Ultrametric Distance in Syntax

Anonymous Version

29th of December 2011

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 isosceles. Here it is shown that $\bar{\text{X}}$ 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.

Contents

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 [21], and Sneath and Sokal (1973) [43]. 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) [42]. 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 [46]. 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) [45]. Ultrametrics are used in the theory of neural nets, Parga and Virasoro [33]. The dynamics of random walks on ultrametric spaces have been studied, Ogielchi and Stein (1985) [32]. Ultrametrics have been applied to the thermodynamics of macromolecules such as RNA, Higgs (1996) [19], the directed polymer problem Perlman and Schwarz (1992)[34], and sociology Schweinberger and Snijders (2003)[41]. 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) [40], a model theoretic approach to ultrametrics is given by Delon (1984) [11], and ultrametric might be related to T-theory, Dress et al (1996) [13]. The relationship between ultrametric distance and hierarchy is further discussed in Guénoche (1997) [15]. Construction of optimal ultrametric trees is discussed by Young and DeSarbo (1995) [47]. Ultrametrics are related to p-adelic quantities, Karwowski and Mendes (1994) [24], Murtagh (2004) [31] and Dragovich (2009) [12]. P-adelic quantities are used in string theory: the way that ultrametrics enters here is explained in §10&§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) [35]. As well as ultrametric trees, there are also *decision trees* Hammer (1998) [18], and the connection between them is still not known. Some of the above ultrametric applications have been reviewed by Rammal *et al* (1986) [36].

1.2 Ultrametric Inequalities

There is the following relationship between trees and ultrametrics. An N-leaf edge(node)-weighted tree corresponds to an $N \times N$ 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
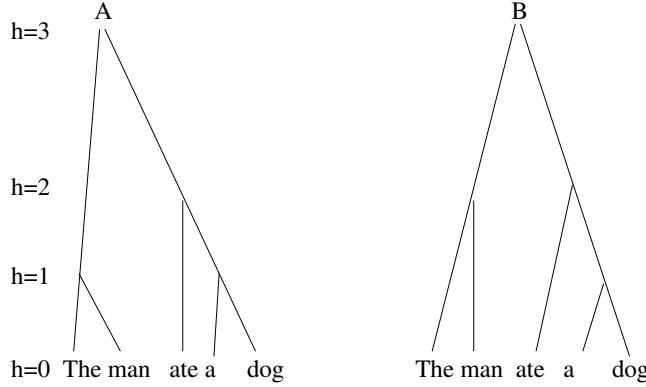
3

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

the usual sense.

$$\forall x, y, z \quad M_{xy} = 0 \quad \text{if} \quad x = y \tag{1}$$

$$M_{xy} > 0 \quad \text{for} \quad x \neq y \tag{2}$$

$$M_{xy} = M_{yx} \tag{3}$$

$$M_{xy} \leq M_{xz} + M_{zy}. \tag{4}$$

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) [22] and Levelt (1970) [27]. The phrase tree approach has critics Botha (1989) [6]; also Evans and Levinson (2009) [14] question the existence of language universals. The examples here mainly follow the examples in Lockward (1972) [28], Kayne (1981) [25], McCloskey (1988) [29], and especially Haegeman (1994) [17]. 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) [29] footnote 6; however consider the ultrametric distance between 'the' and 'man',

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

4

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) [8] 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 §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   Ockam's Razor

Minimum description in science goes back several hundred years to "*Ockam's razor*" or perhaps further, see for example Sorton (1947) [44] page 552. 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) [37] and Zadrozny (2000)
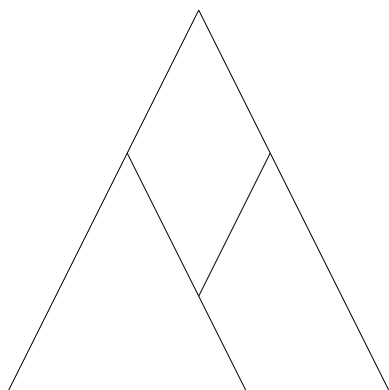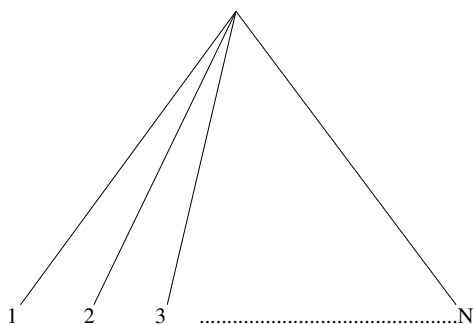
Figure 2: A RETICULATE tree

