Ultrametric Distance in Syntax

Mark D. Roberts
Department of Mathematics and Applied Mathematics,
University of Cape Town,
Rondebosch 7701,
South Africa
roberts@cosmology.mth.uct.ac.za

January 17, 2000

eprint: http://xxx.lanl.gov/abs/cs.CL/9810012

Comments: 29 pages 44480 bytes, Sixteen EPS diagrams, LaTex2e.

Keyword Index:

Phrase Trees, Ultrametricity, Syntax,

Definitions of Government, Features, $\bar{\mathrm{X}}$ Structure.

ACM Classification:

http://www.acm.org/class/1998/overview.html

I.2.7; J.4; I.2.6

Mathematical Review Classification:

http://www.ams.org/msc/

03D65, 68S05, 92K20, 03D55, 92B10.

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 occurrence of hierarchical structure elsewhere in linguistics is mentioned. The phrase tree can be represented by a matrix and the elements of the matrix can be represented by triangles. The height at which branching occurs is not prescribed in previous syntactic models, but it is by using the ultrametric matrix. In other words the ultrametric approach gives a complete description of phrase trees, unlike previous approaches. 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 isoceles, here it is shown that $\bar{\mathrm{X}}$ structure implies that there are no equilateral triangles. Restricting attention to simple syntax a minimum ultrametric distance between lexical categories is calculated. 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 GOVERNMENT.

1 Introduction

Ultrametrics are used to model any system that can be represented by a bifurcating hierarchical tree. To briefly go through some areas where ultrametrics have been applied. Perhaps the most important application is to taxonomy, Jardine and Sibson (1971) Ch.7 [14], and Sneath and Sokal (1973) [30]. Here the end of a branch of the tree represents a species and the ultrametric distance between them shows how closely the species are related. The actual technique is called the hierarchical cluster method, the method classifies species and also shows how closely species are related. This technique has also been used in semantics, Shepard and Arabie (1979) [29]. The technique can become quite complex because they involve statistical analysis with continuous variates. Ultrametrics have been applied frequently in the theory of spin glass, Weissman [32]. 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 Paiginski (1997) [5], also ultrametrics can describe systems with fast relaxation, Vlad (1994) [31]. Ultrametrics are frequently used in the theory of neural nets, Parga and Virasoro [22]. The dynamics of random walks on ultrametric spaces have been studied, Ogielchi and Stein (1985) [21]. Ultrametrics have been applied to the thermodynamics of macromolecules such as RNA, Higgs (1996) [12]. Bounds on the size of ultrametric structure have been discussed by Baldi and Baun (1986) [1]. From a more theoretical angle, a category theory approach has been elucidated by Rutten (1996) [28], and a model theoretic approach to them given Delon (1984) [8]. The relationship between ultrametric distance and hierarchy is further discussed in Guénoche (1997) [9]. Construction of optimal ultrametric trees is discussed by Young and DeSarbo (1995) [33]. Ultrametrics are related to p-adelic quantities, Karwowski and Mendes (1994) [15]. P-adelic quantities are used in string theory, the way that ultrametrics enters here is explained in sections 10 and 13.4 of Bekke and Freund (1993) [2]. There does not seem to be any straightforward connection of any of the above to the optimization techniques of Prince and Smolensky (1997) [24]. As well as ultrametric trees, there are also *decision trees* Hammer (1998) [11], and the connection between them is still not known. Some of the above ultrametric applications have been reviewed by Rammal *et al* (1986) [25]

For the analysis of syntactic phrase trees the necessary technique is quite simple and is illustrated by the examples in section 2. The examples here mainly follow the examples in Lockward (1972) [18], Kayne (1981) [16], McCloskey (1988) [19], and especially Haegeman (1994) [10].

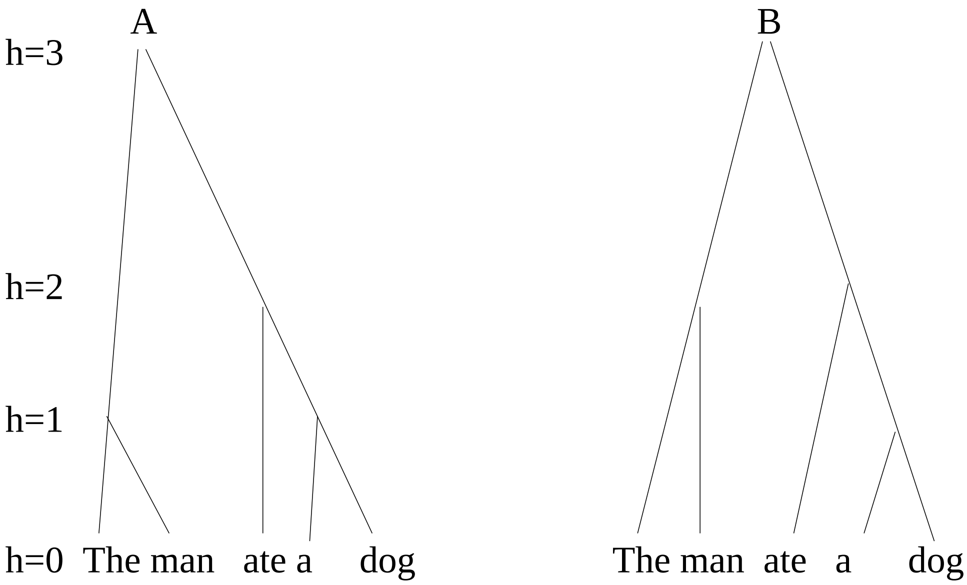

Figure 1: Different syntactic descriptions of “the man ate a dog”

