A & M & J & H \\
A & 0 & 3 & 3 & 3 \\
M & . & 0 & 2 & 2 \\
J & . & . & 0 & 1 \\
H & . & . & . & 0
\end{matrix} \tag{8}
\qquad
\text{Third} = \begin{matrix}
\bullet & A & M & J & H \\
A & 0 & 3 & 3 & 3 \\
M & . & 0 & 1 & 2 \\
J & . & . & 0 & 1 \\
H & . & . & . & 0
\end{matrix} \tag{9}
$$

Figure 5: The 4-ᴀʀʏ tree for "Alf must jump high"

$$\text{Fourth} = \begin{array}{c|cccc} \bullet & A & M & J & H \\ A & 0 & 2 & 2 & 3 \\ M & . & 0 & 1 & 3 \\ J & . & . & 0 & 3 \\ H & . & . & . & 0 \end{array} \quad (10) \qquad \text{Fifth} = \begin{array}{c|cccc} \bullet & A & M & J & H \\ A & 0 & 1 & 2 & 3 \\ M & . & 0 & 2 & 3 \\ J & . & . & 0 & 3 \\ H & . & . & . & 0 \end{array} \quad (11)$$

There are *two* 3-ᴀʀʏ trees with matrices (compare Haegeman (1994) p.142 (Haegeman, 1994) diagrams 84g and 84h): Haegeman (1994, p.142 diagrams 84g and 84h)):

$$\text{Sixth} = \begin{array}{c|cccc} \bullet & A & M & J & H \\ A & 0 & 1 & 1 & 2 \\ M & . & 0 & 1 & 2 \\ J & . & . & 0 & 2 \\ H & . & . & . & 0 \end{array} \quad (12) \qquad \text{Seventh} = \begin{array}{c|cccc} \bullet & A & M & J & H \\ A & 0 & 2 & 2 & 2 \\ M & . & 0 & 1 & 1 \\ J & . & . & 0 & 1 \\ H & . & . & . & 0 \end{array} \quad (13)$$

and finally there is *one* 4-ᴀʀʏ tree (compare Haegeman (1994) p.142 (Haegeman, 1994) diagram 84f) Haegeman (1994, p.142 diagram 84f)) with diagram Figure 5 and matrix:

$$\text{Eighth} = \begin{array}{c|cccc} \bullet & A & M & J & H \\ A & 0 & 1 & 1 & 1 \\ M & . & 0 & 1 & 1 \\ J & . & . & 0 & 1 \\ H & . & . & . & 0 \end{array} \quad (14)$$

2.2. Triangle representation of the proceeding matrices

All ultrametric triangles are isosceles with small base, but only some are equilateral. The previous subsection (2.1) suggests that binary branching implies that there are no equilateral triangles in ultrametric models of syntax. For example from matrix (13), $d(A, M) = 1, d(A, J) = 2, d(M, J) = 2$ has the triangle representation Figure 6, and from matrix (14), $d(A, M) = 1, d(A, J) = 1, d(M, J) = 1$ giving in the triangle representation Figure 7.

In the next section it is proved that $\overline{X}$ structure implies that there are no equilateral triangles.

J

2 / 2

M ———— A
     1
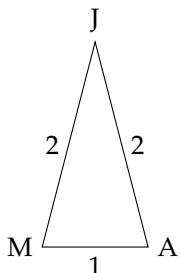
Figure 6: The isosceles triangle

J

1   1

M ———— A
    1
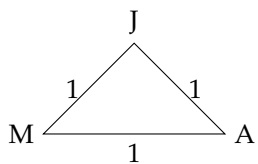
Figure 7: The equilateral triangle

h=i+2   XP

h=i+1 $\overline{\text{X}}$
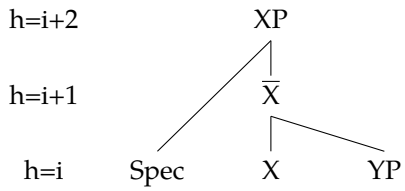
h=i   Spec   X   YP

