Syntactic measuring of linguistic distances

Giuseppe Longobardi
longbard@units.it

(Laboratorio di Linguistica & Antropologia cognitiva, Trieste)

Vienna,October 19th 2010

Darwin's challenge

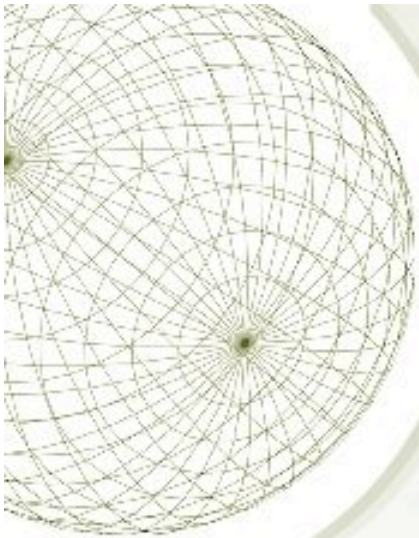
[The Origin of Species, ch. 14]

"If we possessed a perfect pedigree of mankind, a *genealogical arrangement of the races of man* would afford the *best classification of the various languages* now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one"

Cavalli Sforza, Menozzi, and Piazza (1994)

The History and Geography of Human Genes, p.18

"We believe that the major breakthrough in the study of human variation has been the introduction of genetic markers, which are strictly inherited and basically immune to the problem of rapid changes induced by the environment"



Phylogenetic reconstruction and comparative linguistics

Ideal properties of taxonomic characters for linguistic phylogenies:

Universality, discreteness (useful for clustering in relative taxonomies),

improbable similarity (to prove absolute relatedness against chance similarity):

Classical Comparative Method

Based on **lexical** arbitrariness

comparanda **= patently beyond chance probability**

(e.g. 'regular/recurrent' sound correspondences)

Solid phylogenetic conclusions and surprising etymologies (e.g. *full/pieno*)

BUT

so improbable similarities = inevitably rare across languages

OK for Italian/English,

not

Italian/Japanese (long-range comparison)

Mass (multilateral) comparison

[Greenberg's life-long attempt toward a global classification]

Also based on **lexical** arbitrariness

Relies on universal meaning lists

(e.g. Swadesh lists)

Hence, in principle applicable to any set of languages

but…

Mass (multilateral) comparison

supposed cognates are ‘identified’

on mere word resemblance,

not on any precise and improbable phenomena

not safe from chance similarity

Common problem: ***non discreteness*** *of vocabulary comparisons*

- ⇨ **partial identity of** form**:** ***prendo*** **vs.** ***get***
- ⇨ **partial identity of** meaning**:** ***Hund*** **vs.** ***hound/dog***
- ⇨ **identity of** form**, not of** meaning**:** ***klein*** **vs.** ***clean***
- ⇨ **similarity of** meaning shifts, **no correspondence of form:** ***fegato*** **vs. συκώτι**
- ⇨ **difficulty of measuring relative distances:** ***(je) fonds, (ich) giesse, juhomi***

Population Genetics

- ***Comparanda*** **are drawn from a finite** ***universal list of discrete biological options*** **(alleles of genetic polymorphisms)**
- **Comparison is safe w.r.t. chance and universal possible (measuring of taxonomic distances possible between any two populations)**

Parametric Comparison Method (PCM)

Longobardi (2003), Longobardi& Guardiano (2009):

Lexically blind method of **syntactic** comparison

⇨ **universally** applicable

a. mathematically **discrete** results

c. choice of comparanda: **safe from chance**

d. parameter settings: virtually **immune** from natural selection and conscious cultural choice

