Word Order Universals and Information Density

Michael Cysouw & Jelena Prokić

Information ...

- We will be using notions from information theory **as a statistical method**
- Not to measure the information density of **linguistic utterances** itself
- But as a method to test **typological correlations**

doi:10.1038/nature09923

Evolved structure of language shows lineage-specific trends in word-order universals

Michael Dunn[1,2], Simon J. Greenhill[3,4], Stephen C. Levinson[1,2] & Russell D. Gray[3]

Languages vary widely but not without limit. The central goal of linguistics is to describe the diversity of human languages and explain the constraints on that diversity. Generative linguists following Chomsky have claimed that linguistic diversity must be constrained by innate parameters that are set as a child learns a language[1,2]. In contrast, other linguists following Greenberg have claimed that there are statistical tendencies for co-occurrence of traits reflecting universal systems biases[3–5], rather than absolute constraints or parametric variation. Here we use computational phylogenetic methods to address the nature of constraints on linguistic diversity in an evolutionary framework[6]. First, contrary to the generative account of parameter setting, we show that the evolution of only a few word-order features of languages are strongly correlated. Second, contrary to the Greenbergian generalizations, we show that most observed functional dependencies between traits are lineage-specific rather than universal tendencies. These findings support the view that—at least with respect to word order—cultural evolution is the primary factor that determines linguistic structure, with the current state of a linguistic system shaping and constraining future states.

after the noun, whereas dominant object–verb ordering predicts postpositions, relative clauses and genitives before the noun[4]. One general explanation for these observations is that languages tend to be consistent ('harmonic') in their order of the most important element or 'head' of a phrase relative to its 'complement' or 'modifier'[3], and so if the verb is first before its object, the adposition (here preposition) precedes the noun, while if the verb is last after its object, the adposition follows the noun (a 'postposition'). Other functionally motivated explanations emphasize consistent direction of branching within the syntactic structure of a sentence[9] or information structure and processing efficiency[5].

To demonstrate that these correlations reflect underlying cognitive or systems biases, the languages must be sampled in a way that controls for features linked only by direct inheritance from a common ancestor[10]. However, efforts to obtain a statistically independent sample of languages confront several practical problems. First, our knowledge of language relationships is incomplete: specialists disagree about high-level groupings of languages and many languages are only tentatively assigned to language families. Second, a few large language families contain the bulk of global linguistic variation, making sampling purely from unrelated languages impractical. Some balance of related, unre-

THE WORLD ATLAS
OF LANGUAGE STRUCTURES
ONLINE

Welcome to WALS Online

The World Atlas of Language Structures (*WALS*) is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of 55 authors (many of them the leading authorities on the subject).

The first version of *WALS* was published as a book with CD-ROM in 2005 by Oxford University Press. The first online version was published in April 2008. Both are superseeded by the current online version, published in April 2011.

WALS Online is a joint effort of the Max Planck Institute for Evolutionary Anthropology and the Max Planck Digital Library. It is a separate publication, edited by Dryer, Matthew S. & Haspelmath, Martin (Munich: Max Planck Digital Library, 2011) ISBN: 978-3-9813099-1-1. The main programmer is Robert Forkel.

How to use WALS Online

Using *WALS Online* requires a browser (supported by Google Maps) with Javascript enabled.

You find the features or chapters of WALS through the items "Features" and "Chapters" in the navigation bar.

You can also browse and search for languages and language families alphabetically, by map region or by country through the item "Languages" on the navigation bar.

Search ×

WALS News

Previous edition of online WALS
May 10, 2011
The previous edition of the online WALS, the 2008 edition, is available at http://2008.wals.info/.

WALS 2011
Apr 28, 2011
Over the next couple of days we will push WALS 2011 - the new edition of WALS Online - live! While this should mean ...

Commentary Function Fixed
Apr 26, 2011
As we have only learnt recently, the commentary function for WALS online was broken. Comments did not get published, but were ...

Scheduled Server Downtime
Jan 11, 2011
Due to maintenance work, WALS Online will be down on 12 January 2011, from 8am CET (expected duration: < 8h). We apologize ...

Latest Comments

WALS	Feature
82	Order of Subject and Verb
83	Order of Object and Verb
84	Order of Object, Oblique, and Verb
85	Order of Adposition and Noun Phrase
86	Order of Genitive and Noun
87	Order of Adjective and Noun
88	Order of Demonstrative and Noun
89	Order of Numeral and Noun
90	Order of Relative Clause and Noun
91	Order of Degree Word and Adjective
92	Position of Polar Question Particles
93	Position of Interrogative Phrases in Content Questions
94	Order of Adverbial Subordinator and Clause