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 the two trees usually are equivalent (perhaps not always McCloskey (1988) [19] footnote 6): however consider the ultrametric distance between 'the' and 'man',

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

this difference does not occur in current syntactic models, and a purpose of an ultrametric model is to disambiguate this. 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 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 to: *firstly* to the closeness indicated by features, *secondly* to the idea that if a sentence is not sufficiently close then there is a BARRIER Chomsky (1986b) [7]. Only the first 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 a lot of other science, for example those topics described in the first paragraph, 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, and perhaps the most important, is that it might be possible to use some sort of minimum distance principle in syntax, indeed it could be this which would have application in other cognitive processes. Minimum description in science go back several hundred years to "*Ockams razor*" or perhaps further. The principle of least action (see for example Bjorken and Drell (1965) [4] §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 (1982) [26] and Zadrozny (2000) [34]. Briefly they assign a lenght 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 lenght, in bits, of the description of the theory, and ii) the lenght, in bits, of data when encoded with the help of the theory.

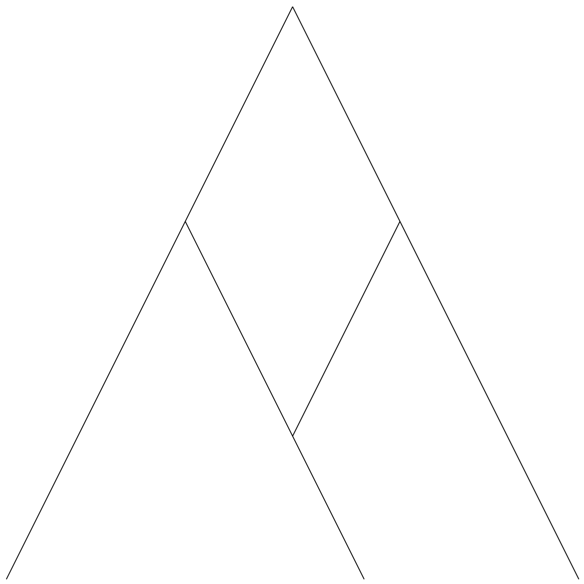

Figure 2: A RETICULATE tree

A RETICULATE tree is a tree in which the branches reconverge, illustrated by **Figure 2**, a NON-RETICULATE tree is a tree in which the branches do not reconverge. N-ARY branching is illustrated by **Figure 3**. BINARY branching is N-ARY branching with $N = 2$. N-ARY branching can be replaced by binary branching if additional layers are used. A SWITCHED tree is a tree in which all the branches are binary. Syntactic phrase trees are NON-RETICULATE and SWITCHED. In most linguistic theories all syntactic phrase trees have $\bar{\mathrm{X}}$ structure, Jackendoff (1977) [13], here attention is restricted to theory which has $\bar{\mathrm{X}}$ structure.

In section 2 it is shown how to represent trees by matrices and triangles. All $\bar{\mathrm{X}}$ triangles are isoceles but not equilateral. In section 3 the matrix $\mathbf{U}$ 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 $\mathbf{F}$ for features is given. $\mathbf{F}$ is not an ultrametric matrix and there appears to be no relation to $\mathbf{U}$. In section 5 it is shown that the notion of C-COMMAND is equivalent to an ultrametric minimum distance. This allows a new definition of government to be

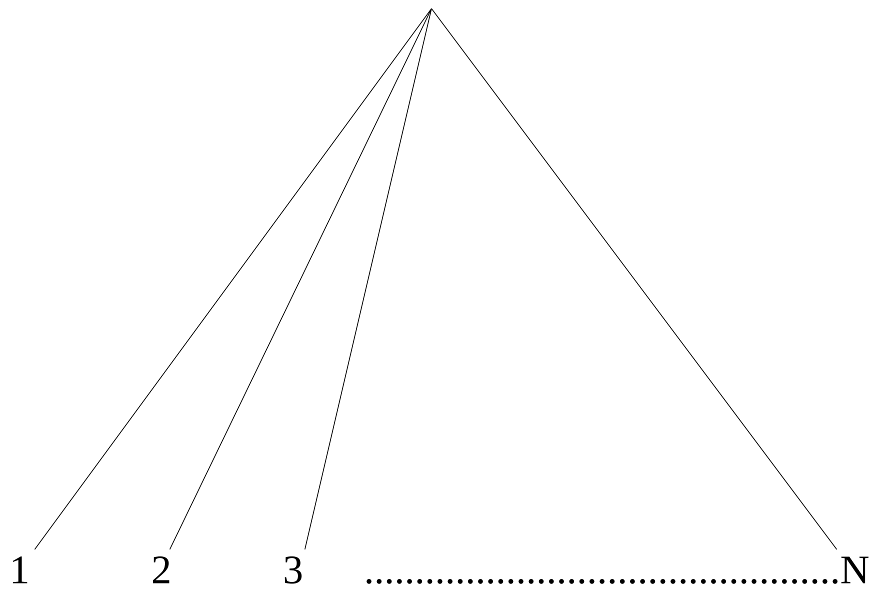

Figure 3: N-ARY branching

given. In the appendix 6 other linguistic hierarchies are discussed. 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*. The appendix 6 very briefly state what these hierarchies are, compare and contrast them, and speculate 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) [20]. A *hierarchy* is an example of a representation as discussed by Roberts (1998) [27].