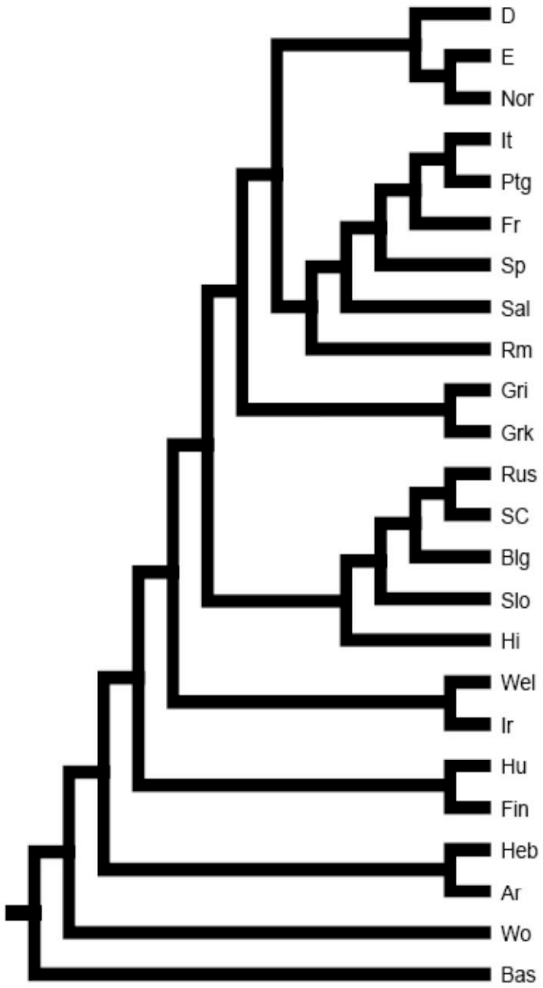
	TABLE A		It	Sal	Sp	Fr	Ptg	Rm	Lat	ClG	NTG	Grl	Grk	Got	OE	E	D	Nor	Blg	SC	Rus	Ir	Wel	Heb	Ar	Wo	Hu	Fin	Hi	Bas	Slo	
1	± gramm. person	FGP	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1
2	± gramm. number +FGP	FGN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	2
3	± gramm. gender +FGN	FGG	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	-	+	3
4	± variable person on D +FGP	FVP	-	-	+	-	-	-	?	+	?	-	+	?	?	-	-	-	-	-	-	?	?	-	+	-	-	-	-	+	+	4
5	± feature spread to N +FGN	FSN	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	+	5
6	± numb. on N (BNs) +FSN	FNN	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	+	+	+	0	+	6
7	± gramm. partial def	DGP	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	-	-	-	7
8	± gramm. def +DGP	DGR	+	+	+	+	+	+	0	+	+	+	+	-	+	+	+	+	+	0	0	+	+	+	+	+	+	0	0	0	0	8
9	± free null partitive Q +FNN	DPQ	-	-	-	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	+	-	0	-	9
10	± gramm. dist. art. -FSN or -FNN or +DGP	DDA	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	+	-	0	0	+	0	10
11	± gramm. top. dem. -FSN or -FNN or +DGP, -DDA	DTA	-	+	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	-	0	-	0	0	0	0	11
12	± def checking N +DGP	DCN	-	-	-	-	-	+	0	-	-	-	-	-	-	-	-	-	+	0	0	-	-	-	-	-	-	0	0	0	0	12
13	± def spread to N +DCN, -NSD	DSN	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	13
14	± def on attrib. +DGP, -DCN	DOA	-	-	-	-	-	0	0	+	+	-	+	+	-	-	-	0	0	0	0	-	-	+	+	-	-	0	0	0	0	14
15	± def on relatives +DGP	DOR	-	-	-	-	-	-	0	-	-	-	-	-	-	-	-	-	-	0	0	-	-	-	+	+	-	0	0	0	0	15
16	± D-controlled infl. on N +FSN	DIN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	0	-	-	-	0	-	16
17	± plural spread from cardinals +FSN	CPS	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	0	-	-	+	0	+	17
18	± gramm. mass-to-count	CFM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	?	+	-	+	-	-	-	-	-	-	18
19	± gramm. boundedness	CGB	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	-	-	19
20	± gramm. partial count -FSN or -FNN or +DGP, -CGB	CGP	+	+	+	+	+	+	0	-	-	+	+	0	+	+	+	+	+	0	0	-	-	-	-	0	0	0	0	+	0	20
21	± gramm. count +CGP	CGR	+	+	+	+	+	+	0	0	0	+	+	0	-	+	+	+	-	0	0	0	0	0	0	0	0	0	0	-	0	21
22	± count checking N +CGB or +CGP	CCN	-	-	-	-	-	-	0	0	0	-	-	0	-	-	-	-	-	0	0	0	0	0	0	-	-	0	-	-	0	22
23	± free Gen	GPR	+	+	+	+	+	-	+	+	-	-	-	-	-	+	+	+	+	0	0	+	+	+	+	-	-	+	+	+	+	23
24	± DP over free Gen +GPR	GIN	+	+	+	+	+	0	+	+	0	0	0	0	0	+	+	+	+	0	0	+	+	+	+	0	0	-	-	-	+	24
25	± GenO -free infl., -GPR or +GIN	GFO	-	-	-	-	-	-	0	0	+	+	+	+	+	-	+	-	-	+	+	+	+	-	-	+	-	0	0	0	+	25
26	± GenS -GPR or *-free infl.*, (+GPR, +GIN) or +GFO, -NPP	GFS	-	-	-	-	-	-	0	0	-	-	-	-	+	+	+	+	-	-	-	-	-	+	+	-	0	0	0	0	-	26
27	± def affix Gen +DCN	GDA	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	0	0	27
28	± feature spread to postp. Gen +*postp. gen.*, +AFS	GSP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	0	+	0	0	28
29	± Poss°-checking N	GCN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	29
30	± structured APs	AST	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	30
31	± feature spread to structured APs +AST	AFS	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	+	+	+	-	+	+	0	-	+	+	-	+	31
32	± feature spread to prod. APs	AFP	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	-	+	+	+	+	+	+	+	+	32
33	± numb. on A +FNN, +AFS or +AFP	AFN	+	+	+	0	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+	+	0	+	+	0	+	+	+	0	+	33
34	± D-controlled infl. on A +AFS, -NSD	ADI	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	+	+	0	0	0	-	0	0	0	0	0	0	0	0	34
35	± relative clauses	ARC	+	+	+	+	+	+	?	?	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	35
36	± DP over relatives +ARC	ADR	+	+	+	+	+	+	?	?	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	0	-	+	36
37	± free APs in Mod +AST, +ARC	AFM	+	-	+	+	+	-	+	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	+	0	-	-	0	-	-	37
38	± APs in Mod -AFM	AMO	0	-	0	0	0	-	0	0	0	-	0	0	+	+	-	-	+	+	+	-	-	+	0	0	-	-	0	-	-	38
39	± overt Mod° -AST or +AFM or +AMO	AOM	-	0	-	-	-	0	-	-	-	0	-	-	-	-	0	0	-	-	-	0	0	-	-	+	0	0	0	0	0	39
40	± adjectival Gen	AGE	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	+	40
41	± N-raising with pied-piping	NPP	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	+	-	41
42	± N over ext.arg. -NPP	NOE	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+	0	0	+	+	+	+	0	+	42
43	± N over GenO +*free infl. Gen* or +GFO, +NOE	NOG	0	0	0	0	0	0	-	-	+	+	+	+	+	0	+	0	0	+	+	+	+	0	0	+	0	-	0	0	+	43
44	± N over As +AST, (-*free infl. Gen* or -GFO, +NOE) or +NOG	NOA	+	+	+	+	+	+	0	0	-	+	-	-	-	0	-	0	-	-	-	+	+	0	0	0	-	0	-	0	-	44
45	± N over M2 As +NOA	NM2	+	+	+	+	+	+	0	0	0	+	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	0	0	0	0	45
46	± N over M1 As +NM2	NM1	-	+	-	-	-	-	0	0	0	+	0	0	0	0	0	0	0	0	0	+	+	0	0	0	0	0	0	0	0	46
47	± N over high As +NM1	NOH	0	+	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	0	0	-	-	0	0	0	0	0	0	0	0	47
48	± N over cardinals -CGN or +CGA, +NPP or +NOH	NOC	0	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	-	0	0	0	-	0	48
49	± NP over D	NOD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	49
50	± strong D (person) +FGP, +DGR, -NOD	NSD	+	+	+	+	+	+	0	+	+	+	+	0	-	-	-	-	+	0	0	-	-	+	+	0	+	0	0	0	0	50
51	± gr. geog .ext. -FSN or -FNN or +DGP, *-general. art. with pr. names*, +NOD or +NSD	NGE	+	+	-	+	+	0	0	?	?	?	0	0	0	0	0	0	-	0	0	0	0	-	-	-	-	0	0	-	0	51
52	± D-Dem	TDD	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+	+	52
53	± A-Dem +AST, +TDD	TAD	-	-	+	-	-	+	+	+	+	-	+	+	-	-	-	-	-	-	-	0	0	0	+	0	-	-	-	-	-	53
54	± D-lic. Dem +*visibleArt*, +TDD or (-TDD, -CGP)	TDL	+	+	+	+	+	+	0	-	-	+	-	0	+	+	+	+	+	0	0	-	-	+	-	+	-	0	0	+	0	54
55	± Gen feature spread to Poss° +GCN, -NOG or -NOA	PGS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	+	0	0	0	55
56	± head feature spread to possessives +*overt poss.*, +AFS or + AFP, -AGE	PHS	+	+	+	+	+	+	+	+	+	-	-	+	+	+	0	+	+	0	0	0	-	0	-	-	-	0	0	0	0	56
57	± D-poss -FSN or -FNN or +DGR, (+CGR,+AFP) or -GFS	PDE	-	-	+	+	-	-	0	-	-	-	-	0	0	0	0	0	-	0	0	+	+	-	-	-	0	0	0	-	0	57
58	± A-poss +AST, -AGE, +PHS or (-AFS,-AFP)	PAD	+	+	+	+	+	-	+	+	+	0	0	+	+	-	+	-	0	0	0	0	-	0	0	0	0	0	-	-	0	58
59	± free high A-poss +PAD	PAH	+	0	-	+	-	0	?	?	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	59
60	± def affix possessives -GDA	PDA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	+	0	0	0	0	0	0	0	0	0	0	0	0	0	60
61	± clitic possessives	PCL	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	+	+	-	-	-	-	-	+	0	0	-	-	-	61
62	± Consisten. Princ. (-NOG or ..., or -NM1, +*A-Compl*) or +NOD	XCP	+	0	+	+	+	+	?	?	?	0	-	?	+	+	+	+	-	-	-	0	0	0	0	0	+	+	0	+	-	62
63	± null-N-licensing art -FSN or-FNN or -DCN, +NSD or +NOD	XNN	-	-	+	-	+	0	0	+	+	-	-	?	0	0	0	0	0	0	0	0	0	-	-	?	-	0	0	+	0	63
			It	Sal	Sp	Fr	Ptg	Rm	Lat	ClG	NTG	Grl	Grk	Got	OE	E	D	Nor	Blg	SC	Rus	Ir	Wel	Heb	Ar	Wo	Hu	Fin	Hi	Bas	Slo	