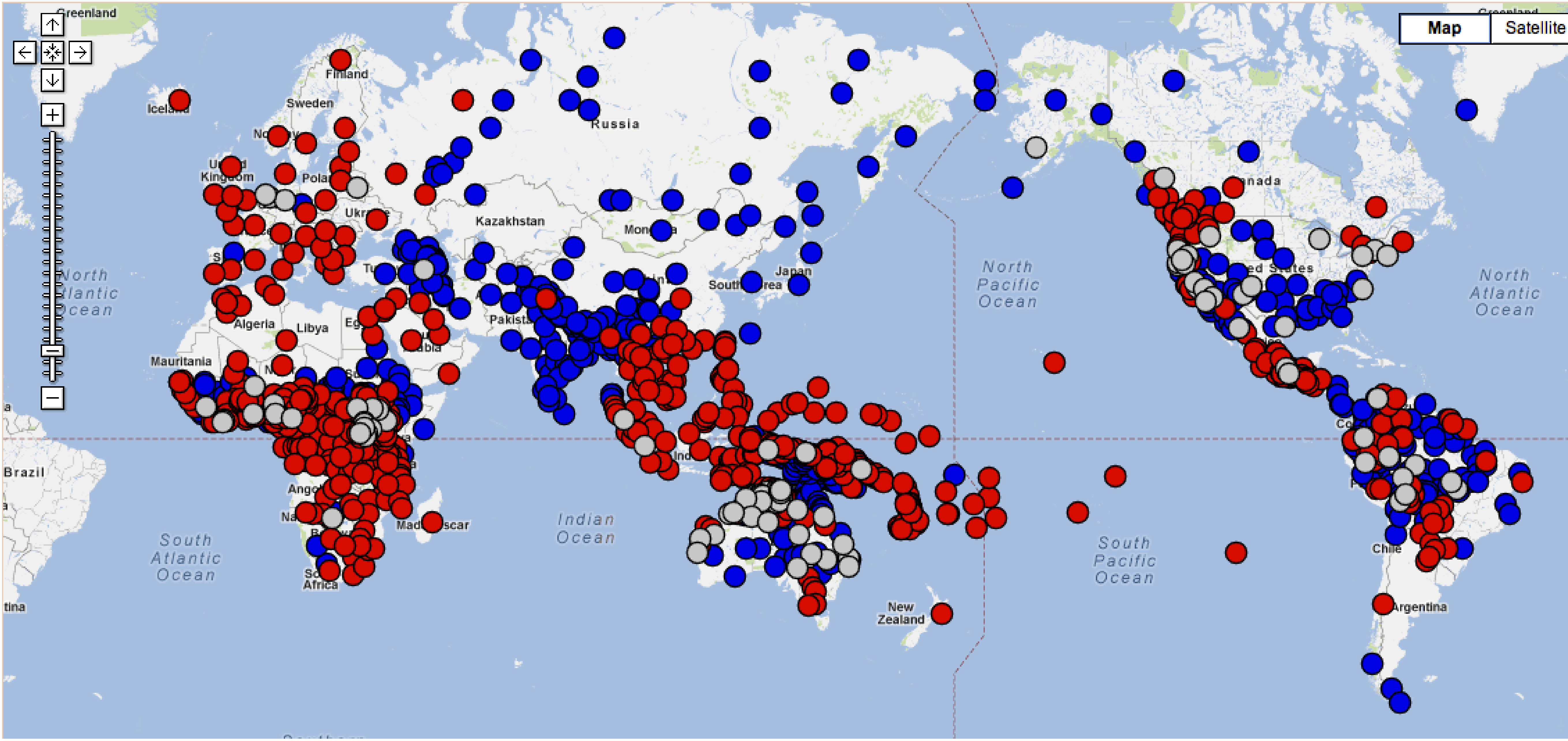
All data from Matthew Dryer

Feature 83A: Order of Object and Verb

by Matthew S. Dryer

get URL for the map currently displayed

show legend



Autocorrelation

(Galton's Problem)

"The difficulty raised by Mr. Galton that some of the concurrences might result from transmission from a common source, so that a single character might be counted several times from its mere duplication, is a difficulty ever present in such investigations [...]. The only way of meeting this objection is to make separate classifications depend on well marked differences, and to do this all over the world" (Taylor 1889: 272).

Dunn et al. (2011)

- Different solution to Galton’s problem
  - based on work by Mark Pagel (see also Elena Maslova)
- Use detailed structure of genealogical tree to investigate changes in types
- Correlated characteristics should co-evolve, i.e change together