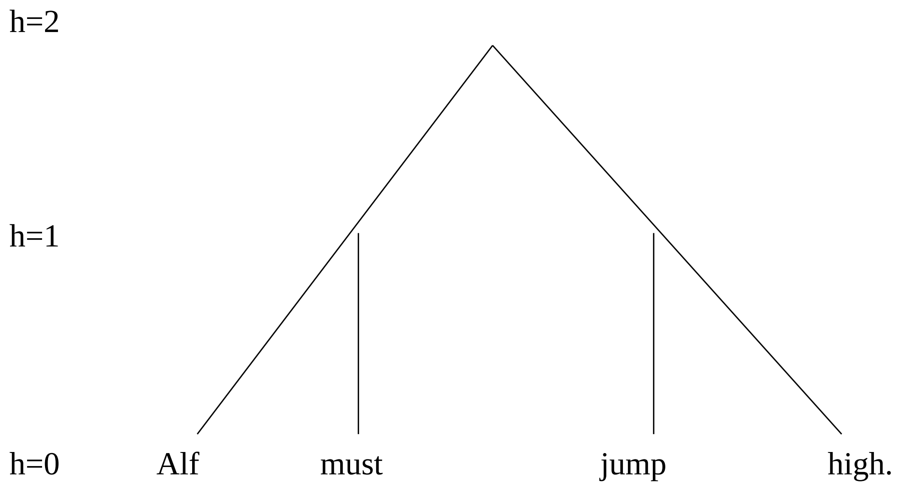

Figure 4: The simplest BINARY tree for "Alf must jump high"

2 $\bar{\mathrm{X}}$ Structure Implies No Equilateral Triangles.

$\bar{\mathrm{X}}$ structure implies BINARY BRANCHING Haegeman (1994) p.139 [10]. To see what this implies for ultrametric distances consider all five species of BINARY BRANCHED tree, the *first* has diagram **Figure 4** and corresponding matrix:

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 1 & 2 & 2 \\ M & & 0 & 2 & 2 \\ J & & & 0 & 1 \\ H & & & & 0 \end{array} \tag{2}$$

respectively. The matrices corresponding to the other *four* BINARY BRANCHED trees are:

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 3 & 3 & 3 \\ M & & 0 & 2 & 2 \\ J & & & 0 & 1 \\ H & & & & 0 \end{array} \tag{3}$$

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 3 & 3 & 3 \\ M & & 0 & 1 & 2 \\ J & & & 0 & 2 \\ H & & & & 0 \end{array} \tag{4}$$

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 2 & 2 & 3 \\ M & & 0 & 1 & 3 \\ J & & & 0 & 3 \\ H & & & & 0 \end{array} \tag{5}$$

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 1 & 2 & 3 \\ M & & 0 & 2 & 3 \\ J & & & 0 & 3 \\ H & & & & 0 \end{array} \tag{6}$$

There are *two* 3-ARY trees with matrices:

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 1 & 1 & 2 \\ M & & 0 & 1 & 2 \\ J & & & 0 & 2 \\ H & & & & 0 \end{array} \tag{7}$$

$$\begin{array}{ccccc} & A & M & J & H \\ A & 0 & 2 & 2 & 2 \\ M & & 0 & 1 & 1 \\ J & & & 0 & 1 \\ H & & & & 0 \end{array} \tag{8}$$

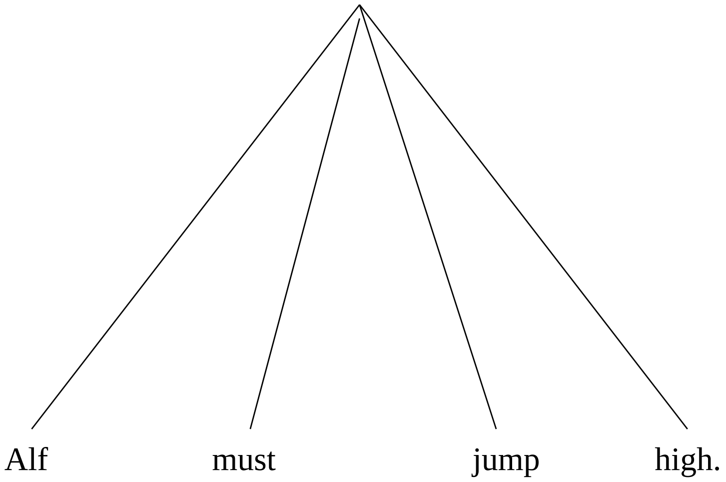

Figure 5: The 4-ARY tree for "Alf must jump high"

and finally there is *one* 4-ARY tree with diagram **Figure 5** and matrix:

$$
\begin{array}{ccccc}
 & A & M & J & H \\
A & 0 & 1 & 1 & 1 \\
M & & 0 & 1 & 1 \\
J & & & 0 & 1 \\
H & & & & 0
\end{array}
\tag{9}
$$