Figure 8: The $\overline{\text{X}}$ Template

2.3. The $\overline{\text{X}}$ Template

The $\overline{\text{X}}$ template Figure 8 is the form that nodes take in syntax.
The matrix representation of this is:

$$\mathbf{\mathcal{X}} = \begin{array}{c|ccc} \bullet & \text{Spec} & \text{X} & \text{YP} \\ \text{Spec} & 0 & i+2 & i+2 \\ \text{X} & . & 0 & i+1 \\ \text{YP} & . & . & 0 \end{array} \tag{15}$$

From this the triangle representation is Figure 9. This is isosceles but not equilateral.

3. The minimum ultrametric distance between lexical categories

3.1. The minimum distance principle

We assume that it is the minimum distance between lexical categories that is important, and refer to this as the MINIMUM DISTANCE PRINCIPLE. In part this is motivated by the discussion in Section 1.4. A current psycholinguistic model of sentence production is the garden path model, see for example Frazier (1987). Part of this model requires *the minimal attachment principle*, which is "do not postulate unnecessary nodes.":
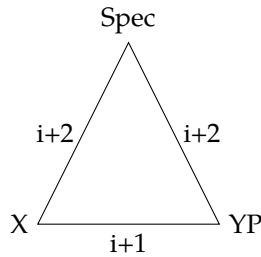
Figure 9: The triangle representation of $\overline{X}$ structure
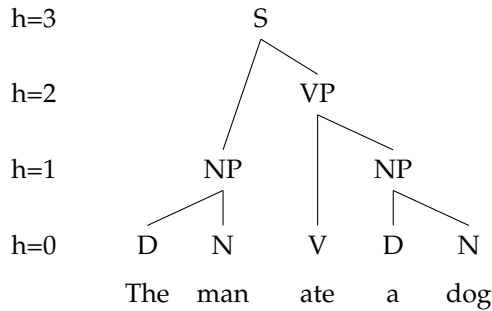
Figure 10: The correct tree for the example in Section 1.

this can be thought of as a minimum principle. The MINIMUM DISTANCE PRINCIPLE implies that the correct tree for equation (6) illustrated by Figure 1 is Figure 10, so that all entries occur at the lowest possible height.

Thus in particular tree **A** is preferred to tree **B**. This assumption does not effect the matrix **U** (16) given and described below, but will have an effect when the analysis is extended to θ-theory, see Haegeman (1994, §3.2.3). From the above $d(N, D) = 1, d(N, V) = d(D, V) = 2$. Similarly from Figure 11, $d(V, A) = 4$.

Constructing other examples gives the ultrametric distance matrix

$$
\mathbf{U} = \begin{array}{c|ccccc}
\bullet & D & N & V & A & P \\
D & 0 & 1 & 2 & 2 & 2 \\
N & . & 0 & 2 & 2 & 2 \\
V & . & . & 0 & 4 & 3 \\
A & . & . & . & 0 & 3 \\
P & . & . & . & . & 0 \\
\end{array}
\tag{16}
$$

h=5   S
h=4   VP
h=3   AP
h=2   PP
h=1   NP   NP
h=0   D   N   V   A   P   D   N
      The man is envious of an elephant

Figure 11: The distance between verbs and adjectives.

Ignoring the determiner D ('a' or 'the') and ordering the matrix NPVA (noun, pronoun, verb, adjective) suggests the pattern

$$
\mathbf{I} =
\begin{matrix}
0 & i & i & i & \ldots \\
. & 0 & i+1 & i+1 & \ldots \\
. & . & 0 & i+2 & \ldots \\
. & . & . & 0 & \ldots
\end{matrix}
\tag{17}
$$