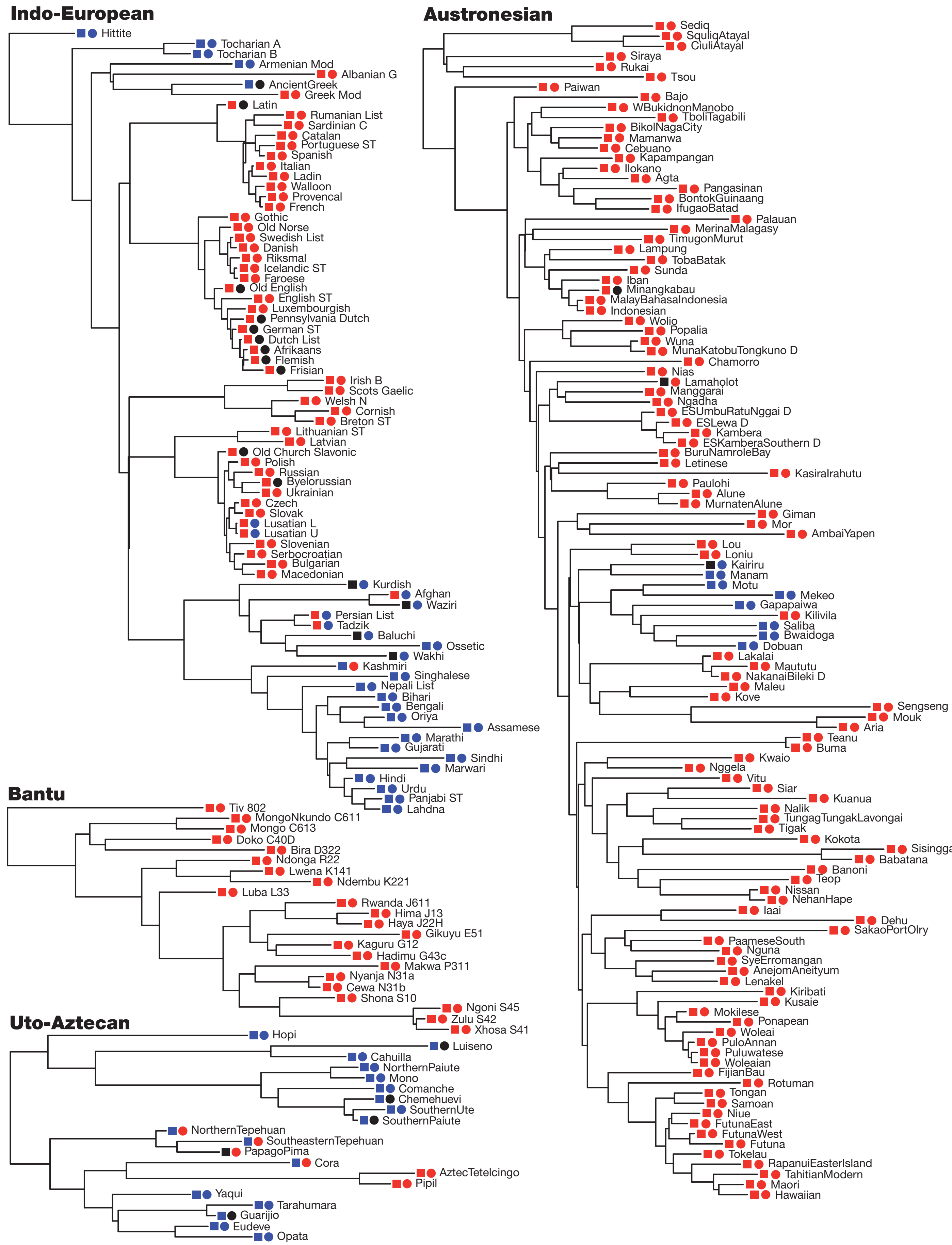


Figure 1 | Two word-order features plotted onto maximum clade credibility trees of the four language families. Squares represent order of adposition and noun; circles represent order of verb and object. The tree sample underlying this tree is generated from lexical data[16,22]. Blue-blue indicates postposition, object–verb. Red-red indicates preposition, verb–object. Red-blue indicates preposition, object–verb. Blue-red indicates postposition, verb–object. Black indicates polymorphic states.