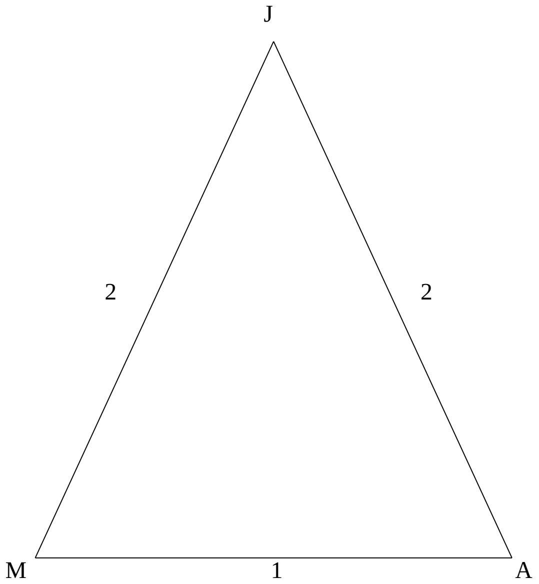

Figure 6: The Isoceles Triangle Representation.

All ultrametric triangles are isoceles, but only some are equilateral. The above suggests that binary branching implies that there are no equilateral triangles. For example from matrix 2, $d(A, M) = 1, d(A, J) = 2, d(M, J) = 2$ giving in the triangle representation **Figure 6**, and from matrix 9, $d(A, M) = 1, d(A, J) = 1, d(M, J) = 1$ giving in the triangle representation **Figure 7**. Formally it is proved that $\bar{\mathrm{X}}$ structure implies that there are no equilateral triangles. Consider the $\bar{\mathrm{X}}$ template **Figure 8**.

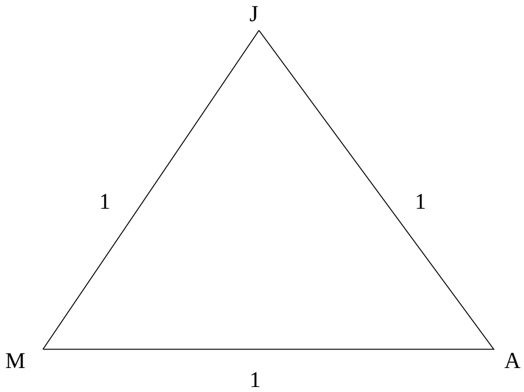

Figure 7: The Equilateral Triangle Representation.

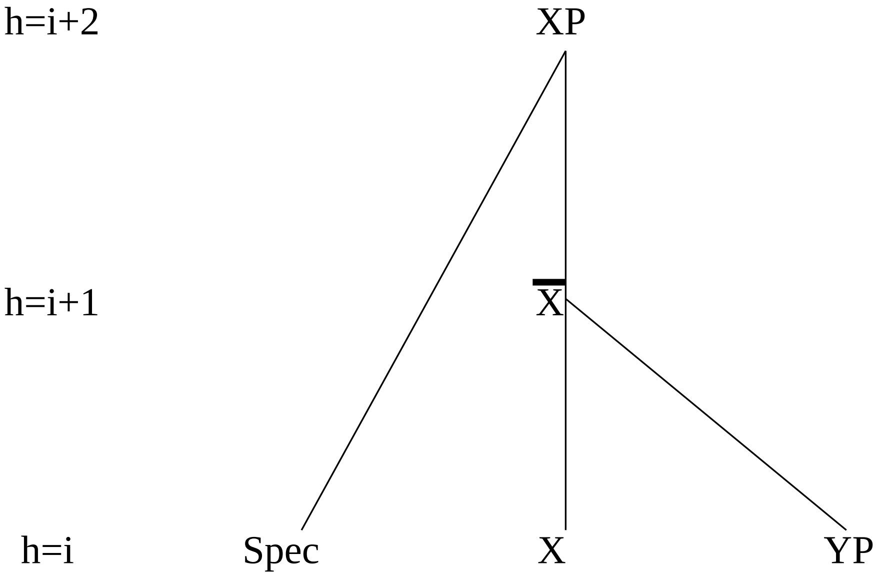

Figure 8: The $\bar{\mathrm{X}}$ Template

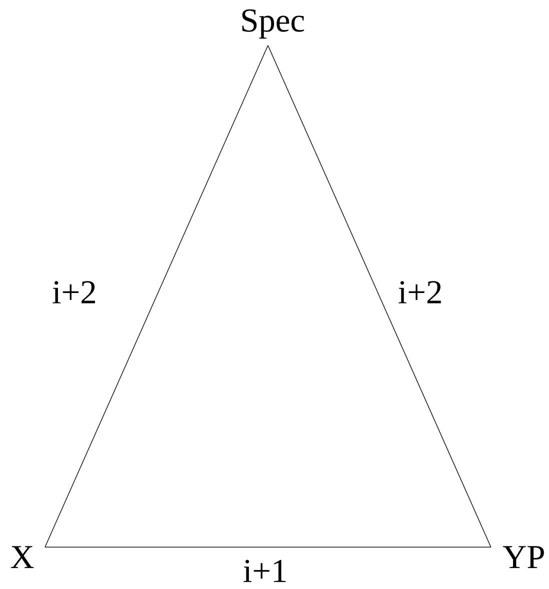

Figure 9: The Triangle Representation of $\bar{\mathrm{X}}$ structure

The matrix representation of this is:

$$\begin{array}{cccc} . & Spec & X & YP \\ Spec & 0 & i+2 & i+2 \\ X & & 0 & i+1 \\ YP & & & 0 \end{array} \tag{10}$$

From this the triangle representation is **Figure 9**. This must be isoceles but not equilateral.