Elaboration of data: ***Jaccard distances***

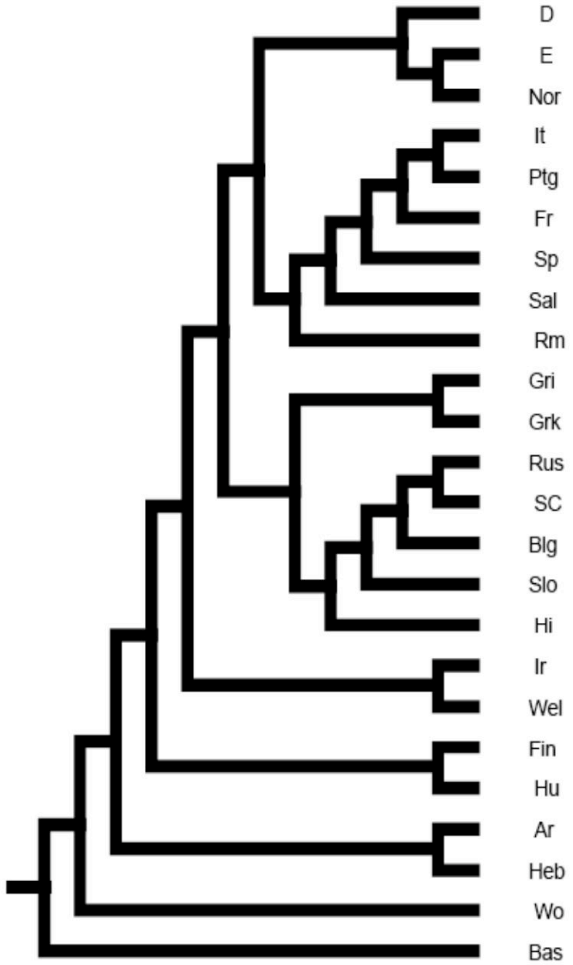
COEFFICIENTS

	It	Sal	Sp	Fr	Ptg	Rm	Gri	Grk	E	D	Nor	Blg	SC	Rus	Ir	Wel	Heb	Ar	Wo	Hu	Fin	Hi	Bas	Slo	
It	0	<46, 3>	<47, 5>	<48, 2>	<50, 2>	<42, 5>	<41, 6>	<37,10>	<37, 7>	<40, 6>	<38, 6>	<38, 7>	<25, 8>	<26, 7>	<33, 9>	<29,13>	<34, 8>	<31,12>	<20,12>	<31, 9>	<22, 7>	<24,6>	<18,14>	<26, 7>	It
Sal	0,0612	0	<42, 7>	<42, 5>	<45, 4>	<41, 6>	<43, 6>	<34,11>	<35, 8>	<40, 6>	<38, 6>	<37, 7>	<25, 7>	<26, 6>	<35, 9>	<31,13>	<33,10>	<28,15>	<21,11>	<31, 9>	<23, 6>	<24,6>	<20,13>	<28, 5>	Sal
Sp	0,0962	0,143	0	<45, 5>	<49, 3>	<41, 6>	<37,10>	<37,10>	<35, 9>	<38, 8>	<36, 8>	<35,10>	<23,10>	<24, 9>	<34, 8>	<30,12>	<31,11>	<31,12>	<18,14>	<28,12>	<20, 9>	<22,8>	<18,14>	<26, 7>	Sp
Fr	0,04	0,106	0,1	0	<46, 4>	<38, 7>	<37, 8>	<33,12>	<35, 8>	<37, 7>	<35, 7>	<34, 9>	<22, 9>	<23, 8>	<31, 9>	<28,13>	<30,10>	<27,14>	<19,13>	<28,10>	<20, 7>	<21,7>	<17,15>	<23, 8>	Fr
Ptg	0,0385	0,0816	0,0577	0,08	0	<42, 5>	<40, 7>	<36,11>	<37, 7>	<40, 6>	<38, 6>	<38, 7>	<25, 8>	<26, 7>	<33, 9>	<29,13>	<33, 9>	<30,13>	<20,12>	<30,10>	<22, 7>	<24,6>	<19,13>	<26, 7>	Ptg
Rm	0,106	0,128	0,128	0,156	0,106	0	<40, 6>	<35, 9>	<33, 9>	<36, 9>	<39, 6>	<36, 9>	<25, 8>	<26, 7>	<31,10>	<28,13>	<30, 9>	<26,13>	<19,11>	<30, 9>	<22, 7>	<23,6>	<19,12>	<25, 8>	Rm
Gri	0,128	0,122	0,213	0,178	0,149	0,13	0	<40, 6>	<32, 9>	<38, 7>	<34, 8>	<36, 7>	<29, 4>	<28, 5>	<37, 7>	<32,11>	<31,10>	<26,15>	<25, 7>	<32, 8>	<22, 7>	<22,6>	<17,13>	<27, 6>	Gri
Grk	0,213	0,244	0,213	0,267	0,234	0,205	0,13	0	<28,14>	<33,12>	<29,13>	<35, 9>	<29, 5>	<28, 6>	<31, 9>	<26,13>	<30,11>	<32,10>	<22,11>	<29,11>	<18,11>	<21,7>	<14,16>	<27, 6>	Grk
E	0,159	0,186	0,205	0,186	0,159	0,214	0,22	0,333	0	<39, 5>	<39, 4>	<33,10>	<22,10>	<23, 9>	<28, 9>	<30, 9>	<31, 8>	<25,14>	<16,14>	<28,10>	<19,10>	<21,7>	<19,12>	<22,10>	E
D	0,13	0,13	0,174	0,159	0,13	0,2	0,156	0,267	0,114	0	<41, 5>	<34,10>	<26, 8>	<27, 7>	<34, 8>	<32, 9>	<30,10>	<25,15>	<19,12>	<30,10>	<23, 8>	<24,6>	<18,13>	<29, 6>	D
Nor	0,136	0,136	0,182	0,167	0,136	0,133	0,19	0,31	0,093	0,109	0	<37, 8>	<23, 9>	<24, 8>	<30, 9>	<27,11>	<31, 8>	<26,13>	<15,14>	<28,10>	<23, 7>	<24,6>	<20,11>	<26, 7>	Nor
Blg	0,156	0,159	0,222	0,209	0,156	0,2	0,163	0,205	0,233	0,227	0,178	0	<31, 3>	<30, 4>	<28,11>	<24,15>	<32, 8>	<26,14>	<18,12>	<28,11>	<21, 9>	<23,6>	<17,14>	<29, 5>	Blg
SC	0,242	0,219	0,303	0,29	0,242	0,242	0,121	0,147	0,312	0,235	0,281	0,0882	0	<34, 1>	<23, 7>	<19,11>	<21, 9>	<17,13>	<16, 8>	<21,10>	<21, 9>	<22,5>	<12,12>	<30, 4>	SC
Rus	0,212	0,188	0,273	0,258	0,212	0,212	0,152	0,176	0,281	0,206	0,25	0,118	0,0286	0	<24, 6>	<20,10>	<22, 8>	<18,12>	<15, 9>	<21,10>	<21, 9>	<23,4>	<13,11>	<31, 3>	Rus
Ir	0,214	0,205	0,19	0,225	0,214	0,244	0,159	0,225	0,243	0,19	0,231	0,282	0,233	0,2	0	<40, 3>	<30, 9>	<27,11>	<17,12>	<25,10>	<19, 8>	<19,6>	<12,13>	<27, 4>	Ir
Wel	0,31	0,295	0,286	0,317	0,31	0,317	0,256	0,333	0,231	0,22	0,289	0,385	0,367	0,333	0,0698	0	<25,13>	<23,14>	<15,14>	<25,10>	<17,10>	<16,10>	<13,14>	<23, 8>	Wel
Heb	0,19	0,233	0,262	0,25	0,214	0,231	0,244	0,268	0,205	0,25	0,205	0,2	0,3	0,267	0,231	0,342	0	<36, 7>	<16,15>	<26,10>	<19, 8>	<18,7>	<14,15>	<21, 9>	Heb
Ar	0,279	0,349	0,279	0,341	0,302	0,333	0,366	0,238	0,359	0,375	0,333	0,35	0,433	0,4	0,289	0,378	0,163	0	<15,16>	<23,13>	<16,11>	<16,10>	<13,16>	<20,10>	Ar
Wo	0,375	0,344	0,438	0,406	0,375	0,367	0,219	0,333	0,467	0,387	0,483	0,4	0,333	0,375	0,414	0,483	0,484	0,516	0	<19, 8>	<12, 8>	<12,8>	<14,10>	<13,10>	Wo
Hu	0,225	0,225	0,3	0,263	0,25	0,231	0,2	0,275	0,263	0,25	0,263	0,282	0,323	0,323	0,286	0,286	0,278	0,361	0,296	0	<25, 5>	<20,7>	<15,12>	<20,11>	Hu
Fin	0,241	0,207	0,31	0,259	0,241	0,241	0,241	0,379	0,345	0,258	0,233	0,3	0,3	0,3	0,296	0,37	0,296	0,407	0,4	0,167	0	<20,6>	<17, 7>	<22, 9>	Fin
Hi	0,2	0,2	0,267	0,25	0,2	0,207	0,214	0,25	0,25	0,2	0,2	0,207	0,185	0,148	0,24	0,385	0,28	0,385	0,4	0,259	0,231	0	<15, 8>	<23, 5>	Hi
Bas	0,438	0,394	0,438	0,469	0,406	0,387	0,433	0,533	0,387	0,419	0,355	0,452	0,5	0,458	0,52	0,519	0,517	0,552	0,417	0,444	0,292	0,348	0	<16, 9>	Bas
Slo	0,212	0,152	0,212	0,258	0,212	0,242	0,182	0,182	0,312	0,171	0,212	0,147	0,118	0,0882	0,129	0,258	0,3	0,333	0,435	0,355	0,29	0,179	0,36	0	Slo
	It	Sal	Sp	Fr	Ptg	Rm	Gri	Grk	E	D	Nor	Blg	SC	Rus	Ir	Wel	Heb	Ar	Wo	Hu	Fin	Hi	Bas	Slo	