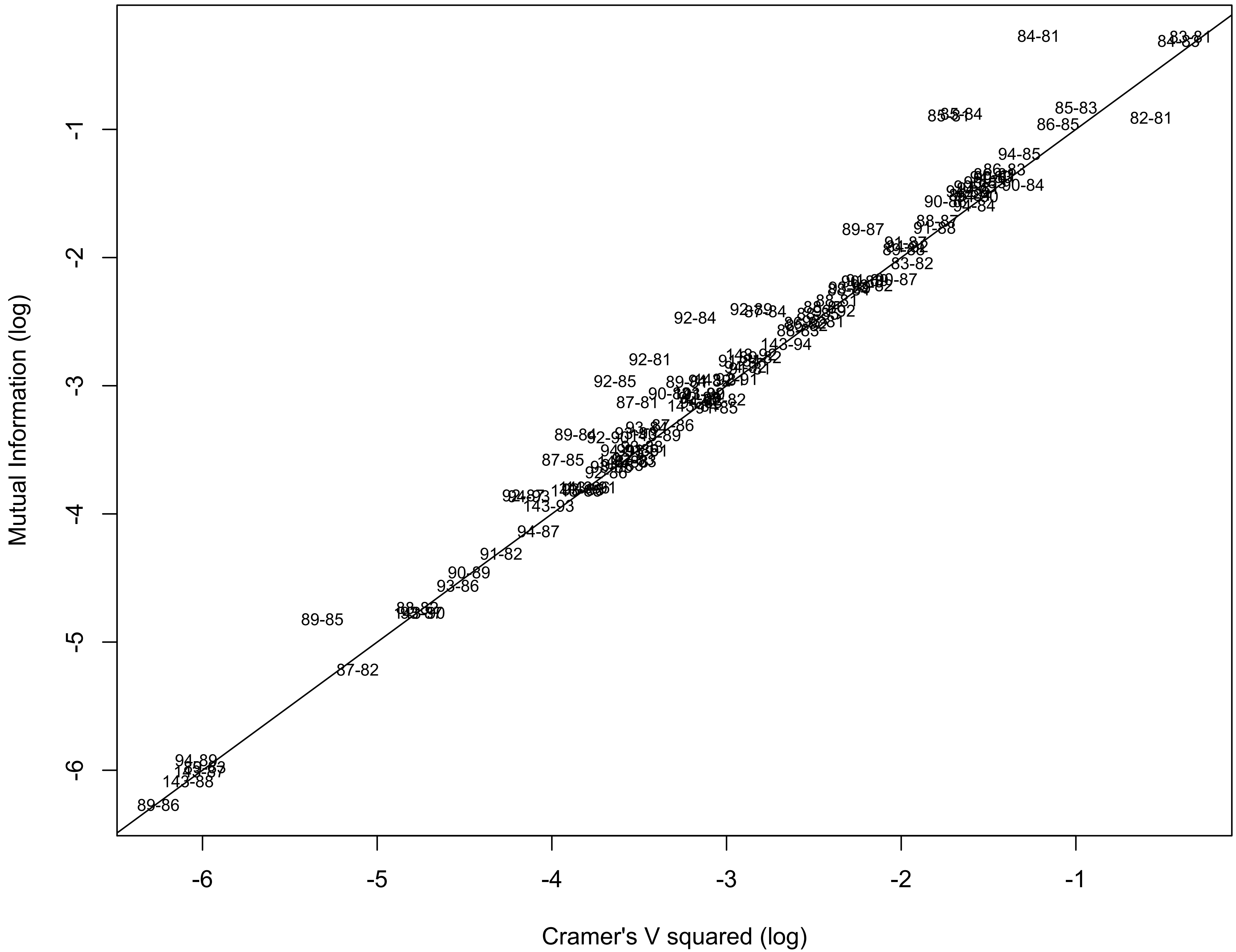
Right in principle, but:

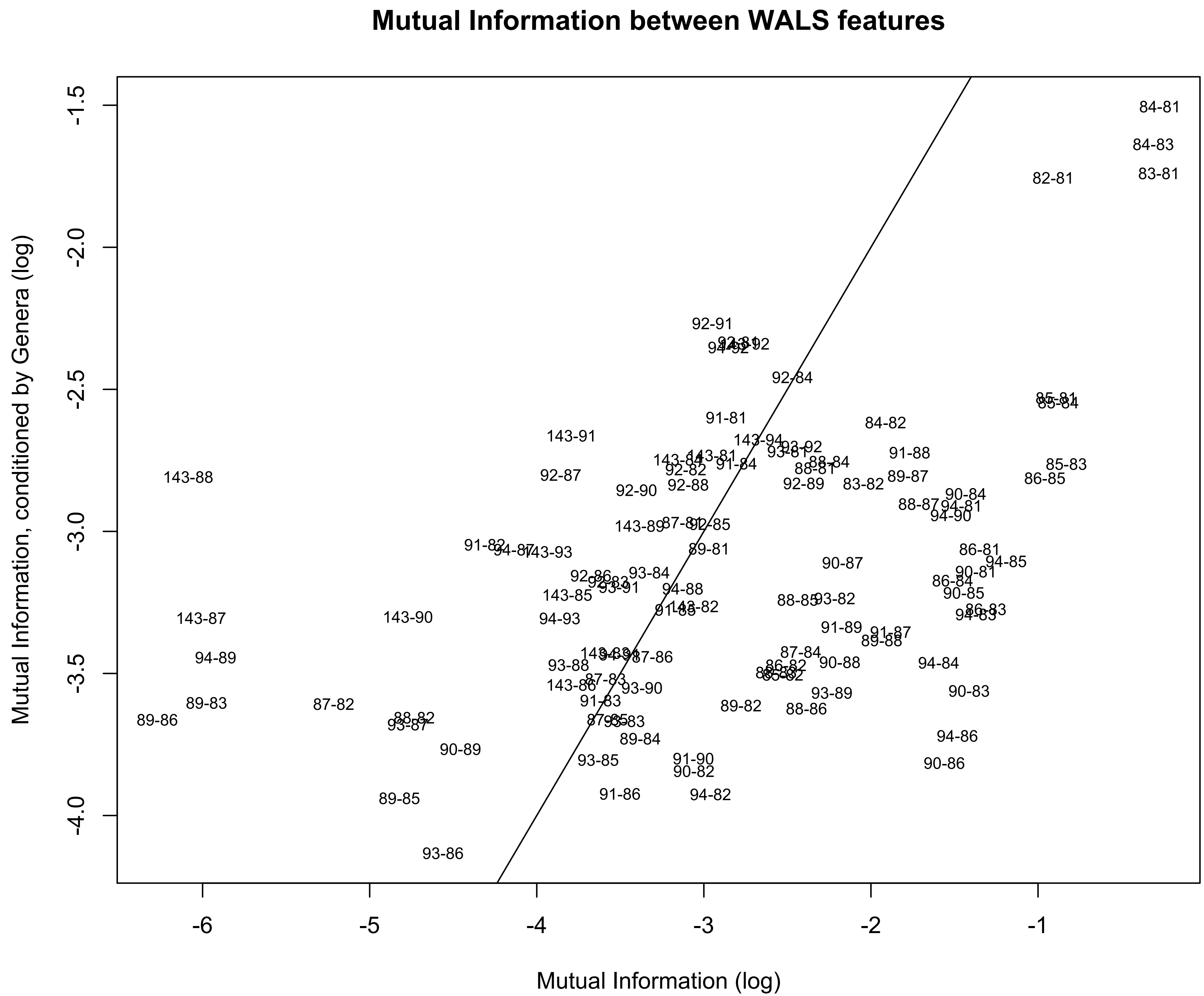
(see reactions in a special issue of Linguistic Typology)