3 The minimum ultrametric distance between lexical categories.

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 principle of least action in physics, see for example Bjorken and Drell (1965) [4] §11.2. The MINIMUM DISTANCE PRINCIPLE implies that the correct tree for the example in the introduction is **Figure**

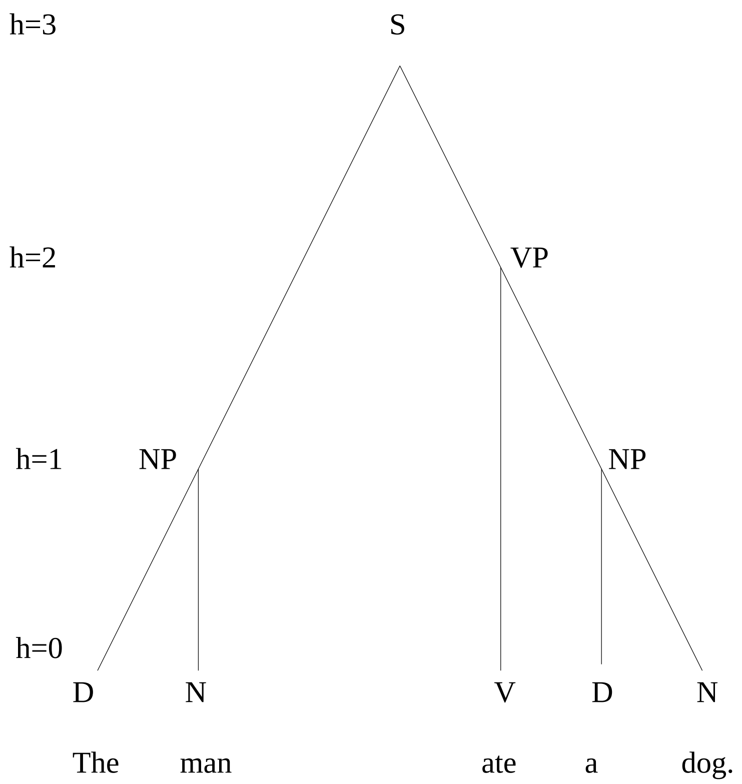

Figure 10: The Correct Tree for the Example in the Introduction.

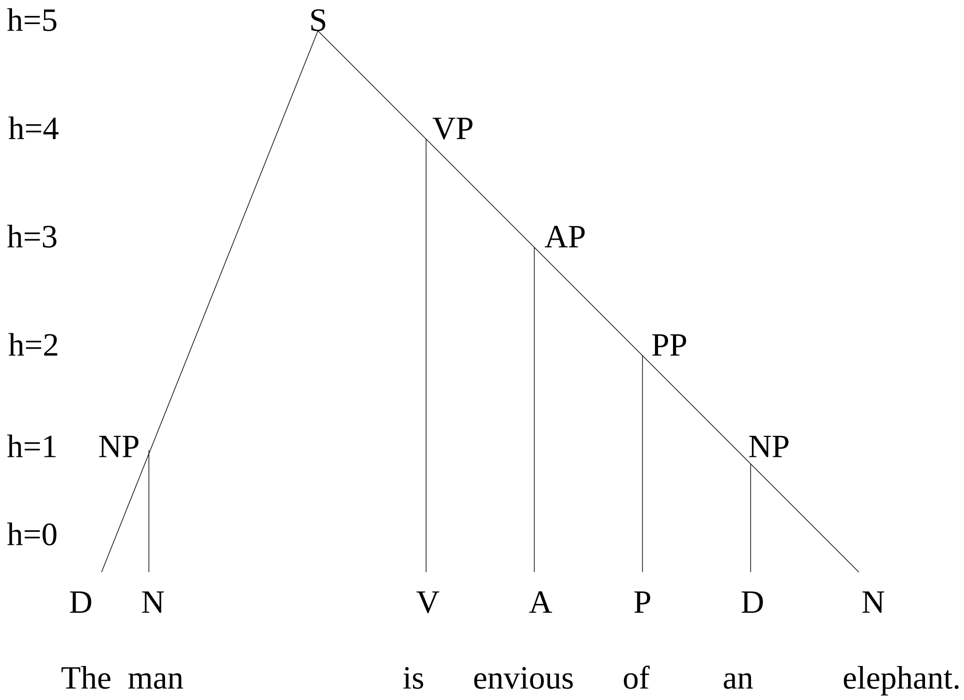

Figure 11: The Distance Between Verbs and Adjectives.