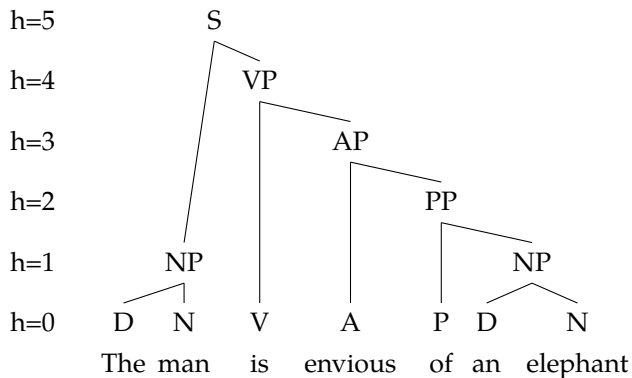which is compatible with the $\overline{X}$ matrix of the last section; however it does not follow by necessity as the $\overline{X}$ case holds for a single sentence and $\mathbf{U}$ is constructed from the syntactical representations of several sentences.

4. Features

4.1. A matrix representation of features

This section investigates whether there is a general framework which can describe both the preceding and also "features". An objective is to reduce the large number of objects that can be 'spec' and so on by using the objects features: in other words to introduce a type of atomic theory. Roughly speaking the idea behind features is that adjectives and prepositions have qualities in them that are associated with verbs and nouns, as can be seen from (18) adjectives have +N & +V and prepositions have -N & -V. Haegeman (1994, p.146) gives the following (18) diagram for features:

$$
\text{Features diagram} =
\begin{matrix}
\text{Noun} & +N & -V \\
\text{Verb} & -N & +V \\
\text{Adj.} & +N & +V \\
\text{Pre.} & -N & -V
\end{matrix}
\tag{18}
$$

in words nouns have features of +noun and -verb, adjectives have features of +noun and +verb, and so on. This can be represented by the matrix

$$
\text{Features matrix} = F = \begin{array}{c|cc}
\cdot & \text{Noun} & \text{Verb} \\
\text{Noun} & +1 & -1 \\
\text{Verb} & -1 & +1 \\
\text{Adj.} & +1 & +1 \\
\text{Pre.} & -1 & -1
\end{array}
\tag{19}
$$

The Pauli matrices (see for example Bjorken and Drell (1965, p.378)) are

$$
I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad
\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad
\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad
\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.
\tag{20}
$$

A square matrix can be constructed by assuming that the matrix is symmetric. This leaves only one unknown $F(A, P)$. Taking $F(A, P) = -1$ gives equal number of positive and negative entries in the matrix

$$
\text{F completion} = \mathbf{FC} = \begin{array}{c|cccc}
\bullet & N & V & A & P \\
N & +1 & -1 & +1 & -1 \\
V & -1 & +1 & +1 & -1 \\
A & +1 & +1 & +1 & -1 \\
P & -1 & -1 & -1 & +1
\end{array}
\tag{21}
$$

which is singular as its determinant vanishes. There appears to be no relation between matrix **FC** (21) and matrix **U** (16). **FC** can be expressed as

$$
\mathbf{FC} = \begin{pmatrix} I - \sigma^1 & -i\sigma^2 + \sigma^3 \\ +i\sigma^2 + \sigma^3 & I - \sigma^1 \end{pmatrix}.
\tag{22}
$$

However this does not correspond in any straightforward way to any of the Dirac matrices (see for example Bjorken and Drell (1965, p.378)) in standard representations.

4.2. Pseudo-inverse of the features matrix

Another way to proceed is to use pseudo-inverses which can be calculated using the program Octave, then

$$
\mathbf{pinv}(\mathbf{F}) = \frac{1}{8} \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{pmatrix},
\tag{23}
$$

then in terms of Pauli matrices (20)

$$
\mathbf{F} = \begin{pmatrix} I - \sigma_1 \\ \sigma_3 + i\sigma_2 \end{pmatrix}, \qquad
\mathbf{pinv}(\mathbf{F}) = \frac{1}{8} (I - \sigma_1, \sigma_3 - i\sigma_2),
\tag{24}
$$

giving products

$$\mathbf{pinv}(\mathbf{F}) * \mathbf{F} = \mathbf{I}, \qquad \mathbf{F} * \mathbf{pinv}(\mathbf{F}) = \frac{1}{2} \begin{pmatrix} \mathbf{I} - \sigma_1 & 0 \\ 0 & \mathbf{I} - \sigma_1 \end{pmatrix}. \qquad (25)$$

Having got an inverse (23) one would hope to be able to use it to predict properties of nouns, verbs and so on, but it is not yet clear how.