DISTANCES



Kitsch



UPGMA

Pairs	Distance
SC-Rus	0,0286
It-Ptg	0,0377
It-Fr	0,0392
It-Sal	0,06
Ir-Wel	0,0698
Sp-Ptg	0,0755
Fr-Ptg	0,0784
Sal-Ptg	0,08
Blg-SC	0,0882
Rus-Slo	0,0882
E-Nor	0,093
Sal-Fr	0,104
It-Rm	0,106
Ptg-Rm	0,106
D-Nor	0,109
It-Sp	0,113
E-D	0,114
Sp-Fr	0,118
Blg-Rus	0,118
SC-Slo	0,118
Gri-SC	0,121
Sal-Gri	0,122
It-Gri	0,128
Sal-Rm	0,128
Sp-Rm	0,128
Ir-Slo	0,129
It-D	0,13
Sal-D	0,13
Ptg-D	0,13
Rm-Gri	0,13
Gri-Grk	0,13
Rm-Nor	0,133
It-Nor	0,136
Sal-Nor	0,136
Ptg-Nor	0,136
Grk-SC	0,147
Blg-Slo	0,147
Rus-Hi	0,148
Ptg-Gri	0,149
Sal-Slo	0,152
Gri-Rus	0,152
Fr-Rm	0,156
Gri-D	0,156
It-E	0,159
Fr-D	0,159
Ptg-E	0,159
Gri-Ir	0,159
Heb-Ar	0,159
Sal-Sp	0,16
Gri-Blg	0,163
Fr-Nor	0,167
Hu-Fin	0,167
D-Slo	0,171
It-Blg	0,174
Sp-D	0,174
Ptg-Blg	0,174
Grk-Rus	0,176
Sal-Blg	0,178
Fr-Gri	0,178
Nor-Blg	0,178
Hi-Slo	0,179
Sp-Nor	0,182
Gri-Slo	0,182
Grk-Slo	0,182
SC-Hi	0,185
Sal-E	0,186
Fr-E	0,186
Sal-Rus	0,188
Sp-Ir	0,19
Gri-Nor	0,19
D-Ir	0,19
Blg-Heb	0,195
It-Hi	0,2
Sal-Hi	0,2
Ptg-Hi	0,2
Rm-D	0,2
Rm-Blg	0,2
Gri-Hu	0,2
D-Hi	0,2
Nor-Hi	0,2
Rus-Ir	0,2
Sal-Ir	0,205
Sp-E	0,205
Rm-Grk	0,205
Grk-Blg	0,205
E-Heb	0,205
Nor-Heb	0,205
D-Rus	0,206
Sal-Fin	0,207
Rm-Hi	0,207
Blg-Hi	0,207
It-Heb	0,209
It-Rus	0,212
It-Slo	0,212
Sp-Slo	0,212
Ptg-Rus	0,212
Ptg-Slo	0,212
Rm-Rus	0,212
Nor-Slo	0,212
It-Grk	0,213
Sp-Gri	0,213
Sp-Grk	0,213
It-Ir	0,214
Ptg-Ir	0,214
Rm-E	0,214
Gri-Hi	0,214
Sp-Blg	0,217
Sal-SC	0,219
Gri-Wo	0,219
Gri-E	0,22
D-Wel	0,22
Fr-Ir	0,225
Grk-Ir	0,225
Fr-Blg	0,227
D-Blg	0,227
Rm-Heb	0,231
Rm-Hu	0,231
E-Wel	0,231
Nor-Ir	0,231
Ir-Heb	0,231
Fin-Hi	0,231
Ptg-Heb	0,233
E-Blg	0,233
Nor-Fin	0,233
SC-Ir	0,233
Ptg-Grk	0,234
D-SC	0,235
Grk-Ar	0,238
Ir-Hi	0,24
It-Fin	0,241
Ptg-Fin	0,241
Rm-Fin	0,241
Gri-Fin	0,241
It-SC	0,242
Ptg-SC	0,242
Rm-SC	0,242
Rm-Slo	0,242
E-Ir	0,243