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** given below, but will have an effect when the analysis is extended to θ theory. 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

$$\mathbf{U} = \begin{array}{c|ccccc} & D & N & V & A & P \\ \hline D & 0 & 1 & 2 & 2 & 2 \\ N & & 0 & 2 & 2 & 2 \\ V & & & 0 & 4 & 3 \\ A & & & & 0 & 3 \\ P & & & & & 0 \end{array} \tag{11}$$

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

$$\mathbf{I} = \begin{array}{ccccc} 0 & i & i & i & \ldots \\ & 0 & i+1 & i+1 & \ldots \\ & & 0 & i+2 & \ldots \\ & & & 0 & \ldots \end{array} \tag{12}$$

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

4 Features

This section investigates whether there is a general framework which can describe the preceding and also features. Haegeman (1994) [10] p.146 gives the following diagram for features:

$$\begin{array}{rcc} Noun: & +N & -V \\ Verb: & -N & +V \\ Adj.: & +N & +V \\ Pre.: & -N & -V \end{array} \tag{13}$$

This can be represented by the matrix

$$
\begin{array}{rcc}
 & N & V \\
Noun: & +1 & -1 \\
Verb: & -1 & +1 \\
Adj.: & +1 & +1 \\
Pre.: & -1 & -1
\end{array}
\tag{14}
$$

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

$$
\mathbf{F} =
\begin{array}{ccccc}
 & N & V & A & P \\
N & +1 & -1 & +1 & -1 \\
V & -1 & +1 & +1 & -1 \\
A & +1 & +1 & +1 & -1 \\
P & -1 & -1 & -1 & +1
\end{array}
\tag{15}
$$

which is singular as its determinant vanishes. There appears to be no relation between $\mathbf{F}$ and $\mathbf{U}$. Using the Pauli matrices (see for example Bjorken and Drell (1965) [4] p.378)

$$
I =
\begin{array}{cc}
1 & 0 \\
0 & 1
\end{array}
\tag{16}
$$

$$
\sigma^1 =
\begin{array}{cc}
0 & 1 \\
1 & 0
\end{array}
\tag{17}
$$

$$
\sigma^2 =
\begin{array}{cc}
0 & -i \\
i & 0
\end{array}
\tag{18}
$$

$$\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{19}$$

F can be expressed as

$$\mathbf{F} = \begin{pmatrix} I - \sigma^1 & -i\sigma^2 + \sigma^3 \\ +i\sigma^2 + \sigma^3 & I - \sigma^1 \end{pmatrix} \tag{20}$$

however this does not correspond in any straightforward way to any of the Dirac matrices (see for example Bjorken and Drell (1965) [4] page 378) in standard representations.

5 A New Definition of C-Command

Recall the following definitions in Haegeman [10]:
Definition [10] 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 line 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 [10] p.83)
the phrase tree **Figure 12** gives the 'dominates' matrix

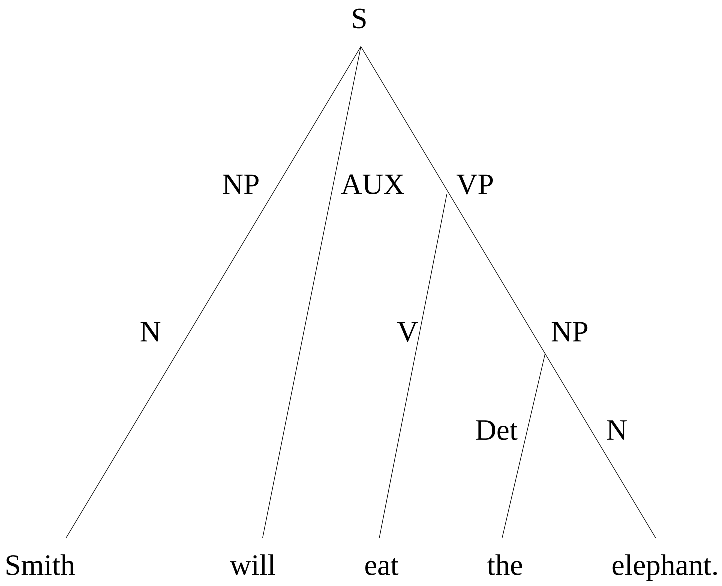

Figure 12: Illustration of DOMINATES

$$\mathbf{D} = \begin{array}{cccccccccc} & S & NP(S) & N(S) & AUX & VP & V & NP(E) & Det & N(E) \\ S & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ NP(S) & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ N(S) & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ AUX & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ VP & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ V & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ NP(E) & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ Det & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ N(E) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \tag{21}$$

where 1 indicates **A** dominates **B** and 0 indicates that it does not.
Definition [10] 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**.

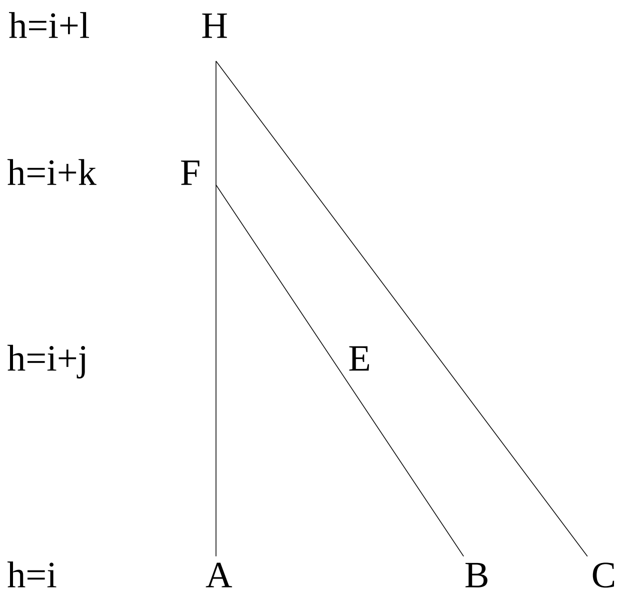

Figure 13: Example of C-COMMANDS.