Figure 3: N-ARY branching

[48]. 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) [9] discusses how constraint handling rules (CHR) can be applied to grammars. This can be thought of as a minimizing procedure.

1.5 Reticulate & N-ary Trees

A RETICULATE tree is a tree in which there are one or more sets of reconvergent branches, 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 syntac-

tic phrase trees have $\bar{\text{X}}$ structure, Jackendoff (1977) [20]. $\bar{\text{X}}$ structure implies binary branching, see subsection 2.3 and figure 8 Here attention is restricted to theory which has $\bar{\text{X}}$ structure.

1.6   Sectional Contents

In section 2 it is shown how to represent trees by matrices and triangles. All $\bar{\text{X}}$ triangles are isosceles 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 given. In appendix 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 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) [30]; for a more recent reference see Cowan (2001) [10]. A *hierarchy* is an example of a representation as discussed by Roberts (2005) [39].

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

2.1   Binary and N-ary Branching for simple sentences

$\bar{\text{X}}$ structure implies BINARY BRANCHING Haegeman (1994) p.139 [17], 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 [17] diagram 84a) and corresponding matrix:

$$First = \begin{array}{c|ccccc} \bullet & A & M & J & H \\ A & 0 & 1 & 2 & 2 \\ M & . & 0 & 2 & 2 \\ J & . & . & 0 & 1 \\ H & . & . & . & 0 \end{array} \tag{7}$$

respectively, where A, M,... are short for 'Alf', 'must'. The matrices corresponding to the other *four* BINARY BRANCHED trees are (compare Haegeman
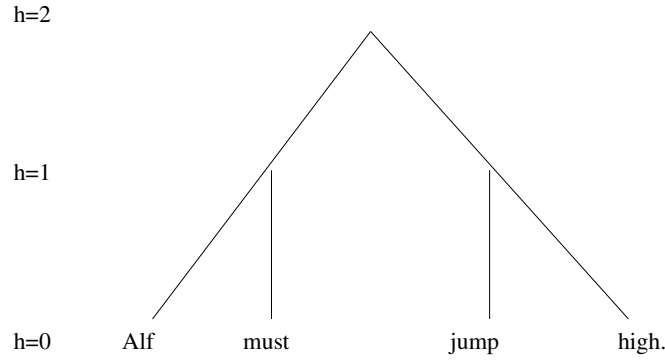
Figure 4: The simplest BINARY tree for "Alf must jump high"

(1994) p.141&142 [17] diagrams 84b,c,d,e):

$$
Second =
\begin{array}{c|cccc}
\bullet & A & M & J & H \\
A & 0 & 3 & 3 & 3 \\
M & . & 0 & 2 & 2 \\
J & . & . & 0 & 1 \\
H & . & . & . & 0 \\
\end{array}
\tag{8}
$$

$$
Third =
\begin{array}{c|cccc}
\bullet & A & M & J & H \\
A & 0 & 3 & 3 & 3 \\
M & . & 0 & 1 & 2 \\
J & . & . & 0 & 1 \\
H & . & . & . & 0 \\
\end{array}
\tag{9}
$$

$$
Fourth =
\begin{array}{c|cccc}
\bullet & A & M & J & H \\
A & 0 & 2 & 2 & 3 \\
M & . & 0 & 1 & 3 \\
J & . & . & 0 & 3 \\
H & . & . & . & 0 \\
\end{array}
\tag{10}
$$