4.3. Truth features matrix

The above approach is not unique; consider instead of objects having $\pm N, \pm V$ having truth values

$$\text{Features truth diagram} = \begin{array}{ccc} \text{Noun} & \text{T} & \text{F} \\ \text{Verb} & \text{F} & \text{T} \\ \text{Adj.} & \text{T} & \text{T} \\ \text{Pre.} & \text{F} & \text{F} \end{array} \qquad (26)$$

with corresponding matrix

$$\text{Features truth matrix} = \mathbf{T} = \begin{array}{ccc} \cdot & \text{Noun} & \text{Verb} \\ \text{Noun} & 1 & 0 \\ \text{Verb} & 0 & 1 \\ \text{Adj.} & 1 & 1 \\ \text{Pre.} & 0 & 0 \end{array} \qquad (27)$$

with inverse

$$\mathbf{pinv}(\mathbf{T}) = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 & 0 \\ -1 & 2 & 1 & 0 \end{pmatrix}, \qquad (28)$$

giving products

$$\mathbf{pinv}(\mathbf{T}) * \mathbf{T} = \mathbf{I}, \quad \mathbf{T} * \mathbf{pinv}(\mathbf{T}) = \frac{1}{3} \begin{pmatrix} 2 & -1 & 1 & 0 \\ -1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (29)$$

which do not simplify in terms of Pauli matrices (20).

The structure of the two approaches does not seem to be the same as can be seen by looking at (19) and (27); and so far it is not clear which is better. It is possible that the above features approach can be extended using more features; for example introducing D=determiner one has the $2^3 = 8$ objects $\{D = (1 \pm 1)/2, N = (1 \pm 1)/2, V = (1 \pm 1)/2\}$, and one could take $\{1, 0, 0\}$ to be a determiner $\{1, 1, 0\}$ to be a noun phrase and so on, then there is the question of how many features have to be introduced to represent all tree objects.
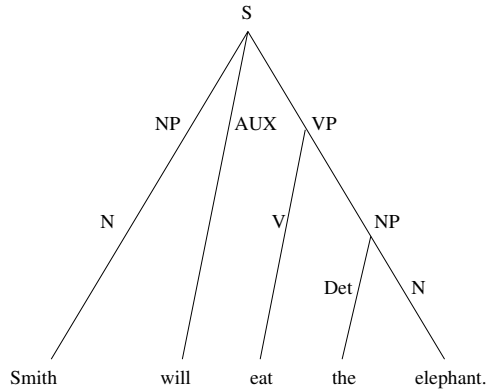
Figure 12: Illustration of DOMINATES

5. Ultrametric Approach to Government

Recall the following definitions in Haegeman (1994):

5.1. Definition of DOMINATES.

Definition Haegeman (1994, p.85) Node **A** DOMINATES node **B** iff:
i) $h(A)$ is higher up or at the same height on the tree as $h(B)$ i.e.$h(A) \geq h(B)$
ii) it is possible to trace a path from **A** to **B** going only downward,
or at most going to one higher node.
Remarks
The *first* requirement is that **A** is at a greater height than **B**. The *second* requirement restricts the possible downward route from **A** to **B** so that it contains at most one upward segment.
Example (compare Haegeman (1994, p.83))The phrase tree in Figure 12 gives the 'dominates' matrix:

$$\mathbf{D} = \begin{array}{c|ccccccccc} \bullet & S & NP(S) & N(S) & AUX & VP & V & NP(E) & Det & N(e) \\ S & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ NP(S) & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ N(S) & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AUX & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ VP & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ V & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ NP(E) & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ Det & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ N(E) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \tag{30}$$

where 1 indicates "**A** dominates **B**" and 0 indicates that it does not.