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 criteria 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 criteria for dominance also needs to be adjusted, if no higher node is allowed the set of C-COMMANDS is again empty. Chomsky (1986a) [6] p.161 approaches the subject in a different manner using maximal projections.
Example: **Figure 13** in the figure $0 < j < k < l$. The corresponding ultrametric matrix is

$$\mathbf{U} = \begin{array}{ccccc} & A & B & C & D \\ A & 0 & k & k & l \\ B & & 0 & j & l \\ C & & & 0 & l \\ D & & & & 0 \end{array} \tag{22}$$

The C-COMMAND matrix **CM** is

$$\mathbf{CM} = \begin{array}{ccccc} A & 1 & 1 & 1 & 0 \\ B & 0 & 1 & 1 & 0 \\ C & 0 & 1 & 1 & 0 \\ D & 1 & 1 & 1 & 1 \end{array} \tag{23}$$

where 1 indicates **A**,**B**,... C-COMMANDS **A**,**B**,... and 0 indicates that it does not.
Definition [10] p.134
The total of all the nodes C-COMMANDED by an element is the C-DOMAIN of that element.
Definition [10] 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'.

Now let $\mathbf{D(A)}$ be the set of all the ultrametric distances to other nodes at the same height and let $\mathbf{M(A)}$ be the set of these which have the smallest value.
Call $\mathbf{M(A)}$ the CU-DOMAIN of **A** and say **A** CU-COMMANDS all $\mathbf{B}\varepsilon\mathbf{M}(\mathbf{A})$, this is illustrated by **Figure 14**.

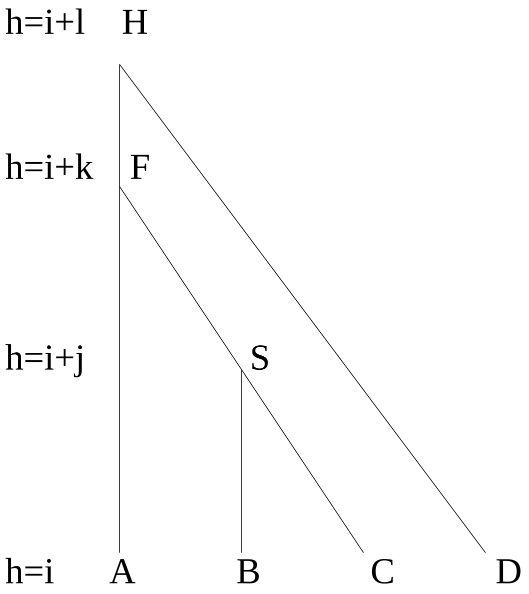

Figure 14: Illustration of the Theorem.

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 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 thus the result holds in general.

This allows a new definition of GOVERNMENT. **A** GOVERNS **B** iff:
i) **A** is a GOVERNOR.
ii) both $\mathbf{A}\varepsilon\mathbf{M}(\mathbf{B})$ and $\mathbf{B}\varepsilon\mathbf{M}(\mathbf{A})$.
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.

SU > DO > IO > OBL > GEN > OCOMP

Figure 15: The *accessibility hierarchy*

6 Appendix: Other Linguistic Hierarchies

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) [17] 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 (*AHC's*) 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 (*PRC's*) 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.

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.

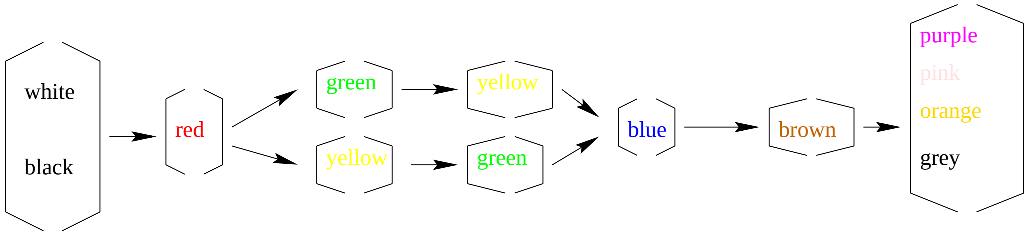

Figure 16: The Berlin-Kay Universal Colour Partial Ordering

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-Kay (1969) [3] 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 speak 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.

7 Acknowledgements

This work has been supported in part by the South African Foundation for Research and Development (FRD).

References

[1] Baldi,P. and BaumE.B.(1986) ♣ 1.1
Bounds on the Size of
Ultrametric Structures.
Phys.Rev.Lett. **56**, 1598-1600.

[2] Brekke,L. and Freund,P.G.O.(1993) ♣ 1.1
P-adelic numbers in Physics.
Math.Rev.94h:11115
Phys.Rep. **233**, 1-66,
especially sections 10 and 13.4.

[3] Berlin,B and Kay,P.(1969) ♣ 6.2
Basic Colour Terms,
University of California Press.

[4] Bjorken,J.D. and Drell,S.D.(1965) ♣ 1.3, 3.1, 4.1, 4.1
Relativistic Quantum Fields,
Math.Rev.66:5092
McGraw Hill, New York. p.378.

[5] Boettcher,S. and Paczuski,M.(1997) ♣ 1.1
Aging in a model of
Self-Organized Criticality.
`cond-mat/9603120`
Phys.Rev.Lett. **79**, 889.