$$
Fifth =
\begin{array}{c|cccc}
\bullet & A & M & J & H \\
A & 0 & 1 & 2 & 3 \\
M & . & 0 & 2 & 3 \\
J & . & . & 0 & 3 \\
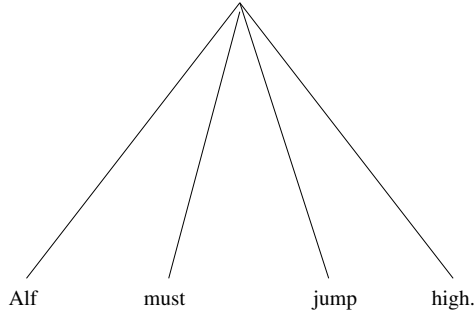H & . & . & . & 0 \\
\end{array}
\tag{11}
$$

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

There are *two* 3-ARY trees with matrices (compare Haegeman (1994) p.142 [17] diagrams 84g and 84h):

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

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

and finally there is *one* 4-ARY tree (compare Haegeman (1994) p.142 [17] diagram 84f) with diagram **Figure 5** and matrix:

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

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 $\bar{X}$ structure implies that there are no equilateral triangles.
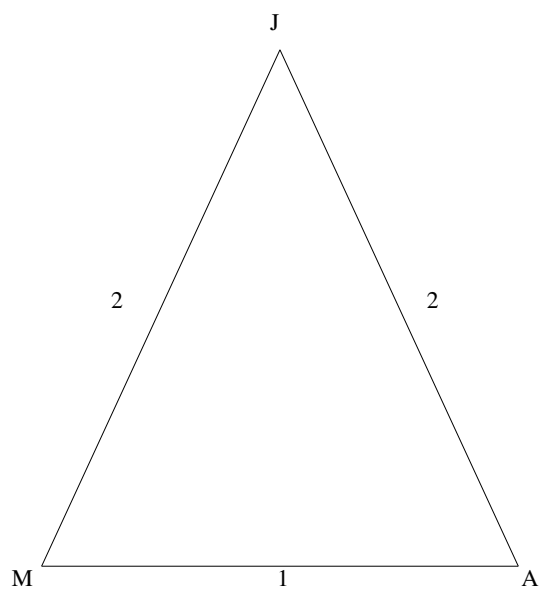
9

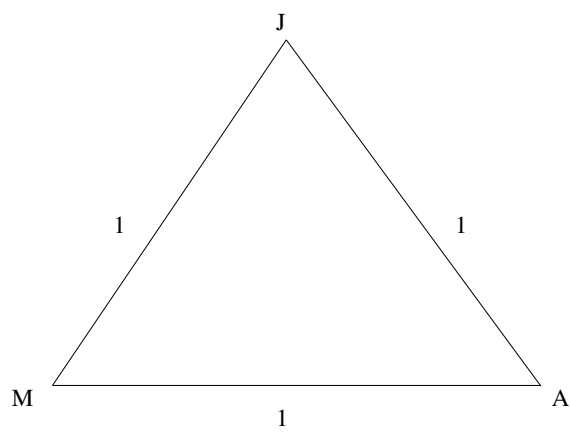Figure 6: The isosceles triangle representation.

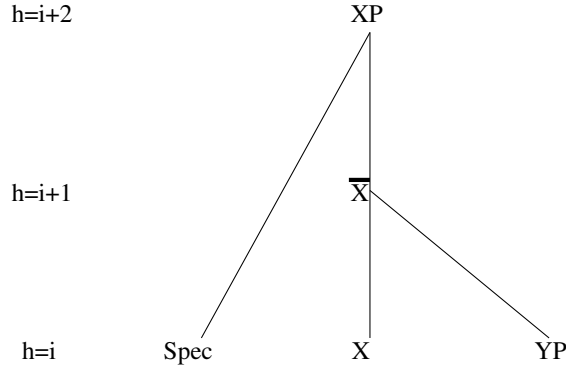Figure 7: The equilateral triangle representation.