- Their interpretation of results is too radical
  - "most observed functional dependencies between traits are lineage-specific rather than universal tendencies" (p.79)
- It is difficult to obtain the necessary data for many families
  - 4 families is not enough to find weaker typological patterns
- Computationally their method is very demanding

Any alternative?

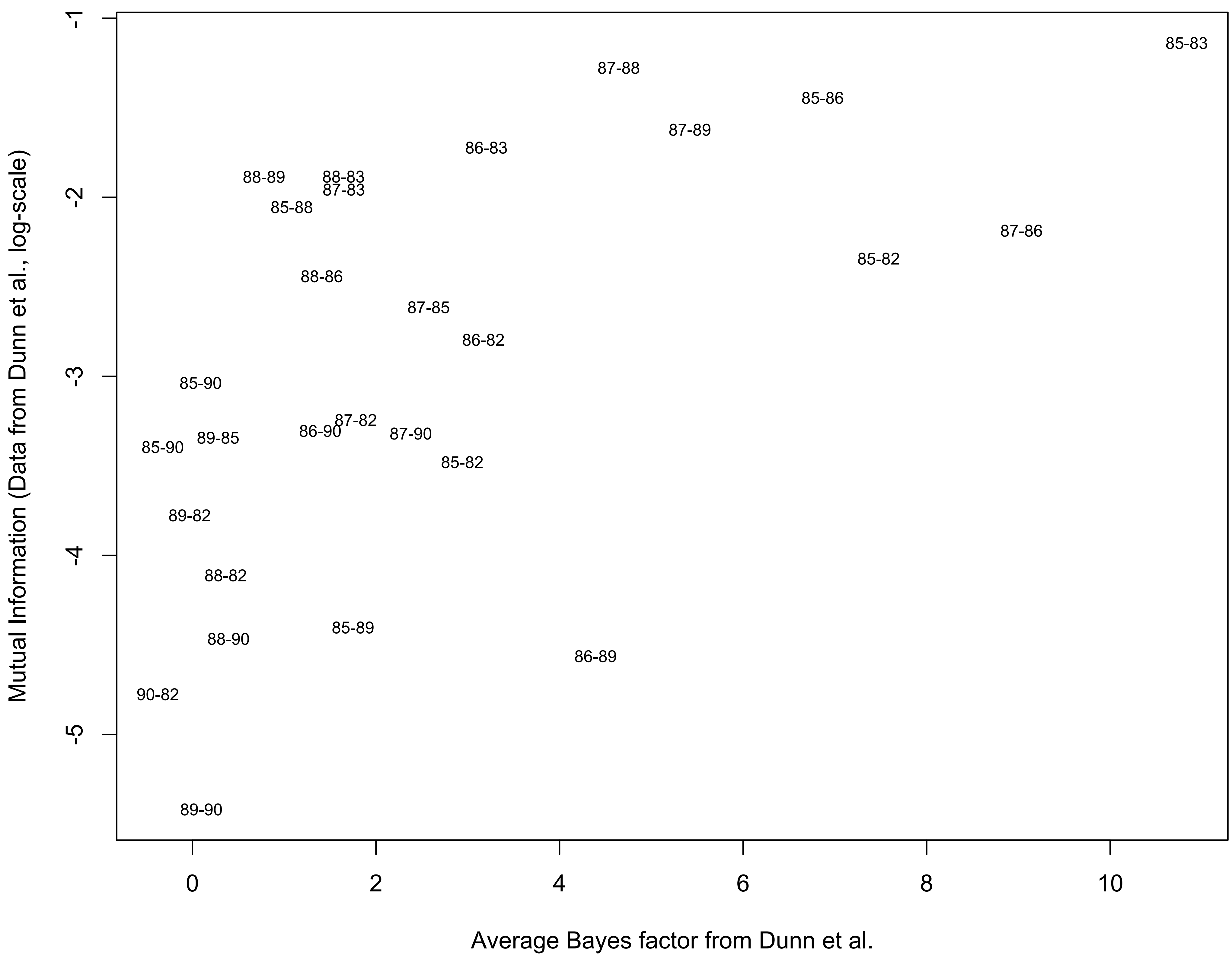
- Conditional Mutual Information
  - **Information** (or entropy) of a typological feature measures ‘fractionality’
  - **Mutual Information** is measure of shared distribution
  - **Conditional Mutual Information** accounts for conditioning factors