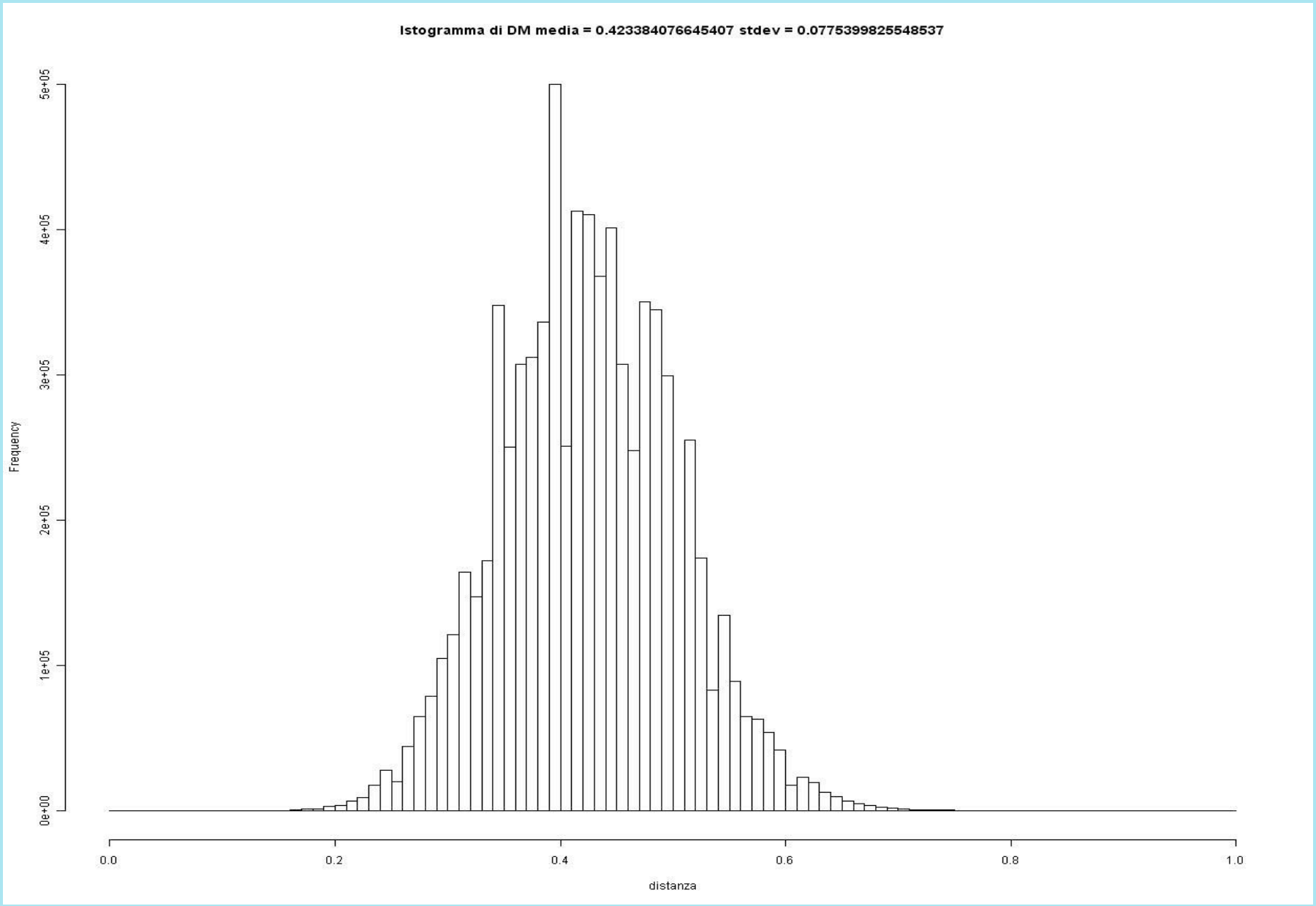
It-Hu	0,244
Sal-Grk	0,244
Sal-Hu	0,244
Rm-Ir	0,244
Gri-Heb	0,244
Sal-Heb	0,25
Fr-Hi	0,25
Grk-Hi	0,25
E-Hi	0,25
D-Heb	0,25
D-Hu	0,25
Nor-Rus	0,25
Sp-Heb	0,256
Gri-Wel	0,256
Fr-Rus	0,258
Fr-Slo	0,258
D-Fin	0,258
Wel-Slo	0,258
Fr-Fin	0,259
Hu-Hi	0,259
E-Hu	0,263
Nor-Hu	0,263
Sp-Hi	0,267
Fr-Grk	0,267
Grk-D	0,267
Rus-Heb	0,267
Fr-Heb	0,268
Ptg-Hu	0,268
Grk-Heb	0,268
Heb-Hu	0,27
Sp-Rus	0,273
Sp-Ar	0,273
Grk-Hu	0,275
Blg-Hu	0,275
Heb-Hi	0,28

E-Rus	0,281
Nor-SC	0,281
Fr-Hu	0,282
Blg-Ir	0,282
Sp-Wel	0,286
Ir-Hu	0,286
Wel-Hu	0,286
Wo-Hu	0,286
Nor-Wel	0,289
Ir-Ar	0,289
Fr-SC	0,29
Fin-Slo	0,29
Fin-Bas	0,292
Sp-Hu	0,293
It-Ar	0,295
Sal-Wel	0,295
Ir-Fin	0,296
Heb-Fin	0,296
Blg-Fin	0,3
SC-Heb	0,3
SC-Fin	0,3
Rus-Fin	0,3
Heb-Slo	0,3
Sp-SC	0,303
It-Wel	0,31
Sp-Fin	0,31
Ptg-Wel	0,31
Grk-Nor	0,31
E-SC	0,312
E-Slo	0,312
Fr-Wel	0,317
Rm-Wel	0,317
Ptg-Ar	0,318
SC-Hu	0,323
Rus-Hu	0,323

Rm-Ar	0,333
Grk-E	0,333
Grk-Wel	0,333
Grk-Wo	0,333
Nor-Ar	0,333
SC-Wo	0,333
Rus-Wel	0,333
Ar-Slo	0,333
Blg-Ar	0,341
Wel-Heb	0,342
E-Fin	0,345
Hi-Bas	0,348
Ar-Hu	0,351
Nor-Bas	0,355
Hu-Slo	0,355
Fr-Ar	0,357
E-Ar	0,359
Bas-Slo	0,36
Sal-Ar	0,364
Sal-Wo	0,364
Gri-Ar	0,366
Rm-Wo	0,367
SC-Wel	0,367
Wel-Fin	0,37
D-Ar	0,375
Rus-Wo	0,375
Wel-Ar	0,378
Grk-Fin	0,379
Blg-Wel	0,385
Wel-Hi	0,385
Ar-Hi	0,385
Rm-Bas	0,387
E-Bas	0,387
D-Wo	0,387
Blg-Wo	0,387

It-Wo	0,394
Ptg-Wo	0,394
Rus-Ar	0,4
Wo-Fin	0,4
Wo-Hi	0,4
Wo-Bas	0,4
Ar-Fin	0,407
Sal-Bas	0,412
Ir-Wo	0,414
D-Bas	0,419
Sp-Wo	0,424
Sp-Bas	0,424
Fr-Wo	0,424
Ptg-Bas	0,424
Hu-Bas	0,429
Gri-Bas	0,433
SC-Ar	0,433
Wo-Slo	0,435
Blg-Bas	0,438
It-Bas	0,455
Rus-Bas	0,458
E-Wo	0,467
Heb-Wo	0,469
Nor-Wo	0,483
Wel-Wo	0,483
Fr-Bas	0,485
SC-Bas	0,5
Heb-Bas	0,5
Ar-Wo	0,5
Wel-Bas	0,519
Ir-Bas	0,52
Grk-Bas	0,533
Ar-Bas	0,533

From Longobardi,Bortolussi, Guardiano, Sgarro: in prep.



'Anti-Babelic principle'

(Guardiano&Longobardi 2005)

Similarities among languages can be due either to ***historical cause*** *(relatedness/convergence) or to* ***chance****, differences can only be due to* ***chance***

(no one ever made grammars diverge on purpose!)