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

2.3 The $\bar{\text{X}}$ Template

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

$$\bar{\mathbf{X}} = \begin{matrix} & \bullet & Spec & X & YP \\ Spec & 0 & i+2 & i+2 \\ X & . & 0 & i+1 \\ YP & . & . & 0 \end{matrix} \tag{15}$$

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 subsection 1.4. A current psycholinguistic model of sentence production is the garden path model, see for example Frazier (1987) [16] .Part of this model requires *the minimal attachment principle*, which is "do not postulate unnecessary nodes.": 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 (1995) §3.2.3 [17]. 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
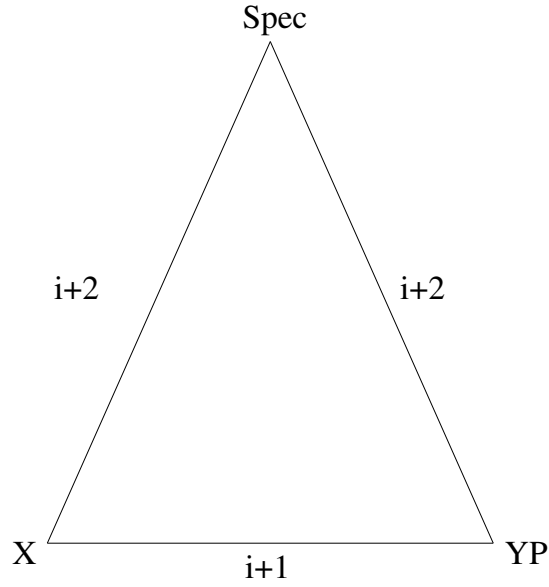
Figure 9: The triangle representation of $\bar{\mathrm{X}}$ structure

distance matrix

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

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

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

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 $\mathbf{U}$ is constructed from the syntactical representations of several sentences.

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

Figure 11: The Distance Between Verbs and Adjectives.

4 Features

4.1 No square matrix representation of features

This section investigates whether there is a general framework which can describe both the preceding and also "features". Roughly speaking the idea behind features is that adjectives and prepositions have qualities in them that are associated with ve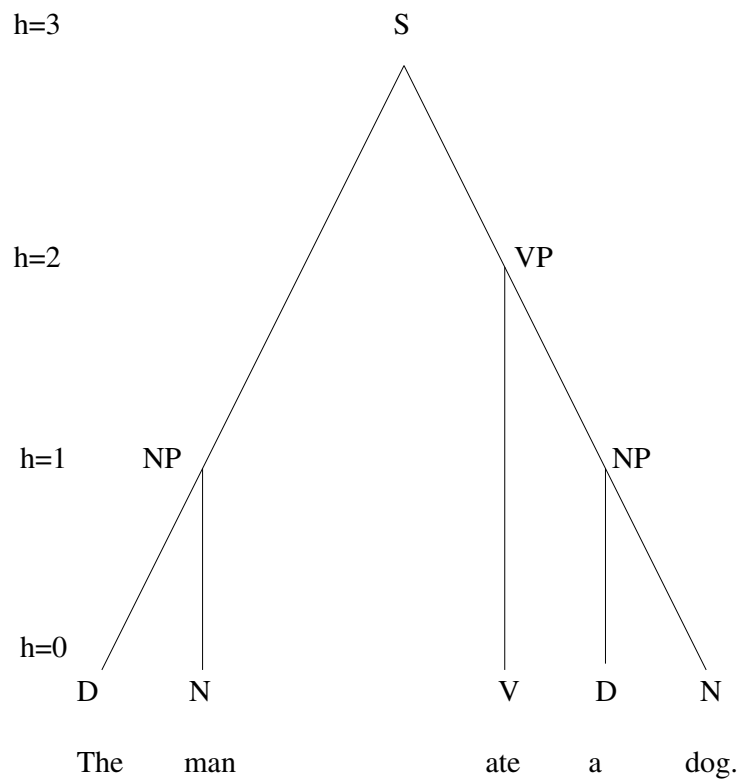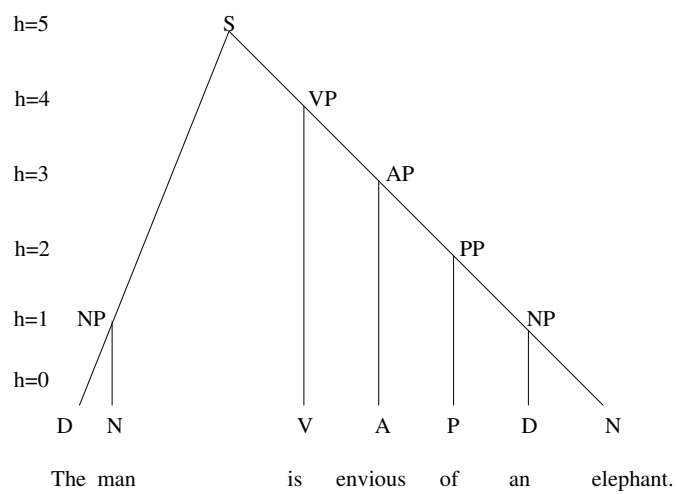rbs and nouns, as can be seen from 18 adjectives have +N & +V and prepositions have -N & -V. Haegeman (1994) [17] p.146 gives the following 18 diagram for features:

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

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

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

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

$$
\mathbf{F} =
\begin{array}{ccccc}
\bullet & N & V & A & P \\
N & +1 & -1 & +1 & -1 \\
V & -1 & +1 & +1 & -1 \\
A & +1 & +1 & +1 & -1 \\
P & -1 & -1 & -1 & +1
\end{array}
\tag{20}
$$

which is singular as its determinant vanishes. There appears to be no relation between matrix $\mathbf{F}$ 20 and matrix $\mathbf{U}$ 16. Using the Pauli matrices (see for example Bjorken and Drell (1965) [4] p.378)

$$
I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \ \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{21}
$$

$\mathbf{F}$ can be expressed as

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

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.
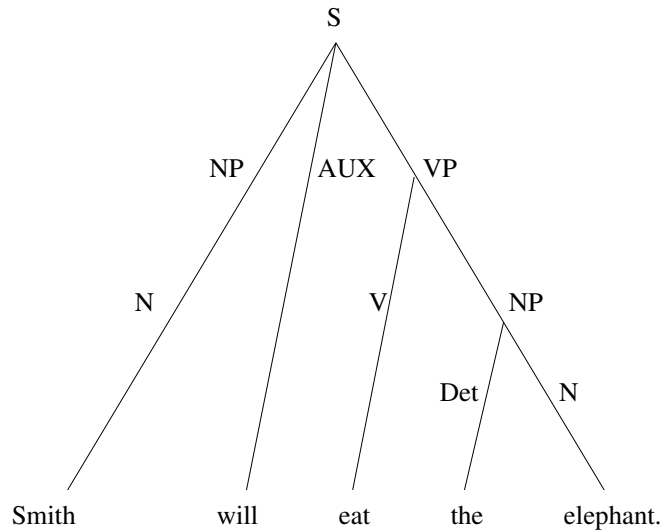
Figure 12: Illustration of DOMINATES

5 Ultrametric Approach to Government

Recall the following definitions in Haegeman [17]:

5.1 Definition of DOMINATES.

Definition [17] 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 [17] p.83)The phrase tree **Figure 12** gives the 'dominates'

matrix:

$$
\mathbf{D} =
\begin{array}{c|ccccccccc}
\bullet & S & NP(S) & N(S) & AUX & VP & V & NP(E) & Det & N(e) \\
S & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
NP(S) & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
N(S) & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
AUX & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
VP & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
V & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
NP(E) & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
Det & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
N(E) & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\tag{23}
$$