5.2. Definition of C-command

Definition Haegeman (1994, p.134)

Node **A** c-commands (constituent-commands) node **B** iff:

i) **A** does not dominate **B** and **B** does not dominate **A**,

ii) The first branching node dominating **A** also dominates **B**.

Remarks

The *first* requirement is that there is no direct route up or down from **A** to **B** passing more than one higher node. The *second* requirement restricts **A** and **B** to be 'close'. Haegeman's first criterion for dominance needs to be adjusted: if it is correct then $h(A) > h(B)$ and $h(B) > h(A)$ so that the set of all c-commands is empty, therefore greater than or equal $\geq$ is used here instead of greater than $>$. Haegeman's second criterion for dominance also needs to be adjusted: if no higher node is allowed the set of c-commands is again empty. Chomsky (1986a, p.161) approaches the subject in a different manner using maximal projections.

For example, in Figure 13, $0 < j < k < l$. The corresponding ultrametric matrix is

$$\mathbf{U} = \begin{array}{c|cccc} \bullet & A & B & C & D \\ A & 0 & k & k & l \\ B & . & 0 & j & l \\ C & . & . & 0 & l \\ D & . & . & . & 0 \end{array} \tag{31}$$

The c-command matrix **CM** is

$$\mathbf{CM} = \begin{array}{c|cccc} \bullet & A & B & C & D \\ A & 1 & 1 & 1 & 0 \\ B & 0 & 1 & 1 & 0 \\ C & 0 & 1 & 1 & 0 \\ D & 1 & 1 & 1 & 1 \end{array} \tag{32}$$

h=i+l   H

h=i+k   F
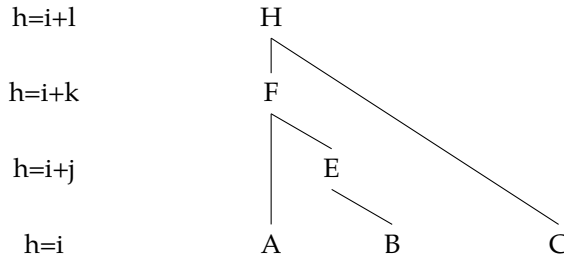
h=i+j   E

h=i   A   B   C

Figure 13: Example of C-COMMANDS.

where 1 indicates **A** C-COMMANDS **A**, and 0 indicates that it does not, similarly for **B, C, D**.

5.3. Definitions of C-Domain & Governs

Definition Haegeman (1994, p.134)
The total of all the nodes C-COMMANDED by an element is the C-DOMAIN of that element.
Definition Haegeman (1994, p.135)
A GOVERNS **B** iff:
i) **A** is a GOVERNOR,
ii)**A** C-COMMANDS **B** and **B** C-COMMANDS **A**.
Remarks:
The *first* requirement is a restriction on the set **A** (in linguistic terminology the category **A**). A GOVERNOR is a part of speech which generalizes the notion of a verb governing an object; unfortunately there does not seem to be a formal definition of it. The *second* requirement is that **A** and **B** should be sufficiently 'close'.

5.4. Definitions of CU-Domain & CU-Command

Now let **D(A)** be the set of all the ultrametric distances to other nodes at the same height and let **M(A)** be the set of these which have the smallest value.
Call **M(A)** the CU-DOMAIN of **A** and say **A** CU-COMMANDS all **B**ε**M**(**A**) (in words **B** is a member of **M(A)**). This is illustrated by Figure 14.