SYNTACTIC DISTANCES

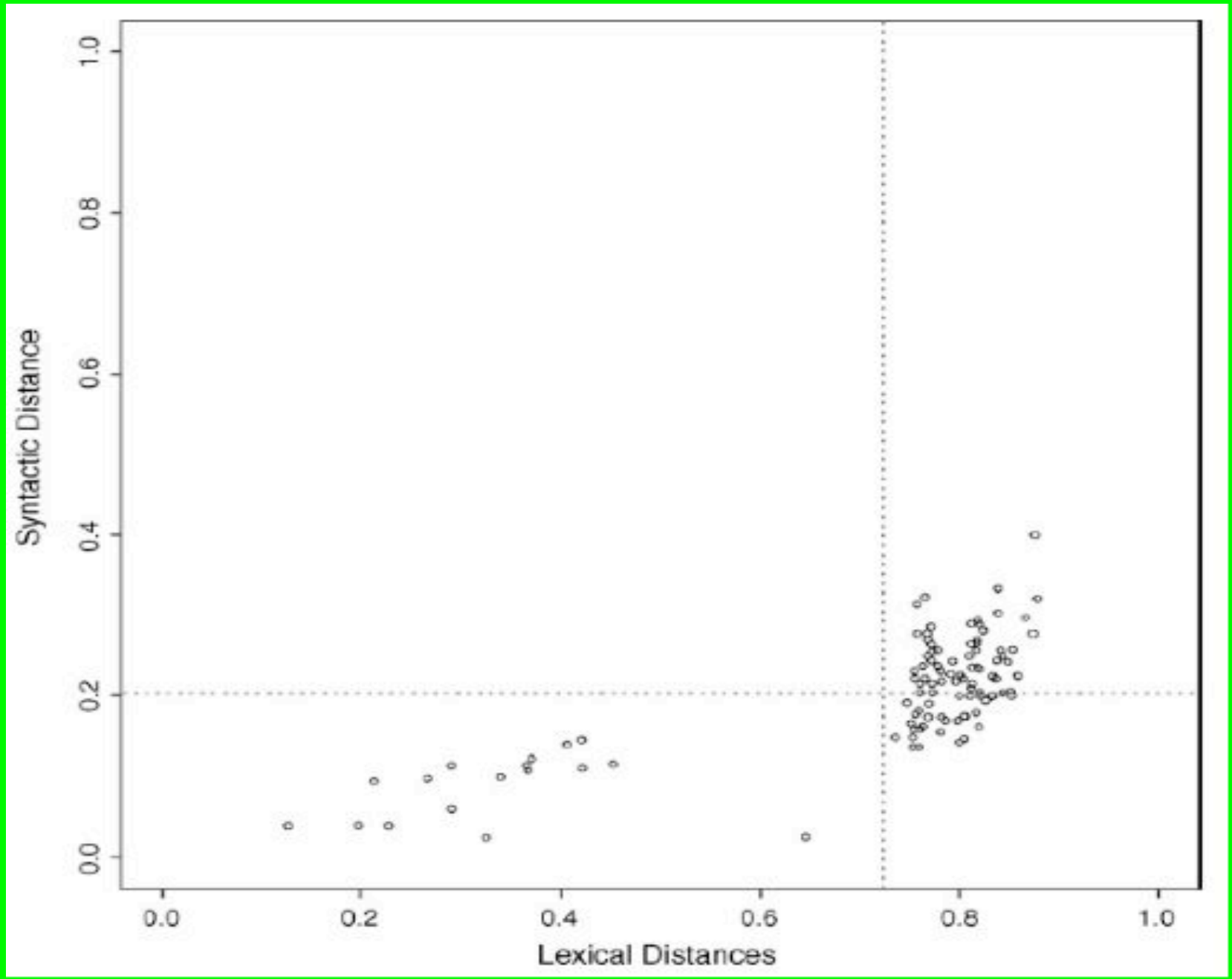
	Bulgarian	English	French	German	Greek	Hindi	Irish	Italian	Norwegian	Portugese	Rumanian	Russian	Serbocroat	Slovenian	Spanish	Welsh	
Bulgarian	0	0,233	0,209	0,227	0,205	0,207	0,282	0,156	0,178	0,156	0,2	0,118	0,0882	0,147	0,222	0,385	Bulgarian
English	0,772	0	0,186	0,114	0,333	0,25	0,243	0,159	0,093	0,159	0,214	0,281	0,312	0,312	0,205	0,231	English
French	0,791	0,764	0	0,159	0,267	0,25	0,225	0,04	0,167	0,08	0,156	0,258	0,29	0,258	0,1	0,317	French
German	0,769	0,422	0,756	0	0,267	0,2	0,19	0,13	0,109	0,13	0,2	0,206	0,235	0,171	0,174	0,22	German
Greek	0,811	0,838	0,843	0,812	0	0,25	0,225	0,213	0,31	0,234	0,205	0,176	0,147	0,182	0,213	0,333	Greek
Hindi	0,801	0,854	0,824	0,853	0,874	0	0,24	0,2	0,2	0,2	0,207	0,148	0,185	0,179	0,267	0,385	Hindi
Irish	0,818	0,817	0,812	0,806	0,859	0,878	0	0,214	0,231	0,214	0,244	0,2	0,233	0,129	0,19	0,0698	Irish
Italian	0,769	0,753	0,197	0,735	0,822	0,818	0,8	0	0,136	0,0385	0,106	0,212	0,242	0,212	0,0962	0,31	Italian
Norwegian	0,773	0,452	0,77	0,367	0,821	0,852	0,836	0,754	0	0,136	0,133	0,25	0,281	0,212	0,182	0,289	Norwegian
Portugese	0,781	0,76	0,291	0,753	0,833	0,813	0,817	0,227	0,761	0	0,106	0,212	0,242	0,212	0,0577	0,31	Portugese
Rumanian	0,798	0,773	0,421	0,751	0,843	0,827	0,837	0,34	0,786	0,371	0	0,212	0,242	0,242	0,128	0,317	Rumanian
Russian	0,365	0,758	0,778	0,755	0,832	0,8	0,782	0,761	0,758	0,773	0,781	0	0,0286	0,0882	0,273	0,333	Russian
Serbocroat	0,291	0,766	0,772	0,764	0,821	0,805	0,796	0,755	0,772	0,766	0,778	0,325	0	0,118	0,303	0,367	Serbocroat
Slovenian	0,385	0,751	0,782	0,733	0,821	0,8	0,809	0,76	0,762	0,781	0,79	0,386	0,316	0	0,212	0,258	Slovenian
Spanish	0,782	0,76	0,266	0,747	0,833	0,819	0,805	0,212	0,761	0,126	0,406	0,769	0,768	0,772	0	0,286	Spanish
Welsh	0,838	0,841	0,81	0,82	0,867	0,876	0,645	0,793	0,849	0,804	0,812	0,818	0,821	0,838	0,813	0	Welsh
	Bulgarian	English	French	German	Greek	Hindi	Irish	Italian	Norwegian	Portugese	Rumanian	Russian	Serbocroat	Slovenian	Spanish	Welsh	