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

5.2 Definition of C-command

Definition [17] 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) [7] p.161 approaches the subject in a different manner using maximal projections.
Example:Figure 13 in the figure $0 < j < k < l$. The corresponding ultra-

metric matrix is

$$
\mathbf{U} =
\begin{array}{c|cccc}
\bullet & A & B & C & D \\
A & 0 & k & k & l \\
B & . & 0 & j & l \\
C & . & . & 0 & l \\
D & . & . & . & 0 \\
\end{array}
\tag{24}
$$

The C-command matrix **CM** is

$$
\mathbf{CM} =
\begin{array}{c|cccc}
\bullet & A & B & C & D \\
A & 1 & 1 & 1 & 0 \\
B & 0 & 1 & 1 & 0 \\
C & 0 & 1 & 1 & 0 \\
D & 1 & 1 & 1 & 1 \\
\end{array}
\tag{25}
$$
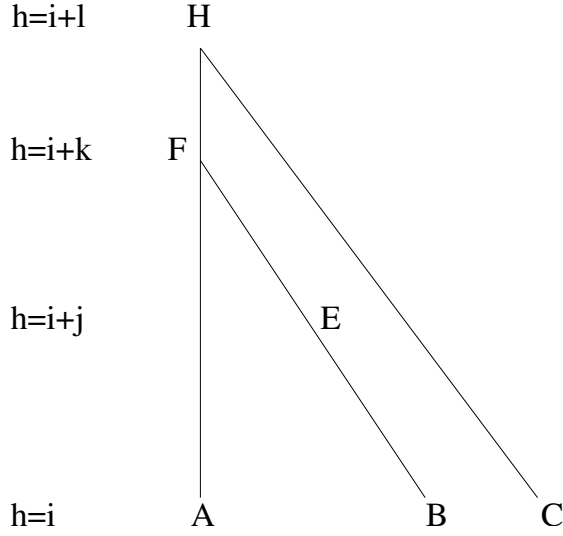
Figure 13: Example of C-COMMANDS.

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 [17] p.134
The total of all the nodes C-COMMANDED by an element is the C-DOMAIN of that element.
Definition [17] 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**.

17

Figure 14: Illustration of the Theorem.

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

Theorem:
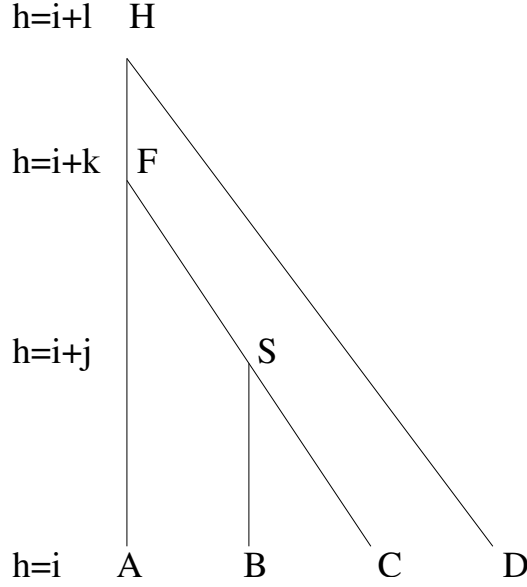The sets **A** C-COMMANDS **B** and A CU-COMMANDS **B** are identical, likewise the C-DOMAIN and the CU-DOMAIN.

Proof:
From the i) part of the definition of C-COMMAND $h(A) = h(B)$, so that we are only 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**ε**M(B)** and **B**ε**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 ultra-

$$SU \quad > \quad DO \quad > \quad IO \quad > \quad OBL \quad > \quad GEN \quad > \quad OCOMP$$

Figure 15: The *accessibility hierarchy*

metric requirement that distances be minimal.