5.5. Theorem showing the identity between C-Domain & CU-Domain

Theorem:
The sets **A** C-COMMANDS **B** and A CU-COMMANDS **B** are identical, likewise the C-DOMAIN and the CU-DOMAIN.
Proof:
From the i) part of the definition of C-COMMAND $h(A) = h(B)$, so that we are only

h=i+l   H

h=i+k   F
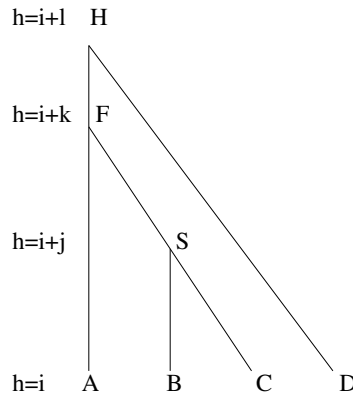
h=i+j   S

h=i   A   B   C   D

Figure 14: Illustration of the Theorem.

concerned with nodes at the same height $h(A) = i$. Let the first branching node above **A** be **F**, with $h(F) = i+k$. Let **H** be any node dominating **F**, with $h(H) = i+l$. Let **E** be the subsidiary node dominating **B** and **C** and dominated by **F**, with $h(E) = i+j$. The closest nodes to **A** are **B** and **C** both with an ultrametric distance k. The sets **D(A)** and **M(A)** are **D(A)**={**A,B,C,D**}, **M(A)**={**A,B,C**}. **A** both c-commands and cu-commands itself and **B** and **C**. The actual integer values $i, j, k, \ldots$ are arbitrary and thus the result holds in general.

5.6. A New Definition of Government

This allows a new definition of GOVERNMENT. **A** GOVERNS **B** iff:
i) **A** is a GOVERNOR.
ii) both $A \varepsilon M(B)$ and $B \varepsilon M(A)$ (in words **A** is a member of **M(B)** and vice versa). This definition of GOVERNMENT is the same as the previous definition of GOVERNMENT, but with the c-command requirement replaced by an ultrametric requirement that distances be minimal.

6. Conclusion

The definition of government in Section 5.6 might at sometime in the future allow the five points in Section 1.3 to be addressed.

$$\boxed{\text{SU > DO > IO > OBL > GEN > OCOMP}}$$

Figure 15: The *accessibility hierarchy*

7. Appendix: Other Linguistic Hierarchies

7.1. The Accessibility Hierarchy

A RELATIVE CLAUSE (*RC*) is a clause that modifies a noun or pronoun that occurs elsewhere in a sentence. The *accessibility hierarchy* (*AH*) for relative clauses is given by Keenan and Comrie (1977) and illustrated in Figure 15.

Noun phrases (NP) occurring to the left of ">" are more accessible than those on the right. SU is short for subject, DO for direct object, IO for indirect object, OBL for major oblique case NP, GEN for genitive NP, OCOMP for object of comparison. The properties of the accessible hierarchy are contained in two sets of constraints.

The accessible hierarchy constraints (*AHCs*) are:

AHC1) A language must be able to relativize subjects.

ACH2) Any RC forming strategy must apply to a continuous segment of the AH.

ACH3) Strategies that apply at one point of the AH may in principle cease to apply at any lower point.

The primary relativization constraints (*PRCs*) are

PRC1) A language must have a primary RC-forming strategy.

PRC2) If a primary strategy in a given language can apply to a low position on the AH, then it can apply to all higher positions.

PRC3) A primary strategy may cut off at any point on the AH.

For a given language a deployment that can be used to relativize a clause at a specified place on the AH can also be used to relativize all more accessible clauses. The type of relativization varies from language to language. There appears to be nothing known on how the skill to deploy a relativization develops in an individual. One would expect that when a given method is applied the less accessible would take longer to process. There seems to be no psycholinguistic tests done to see if this is indeed the case.