[6] Chomsky, N.(1986a) ♣ 5.1
Knowledge of Language, Its Nature, Orgin and Use.
Praeger Publishers, New York.

[7] Chomsky, N.(1986b) ♣ 1.3
Barriers,
MIT Press, Cambridge, MA.

[8] Delon,F.(1984) ♣ 1.1
Espaces Ultramétriques.
Math.Rev.85i:03116
J. Symbolic Logic **49**, 405.

[9] Guénoche,A.(1997) ♣ 1.1
Order distance associated with hierarchy.
Math.Rev.98d:62108
J.Classification **14**, 101.

[10] Haegeman, L.(1994) ♣ 1.2, 2.1, 4.1, 5.1
Introduction to Government
and Binding Theory,
First Edition (1991),
Second Edition (1994),
Blackwell, Oxford.

[11] Hammer,H.(1998) ♣ 1.1
Tree structure, Entropy, and the Action Principle for Neighbourhood Topologies.
`hep-th/9811118`

[12] Higgs,P.G.(1996) ♣ 1.1
Overlaps between RNA
Secondary Structure.
Phys.Rev.Lett. **76**, 704-707.

[13] Jackendoff,R.S(1977) ♣ 1.4
$\bar{\mathrm{X}}$ Syntax. A Study of Phrase Structure.
MIT Press, Cambridge, Mass.

[14] Jardine,N. and Sibson,R.(1971) ♣ 1.1
Mathematical Taxonomy,
John Wiley and Sons.

[15] Karwowski,W. and Mendes,R.V.(1994) ♣ 1.1
Hierarchical structures and asymmetric stochastic processes on p-adics and adeles.
Math.Rev.95h:60166
J.Math.Phys. **35**, 4637.

[16] Kayne,R.S.(1981) ♣ 1.2
Unambiguous Paths,
Ch.5 in
Levels of Syntactic Representation,
Edited by Robert May and Jan Koster,
Foris Publications, Dordrecht.

[17] Keenan,E.L. and Comrie,B.(1977) ♣ 6.1
Noun Phrase Accessibility
and Universal Grammar.
Linguistic Inquiry **8**, 63-99.

[18] Lockward,D.G.(1972) ♣ 1.2
Introduction to Stratification Grammar.
Harcourt Brace Jovanovich Inc.,
New York.

[19] McCloskey,J.(1988) ♣ 1.2, 1.3
Syntatic Theory,
p.18 in Linguistics:
The Cambridge Survey,
Cambridge University Press, Cambridge.

[20] Miller,G.A.(1956) ♣ 1.5
The magical number seven, plus or minus two: some limits on our capacity to process information.
`http://cogprints.soton.ac.uk`
`/abs/psy/199807022`
Psy.Rev. **63**,81-97.

[21] Ogielchi,A.T. and Stein,D.L.(1985) ♣ 1.1
Dynamics on Ultrametric Spaces.
Math.Rev.86j:82051
Phys.Rev.Lett. **55**, 1634-1637.

[22] Parga,N. and Virasoro,M.A.(1986) ♣ 1.1
The Ultrametric Organization
of Neural Net Memories.
J.de Physique **47**, 1857.

[23] Perlman,E.M. and Schwarz,M.(1992) ♣ 1.1
The Directed Polymer Problem.
Europhys.Lett. **17**, 227.

[24] Prince,A. and Smolensky,P.(1997) ♣ 1.1
Optimality: From Neural Networks to Universal Grammar.
Science **275**,1604.

[25] Rammal,R., Toulouse,G. and Virasoro,M.A.(1986) ♣ 1.1
Ultrametricity for Physicists.
Math.Rev.87k:82105
Rev.Mod.Phys. **58**, 765-788.

[26] Rissanen,Jorma(1983) ♣ 1.3
A universal prior for integers and estimation by minimum description lenght.
Math.Rev.85i:62002
Annals of Statistics **11**,416-431.

[27] Roberts.M.D. ♣ 1.5
Name Strategy:
Its Existence and Implications.
`cs.CL/9812004`

[28] Rutten,J.,J.,M.,M.,(1996) ♣ 1.1
Elements of Generalized Ultrametric Domain Theory.
Theor.Comp.Sci. **170**, 349.

[29] Shepard,R.N. and Arabie,P.(1979) ♣ 1.1
Additive Clustering: Representative of Similarities as Combinations of Discrete Overlapping Properties.
Psychological Review **86**, 87-123, especially page 90.

[30] Sneath,P.H. and Sokal,R.R.(1973) ♣ 1, 1
The Principles and Practice
of Numerical Classification.
W.H.Freeman and Company,
San Francisco.

[31] Vlad, M.O.(1994) ♣ 1.1
Fractal time, ultrametric topology
and fast relation.
Math.Rev.95a:82061
Phys.Lett. **A189**, 299-303.

[32] Weissman,M.,B.(1993) ♣ 1.1
What is Spin Glass.
Rev.Mod.Phys. **65**, 829.

[33] Young,M.R. and DeSarbo,W.S.(1995) ♣ 1.1
A Parametric procedure for Ultrametric tree estimation from conditional rank order proximity data.
Psychometrika **60**,47.

[34] Zadrozny,Wlodek (2000) ♣ 1.3
Minimum Description Lenght and Compositionality.
`cs.CL/0001002`