Mutual Information (log)
Cramer's V squared (log)
-1
-2
-3
-4
-5
-6
84-81
82-81
85-83
86-85
94-85
90-84
89-87
83-82
90-87
92-84
143-94
92-81
92-85
87-81
89-84
87-85
143-93
94-87
91-82
90-89
93-86
89-85
87-82
143-88
89-86

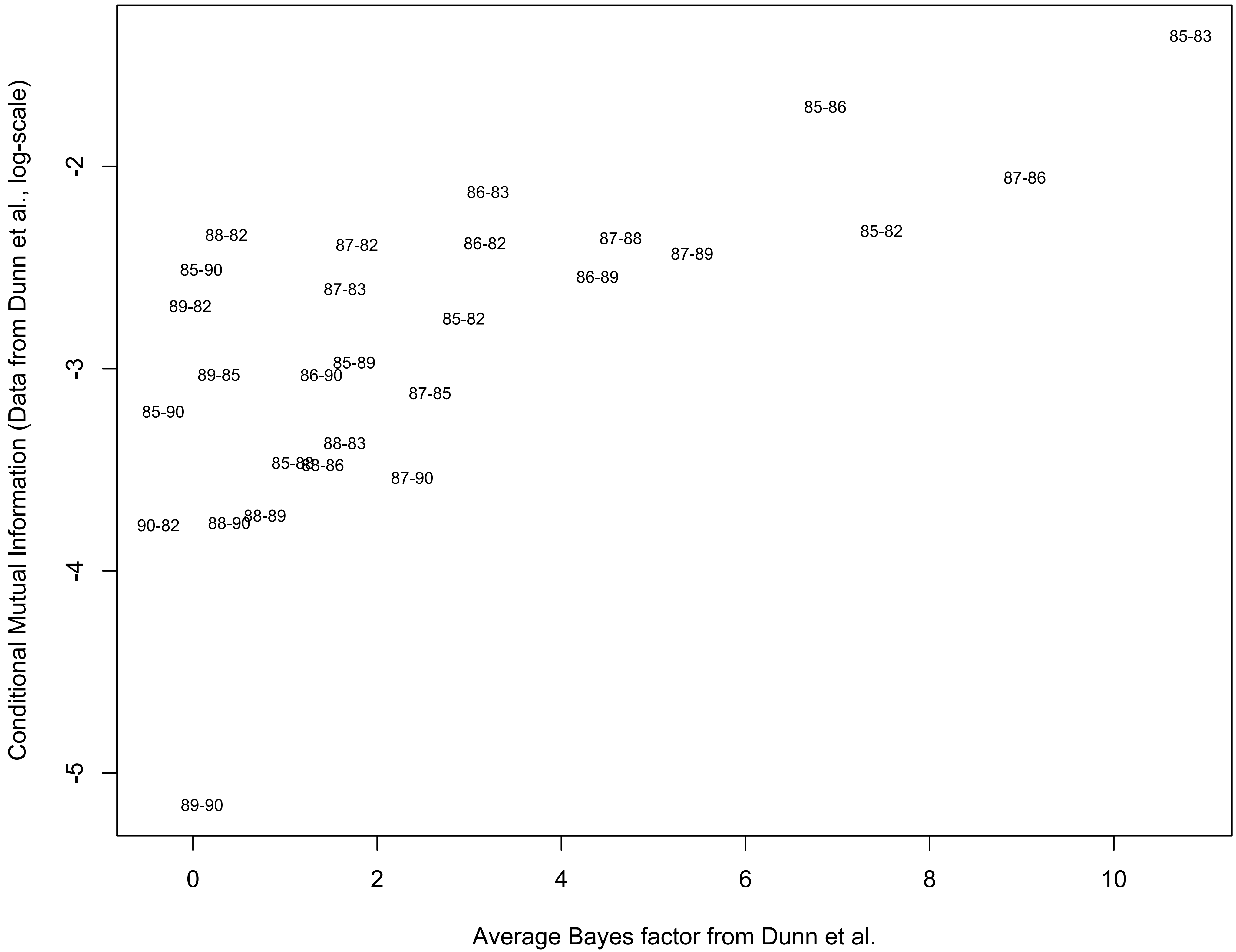

Mutual Information between WALS features
Mutual Information, conditioned by Genera (log)
Mutual Information (log)
-1.5
-2.0
-2.5
-3.0
-3.5
-4.0
-6
-5
-4
-3
-2
-1
84-81
84-83
83-81
82-81
92-91
92-84
85-81
85-84
91-81
84-82
143-91
143-94
93-92
93-81
91-88
143-81
88-84
85-83
143-88
92-87
92-82
91-84
88-81
89-87
86-85
92-90
92-88
92-89
83-82
90-84
88-87
94-81
94-90
143-89
92-85
91-82
143-93
89-81
86-81
94-85
90-87
92-86
93-84
90-81
86-84
93-91
94-88
90-85
143-85
88-85
93-82
143-82
86-83
94-83
143-87
143-90
94-93
91-89
91-87
89-88
94-89
87-84
87-86
93-88
90-88
94-84
87-83
143-86
93-90
90-83
93-89
89-83
87-82
91-83
89-82
88-86
89-86
88-82
93-87
93-83
94-86
89-84
90-89
91-90
90-86
93-85
90-82
91-86
94-82
89-85
93-86

Comparison to Dunn et al.