7.2. The Berlin-Kay Universal Colour Partial Ordering

The perception of colour often involves the deployment of a colour name strategy. The effect of this is to alter the way the colour is perceived. The five principles of colour perception are:

CP1) The communicability of a referent in an array and for a particular community is very closely related to the memorability of that referent in the same array and for members of the same community.
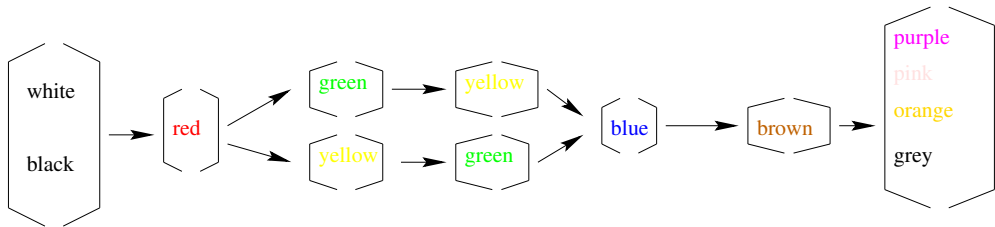
Figure 16: The Berlin-Kay Universal Colour Partial Ordering

CP2) In the total domain of colour there are eleven small focal areas in which are found the best instances of the colour categories named in any particular language. The focal areas are human universals, but languages differ in the number of basic colour terms they have: they vary from two to eleven.

CP3) Colour terms appear to evolve in a language according to the Berlin and Kay (1969) universal partial ordering illustrated by Figure 16.

CP4) Focal colours are more memorable and easier to recognize than any other colours, whether or not the subject speaks a language having a name for the colour.

CP5) The structure of the colour space determined by multi-dimensional scaling of perceptual data is probably the same for all human communities and it is unrelated to the space yielded by naming data.

Again there is a *culturally determined linguistic partial ordering* (or hierarchy). On this occasion it determines the semantic content of individual words rather than syntax rules. Again there appears to be nothing known on how the skill develops in an individual, or any timing tests on the possession of a colour name strategy. The existence of two separate hierarchical *partial orderings* suggests that there is a general mechanism for there construction. Most members of a community seem to develop these culturally determined skills suggesting that the capacity to develop them is usually innate but their manifestation depends on environment.

Bibliography

Andreas Dress, Vincent Moulton and Werner Terhalle. T-theory: An overview. *Europ. J. Combinatorics*, 17:161–175, 1996.

B. Dragovich, A.Yu. Khrennikov, S.V. Kozyrev and I.V. Volovich. On p-adic mathematical physics, p-adic numbers. *Ultrametric Analysis, and Applications Journal*, 1:1–90, 2009.

Backofen, Rogers and Vijay-Shankar. A first-order axiomatization of the theory of finite trees. *Journal of Logic, Language and Information*, 4(1):5–39, 1995.

Baldi, P. and E.B. Baum. Bounds on the size of ultrametric structures. *Phys.Rev.Lett.*, 56:182–184, 1986.

Berlin, Brent and Paul Kay. *Basic Colour Terms*. University of California Press, 1969.

Bjorken, James D. and Sidney D. Drell. *Relativistic Quantum Fields*. McGraw Hill, 1965.

Boettcher, S. and M. Paczuski. Aging in a model of self-organized criticality. *Phys.Rev.Lett.*, 79: 889, 1997.

Botha, Rudolf P. *Challenging Chomsky: The Generative Garden game*. Blackwell Oxford, 1965.

Brekke, L. and Peter G.O. Freund. P-adelic numbers in physics. *Phys.Rep.*, 233:1–166, 1993.

Chomsky, Noam. *Knowledge of Language, Its Nature, Origin and Use*. Praeger Publishers, New York, 1986a.

Chomsky, Noam. *Barriers*. MIT Press,Cambridge, MA, 1986b.

Christiansen, Henning. Chr as grammar formalism. 2001.

Cowan, Nelson. The magical number 4 in short-term memory:a reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, 24(1), 2001.

Delon, F. Espaces ultramétriques. *J. Symbolic Logic*, 49:405, 1984.

Evans, Nicholas and Stephen Levison. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences*, 32(05):429–448, 2009.

Frazier, Lyn. Sentence processing: A tutorial review. *Attention & Performance*, XII:559–586, 1987.

Guénoche, A. Order distance associated with hierarchy. *J.Classification*, 14:101, 1997.

Haegeman, Liliane. *Introduction to Government and Binding Theory*. Blackwell, Oxford, 1994.