6 Conclusion

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

7 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) [26] 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.
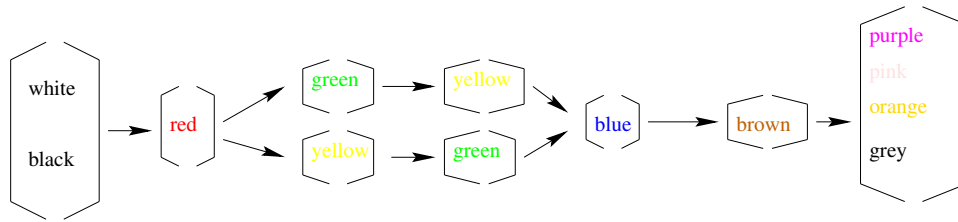
Figure 16: The Berlin-Kay Universal Colour Partial Ordering

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.

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 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.

References

[1] P. Baldi and E.B. Baum (1986), Bounds on the Size of Ultrametric Structures, *Phys.Rev.Lett.* **56**, 1598-1600. 3

[2] L. Brekke and Peter G.O. Freund (1993), P-adelic numbers in Physics, *Phys.Rep.* **233**, 1-66, especially sections 10 and 13.4. 3

[3] Brent Berlin and Paul Kay (1969), Basic Colour Terms, University of California Press. 20

[4] James D. Bjorken and Sidney D. Drell (1965), Relativistic Quantum Fields, McGraw Hill, New York. p.378. 5, 14

[5] S. Boettcher and M. Paczuski (1997), Aging in a model of Self-Organized Criticality, `cond-mat/9603120` *Phys.Rev.Lett.* **79**, 889.3

[6] Rudolf P. Botha (1989), Challenging Chomsky: The Generative Garden game, Blackwell Oxford, ISBN 0-631-16621-1. 4

[7] Noam Chomsky (1986a), Knowledge of Language, Its Nature, Orgin and Use, Praeger Publishers, New York. 16

[8] Noam Chomsky (1986b), Barriers, MIT Press, Cambridge, MA. 5

[9] Henning Christiansen (2001), CHR as Grammar Formalism. `cs.CL/0106059` 6

[10] Nelson Cowan (2001), The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity, *Behavioral & Brain Sciences* **24 (1)**: XXX-XXX. 7

[11] F. Delon (1984), Espaces Ultramétriques, *Math.Rev.85i:03116 J. Symbolic Logic* **49**, 405. 3

[12] B. Dragovich, A.Yu. Khrennikov, S.V. Kozyrev and I.V. Volovich (2009), On p-adic mathematical physics, p-Adic Numbers, *Ultrametric Analysis, and Applications Journal*,**1**, 1-90. 3

[13] Andreas Dress, Vincent Moulton and Werner Terhalle (1996)
T-theory: An Overview.
Europ. J. Combinatorics **17**, 161-175. 3

[14] Nicholas Evans & Stephen Levison
The Myth of Language Universlas: Language diversity and its importance for cognitive science.
To be published in *Behavioural and Brain Sciences.* 4

[15] A. Guénoche (1997), Order distance associated with hierarchy, *J.Classification* **14**, 101. 3

[16] Lyn Frazier (1987), Sentence Processing: A Tutorial Review, *Attention & Performance***XII**,p.559-586, The Psychology of Reading, edited Max Coltheart. 11

[17] Liliane Haegeman (1994), Introduction to Government and Binding Theory, First Edition (1991), Second Edition (1994), Blackwell, Oxford. 4, 7, 8, 9, 11, 14, 15, 16, 17

[18] Hanno Hammer (1998), Tree structure, Entropy, and the Action Principle for Neighbourhood Topologies, `hep-th/9811118` 3

[19] P.G. Higgs (1996), Overlaps between RNA Secondary Structure, *Phys.Rev.Lett.* **76**, 704-707. 3

[20] R.S. Jackendoff (1977), $\bar{\text{X}}$ Syntax. A Study of Phrase Structure, MIT Press, Cambridge, Mass. 7

[21] N. Jardine and R. Sibson (1971), Mathematical Taxonomy, John Wiley and Sons. 3

[22] N.F. Johnson (1965), The psychological reality of phrase-structure rules, *Journal of Verbal Learning and Verbal Behaviour***4**,469-475. 4

[23] Irving Kaplansky (1977), Set Theory and Metric Spaces, AMS Chelsea Publishing, ISBN 0-8218-2694-8