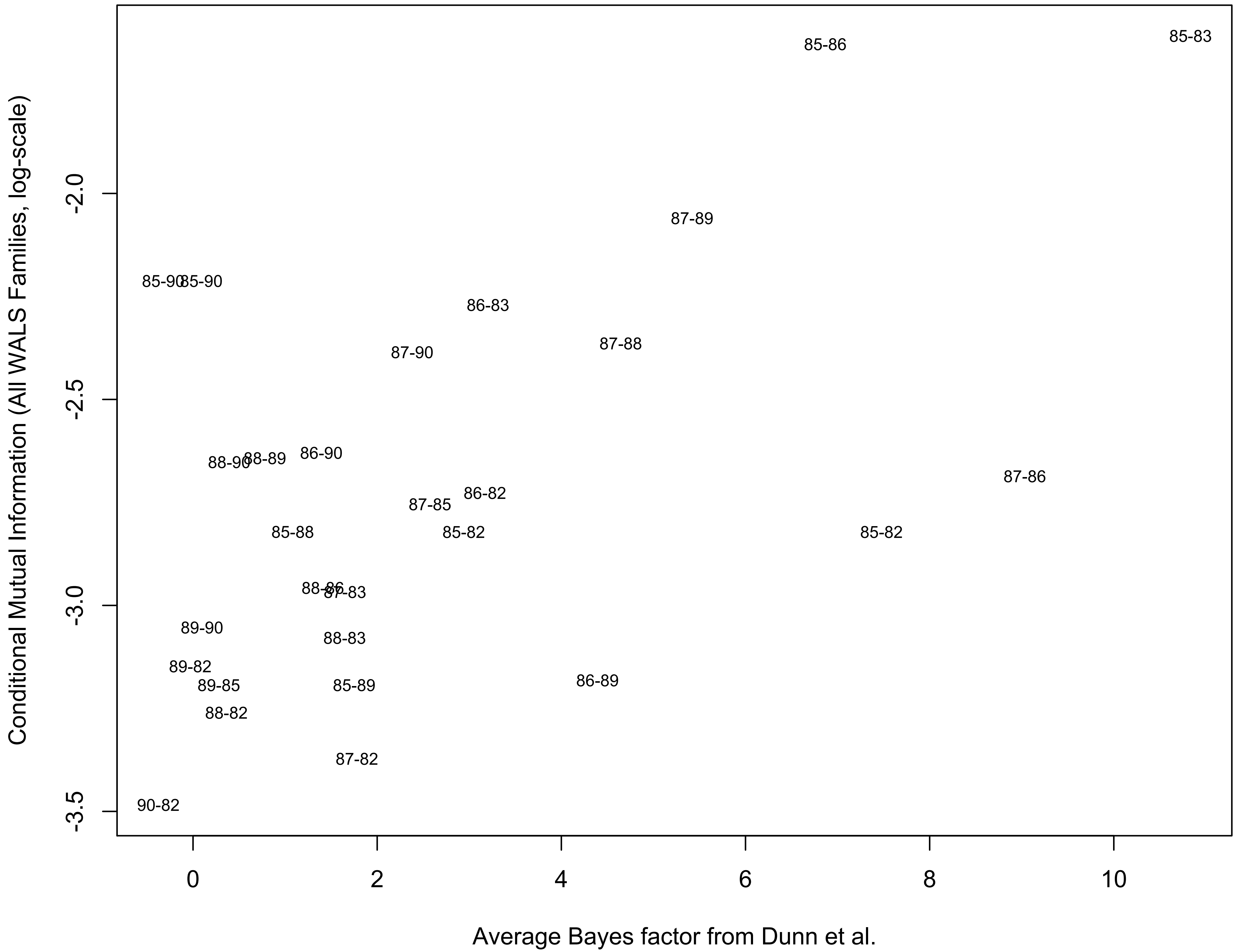
- First, using only the data from Dunn et al.
  - compare with Mutual Information: approximate match
  - compare with Conditional Mutual Information (CMI) conditioned by families: good match
- Second, using all WALS data, CMI by family



Mutual Information (Data from Dunn et al., log-scale)
Average Bayes factor from Dunn et al.
-1
-2
-3
-4
-5
0
2
4
6
8
10
85-83
87-88
85-86
87-89
86-83
88-89
88-83
87-83
85-88
87-86
85-82
88-86
87-85
86-82
85-90
87-82
86-90
87-90
89-85
85-90
85-82
89-82
88-82
85-89
88-90
86-89
90-82
89-90



85-83
85-86
87-86
86-83
88-82
85-82
87-82
86-82
87-88
87-89
85-90
86-89
87-83
89-82
85-82
85-89
89-85
86-90
87-85
85-90
88-83
85-88
88-86
87-90
88-89
90-82
88-90
89-90
-2
-3
-4
-5
0
2
4
6
8
10
Conditional Mutual Information (Data from Dunn et al., log-scale)
Average Bayes factor from Dunn et al.