Hammer, Hanno. Tree structure, entropy, and the action principle for neighbourhood topologies. Technical report, Cambridge, DAMTP, 1998.

Higgs, P.G. Overlaps between rna secondary structure. *Phys.Rev.Lett.*, 76:704–707, 1996.

Jackendoff, R.S. *X Syntax. A Study of Phrase Structure*. MIT Press, Cambridge, Mass, 1977.

Jardine, N. and R. Sibson. *Mathematical Taxonomy*. John Wiley and Sons, 1971.

Johnson, N.F. The psychological reality of phrase-structure rules. *Journal of Verbal Learning and Verbal Behaviour*, 4:469–475, 1965.

Karwowski, W. and R.V. Mendes. Hierarchical structures and asymmetric stochastic processes on p-adics and adeles. *J.Math.Phys.*, 35:4637, 1994.

Kayne, Richard S. Unambiguous paths. *Levels of Syntactic Representation*, 5, 1981.

Keenan, E.L. and Bernard Comrie. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, 8:63–99, 1977.

Levelt, W.J.M. Hierarchical clustering algorithms in the psychology of grammar. *Advances in psycholinguistics*, 1970.

Lockward, D.G. *Introduction to Stratification Grammar*. Harcourt Brace Jovanovich Inc.,New York., 1972.

McCloskey, J. *Syntatic Theory*. Cambridge University Press, Cambridge, 1988.

Miller, George A. The magical number seven, plus or minus two:some limits on our capacity to process information. *Psy.Rev.*, 63:81–97, 1956.

Murtagh, F. On ultrametricity, data coding, and computation. *Journal of Classification*, 21:167–184, 2004.

Ogielchi, A.T. and D.L. Stein. Dynamics on ultrametric spaces. *Phys.Rev.Lett.*, 55:1634–1637, 1985.

Parga, N. and M.A. Virasoro. The ultrametric organization of neural net memories. *J.de Physique*, 47:1857, 1986.

Perlman, E.M. and M. Schwarz. The directed polymer problem. *Europhys.Lett.*, page 227, 1992.

Prince, A. and P. Smolensky. Optimality: From neural networks to universal grammar. *Science*, page 1604, 1997.

R. Rammal, G. Toulouse and M.A. Virasoro. Ultrametricity for physicists. *Rev.Mod.Phys.*, 58: 765–788, 1986.

Rissanen, Jorma. A universal prior for integers and estimation by minimum description length. *Annals of Statistics*, 11:416–431, 1983.

Roberts, Mark D. Name strategy: Its existence and implications. *Int.J.Computational Cognition*, 3:1–14, 2005.

Rutten, J.J.M.M. Elements of generalized ultrametric domain theory. *Theor.Comp.Sci.*, 170:349, 1996.

Schweinberger, M. and T.A.B. Snijders. Settings in social networks: A measurement model. *Sociological Methodology*, 23:307–341, 2003.

Shepard, R.N. and P. Arabie. Additive clustering: Representative of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86:87–123, 1979.

Sneath, P.H. and R.R. Sokal. *The Principles and Practice of Numerical Classification*. W.H.Freeman and Company, San Francisco., 1973.

Sorton, George. Introduction to the history of science. *Introduction to the History of Science*, III: 552, 1947.

Vlad, M.O. Fractal time, ultrametric topology and fast relation. *Phys.Lett.*, A189:299–303, 1994.

Weissman, M.B. What is spin glass. *Rev.Mod.Phys.*, 65:829, 1993.

Young, M.R. and W.S. DeSarbo. A parametric procedure for ultrametric tree estimation from conditional rank order proximity data. *Psychometrika*, 60:47, 1995.

Zadrozny, Wlodek. Minimum description length and compositionality. *Computing Meaning*, 1: 113–128, 1999.

Address for correspondence:
Mark D. Roberts
`robemark@gmail`
Flat 44, The Cloisters, 83 London Road,
Guildford, GU1 1FY, United Kingdom