[24] W. Karwowski and R.V. Mendes (1994), Hierarchical structures and asymmetric stochastic processes on p-adics and adeles, *J.Math.Phys.* **35**, 4637. 3

[25] Richard S. Kayne (1981), Unambiguous Paths, **Ch.5** in *Levels of Syntactic Representation*, Edited by Robert May and Jan Koster, Foris Publications, Dordrecht. 4

[26] E.L. Keenan and Bernard Comrie (1977), Noun Phrase Accessibility and Universal Grammar, *Linguistic Inquiry* **8**, 63-99. 19

[27] W.J.M. Levelt (1970), Hierarchical clustering algorithms in the psychology of grammar, In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.) Advances in psycholinguistics. Amsterdam: North Holland. 4

[28] D.G. Lockward (1972), Introduction to Stratification Grammar, Harcourt Brace Jovanovich Inc.,New York. 4

[29] J. McCloskey (1988), Syntatic Theory, p.18 in Linguistics: The Cambridge Survey, Cambridge University Press, Cambridge. 4

[30] George A. Miller (1956), 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. 7

[31] F. Murtagh (2004), On ultrametricity, data coding, and computation, *Journal of Classification*, **21**, 167-184. 3

[32] A.T. Ogielchi and D.L. Stein (1985), Dynamics on Ultrametric Spaces, *Phys.Rev.Lett.* **55**, 1634-1637. 3

[33] N. Parga and M.A. Virasoro (1986), The Ultrametric Organization of Neural Net Memories, *J.de Physique* **47**, 1857. 3

[34] E.M. Perlman and M. Schwarz (1992), The Directed Polymer Problem, *Europhys.Lett.* **17**, 227. 3

[35] A. Prince and P. Smolensky (1997), Optimality: From Neural Networks to Universal Grammar, *Science* **275**, 1604. 3

[36] R. Rammal, G. Toulouse and M.A. Virasoro (1986), Ultrametricity for Physicists, *Rev.Mod.Phys.* **58**, 765-788. 3

[37] Jorma Rissanen (1983), A universal prior for integers and estimation by minimum description length, *Annals of Statistics* **11**, 416-431. 5

[38] Mark D. Roberts (2004), P-model Alternative to the T-model, *Formal, Computational and Logical Linguistics* **5** 1-18.

[39] Mark D. Roberts (2005), Name Strategy: Its Existence and Implications, *Int.J.Computational Cognition* **3** 1-14. 7

[40] J.J.M.M. Rutten (1996), Elements of Generalized Ultrametric Domain Theory, *Theor.Comp.Sci.* **170**, 349. 3

[41] M. Schweinberger & T.A.B. Snijders (2003), Settings in Social Networks: A Measurement Model, *Sociological Methodology* **23**, 307-41. 3

[42] R.N. Shepard and P. Arabie (1979), Additive Clustering: Representative of Similarities as Combinations of Discrete Overlapping Properties, *Psychological Review* **86**, 87-123, especially page 90. 3

[43] P.H. Sneath and R.R. Sokal (1973), The Principles and Practice of Numerical Classification, W.H.Freeman and Company, San Francisco. 3

[44] George Sorton (1947), Introduction to the History of Science, Vol.**III**, page 552, The Williams and Wilkins Company, Baltimore. 5

[45] M.O. Vlad (1994), Fractal time, ultrametric topology and fast relation, *Phys.Lett.* **A189**, 299-303. 3

[46] M.B. Weissman (1993), What is Spin Glass, *Rev.Mod.Phys.* **65**, 829. 3

[47] M.R. Young and W.S. DeSarbo (1995), A Parametric procedure for Ultrametric tree estimation from conditional rank order proximity data, *Psychometrika* **60**, 47. 3

[48] Wlodek Zadrozny (2000), Minimum Description Length and Compositionality, `cs.CL/0001002` H. Blunt and R. Muskens (Eds) *Computing Meaning* Vol. **1**, Kluwer 1999, pp113-128. 6