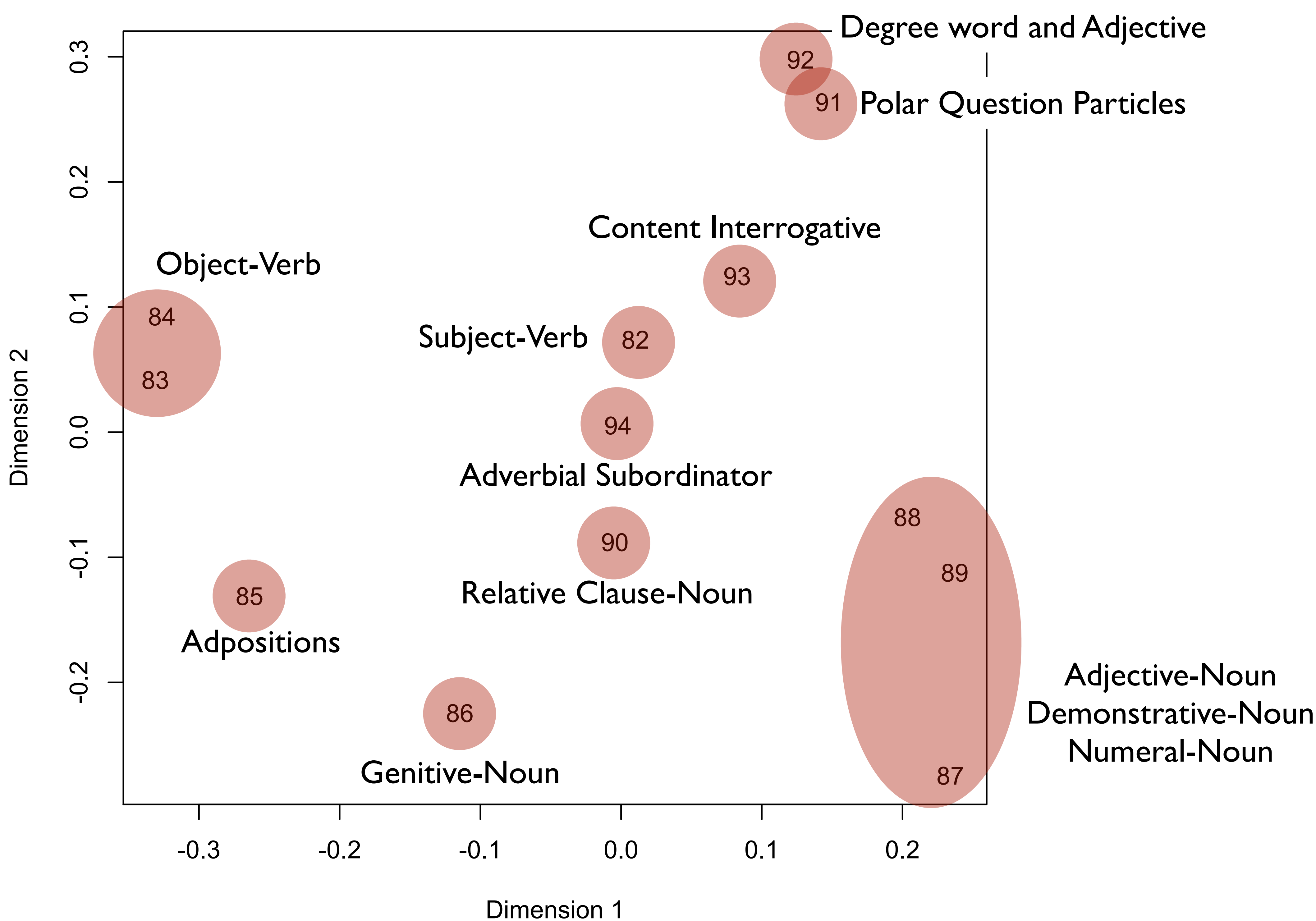
Conditional Mutual Information (All WALS Families, log-scale)
-2.0
-2.5
-3.0
-3.5
Average Bayes factor from Dunn et al.
0
2
4
6
8
10
85-86
85-83
87-89
85-90
85-90
86-83
87-88
87-90
88-90
88-89
86-90
87-86
86-82
87-85
85-88
85-82
85-82
88-86
87-83
89-90
88-83
89-82
89-85
85-89
86-89
88-82
87-82
90-82

Conclusion

- Mutual Information conditioned by linguistic families (CMI) is highly similar to the Dunn et al. measure
- CMI is easier to apply for many families
- Dunn et al. data shows influence from limited selection of families

MDS of Conditional MI

(using Family as condition)



Next steps

- Conditional Mutual Information uses a classification as condition
  (e.g. genera, families, areas, ...)
- Many classifications can be combined as multiple conditioning factors
- But: hierarchically ordered classifications are identical to the most detailed classification
  (e.g. in WALS: genera $\subset$ families $\subset$ areas $\equiv$ genera)
- New work by Dress & Albu: Conditional Mutual Information, conditioned by a tree