LEXICAL DISTANCES

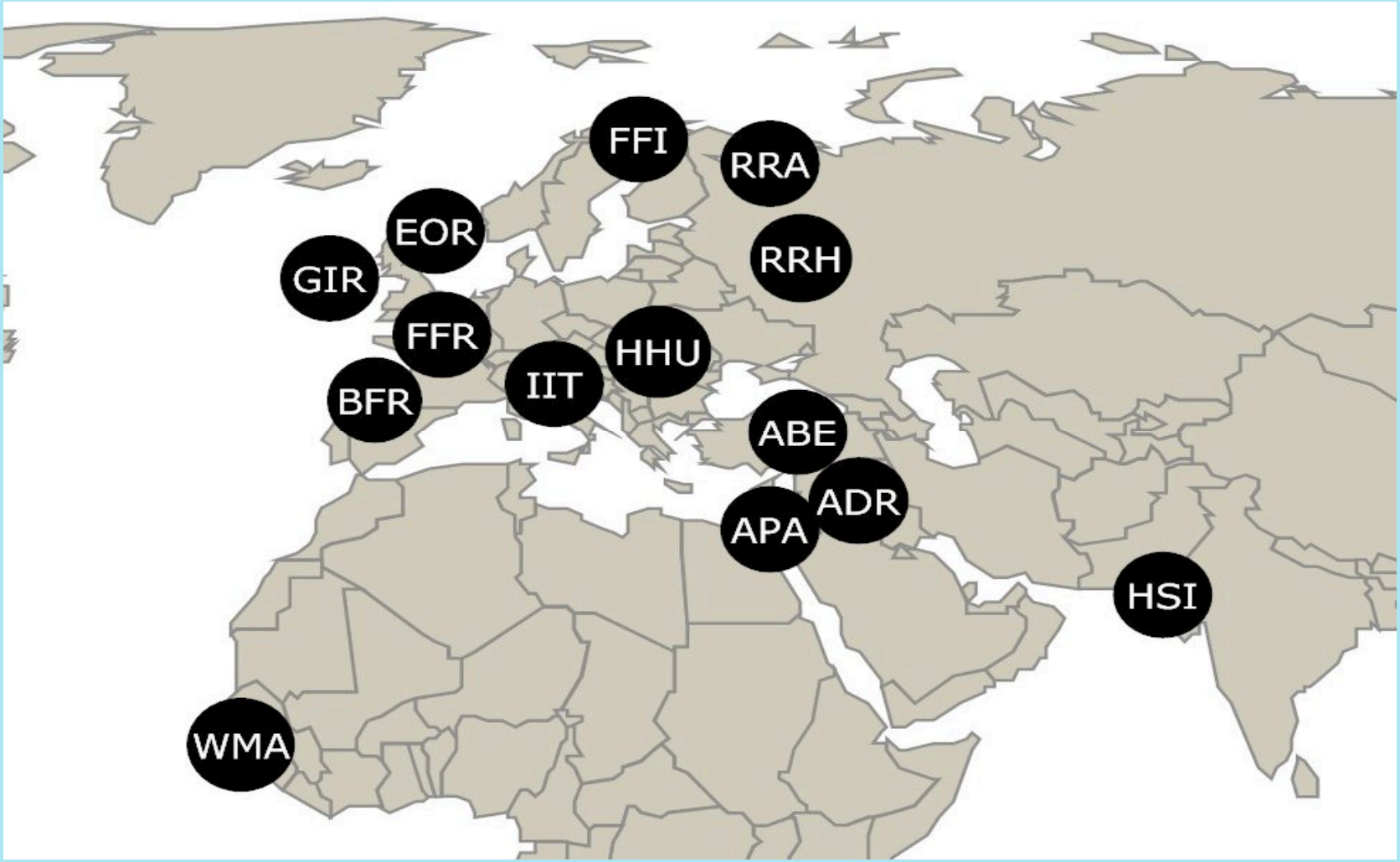
Syntactic distances are about 1/4 the corresponding lexical distances (attended ratio 2/3)

Indirect argument for some version of an Inertial Theory (Keenan 1994) of diachronic syntax

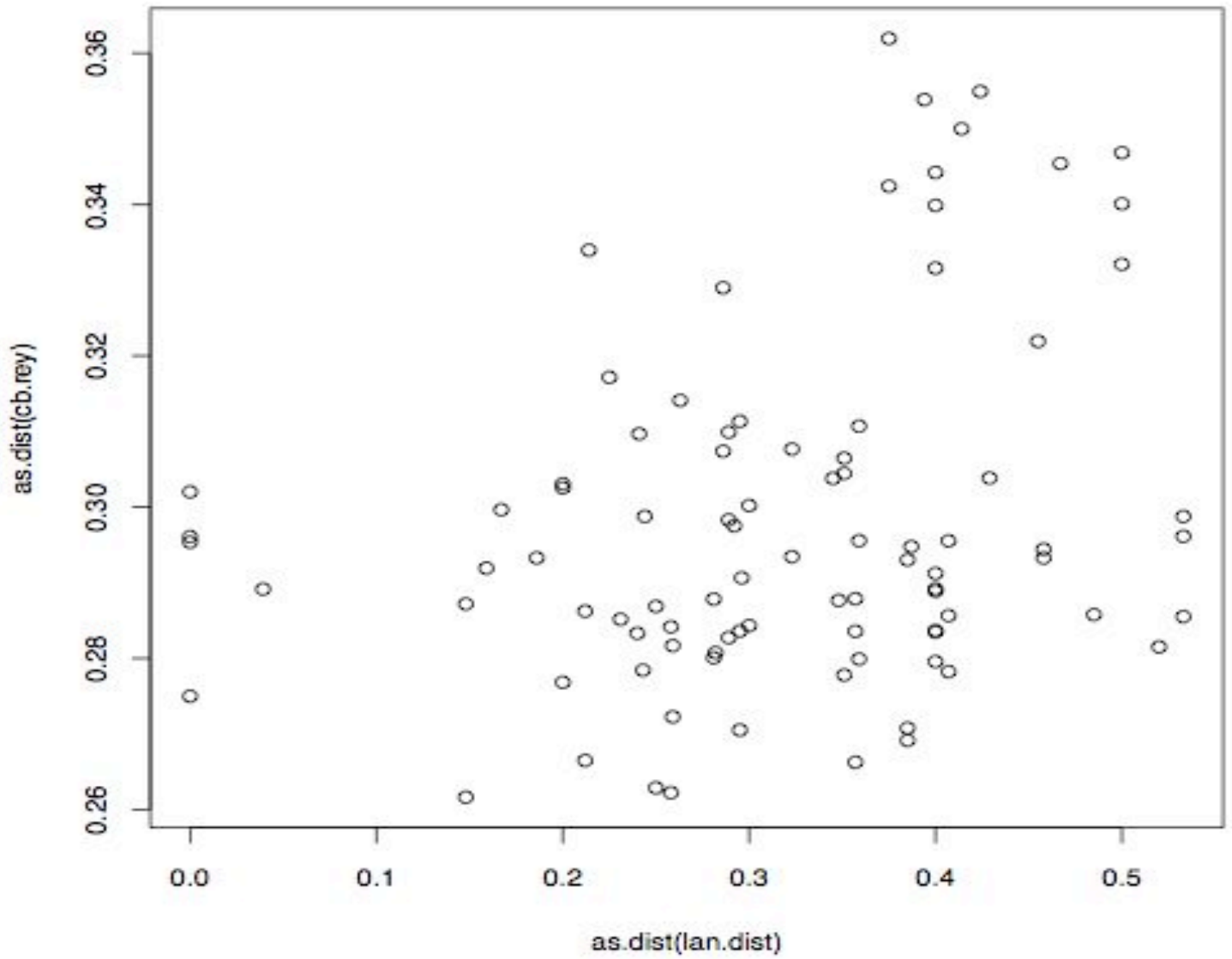
Scatter plot: syntax/lexicon



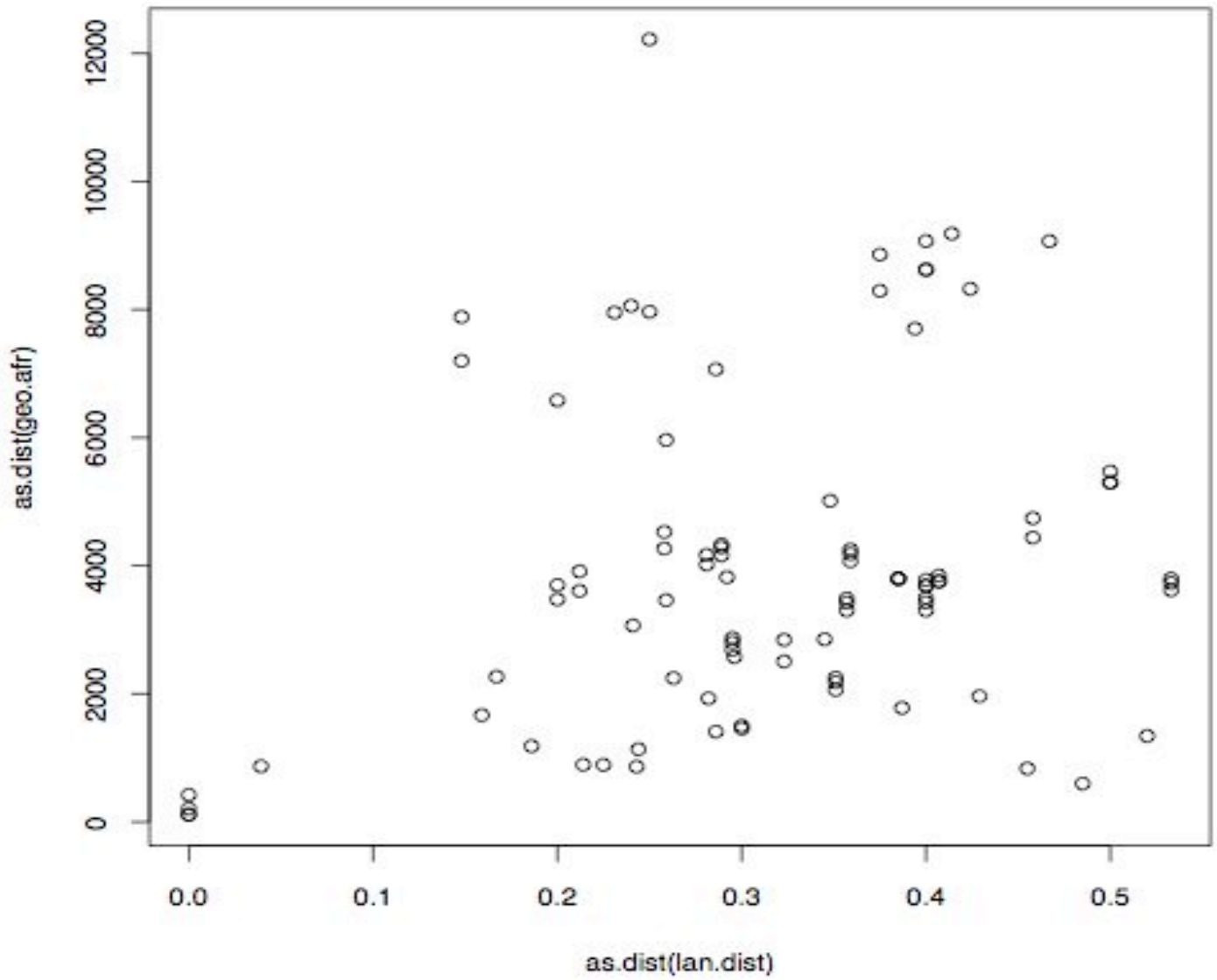
From Colonna, Boattini, Dall'Ara, Guardiano, Pettener, Longobardi, Barbujani: in *Human Heredity*, 2010



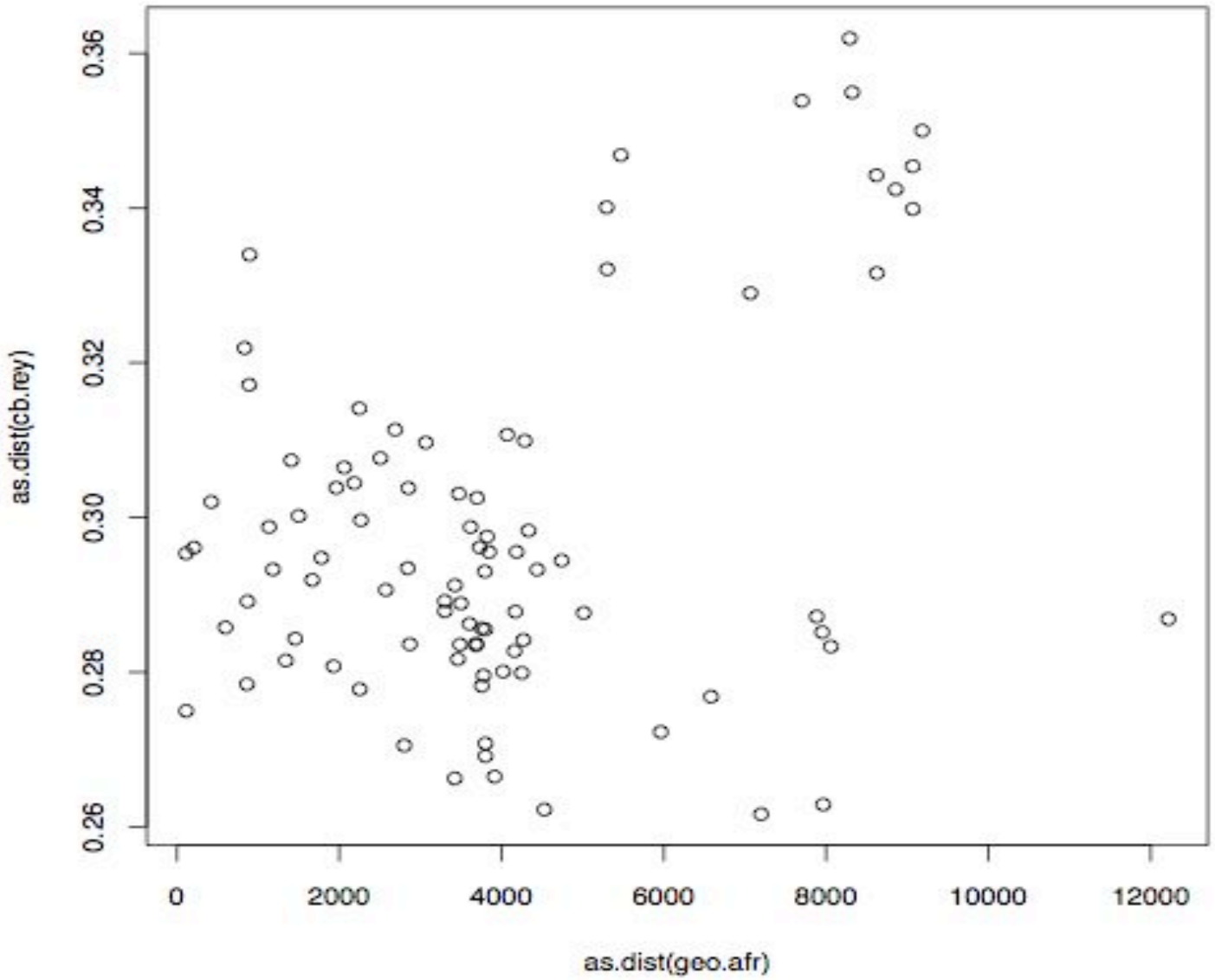
Languages and genes



Languages and geography



Genes and geography



Grammar as a population science: overcoming a XX century bias

In biology a sort of Galilean revolution (molecular analyses) has affected both theoretical and evolutionary work leading to formal theories of biological diversity

In linguistics a Galilean revolution in the study of grammatical theory has been proposed by Chomsky but it has not yet been fully exploited for a mathematical approach to linguistic diversity