

Computations on Language Family Dynamics

Harald Hammarström

Chalmers University

Gothenburg, Sweden

harald2@cs.chalmers.se

Languages Families and their Sizes

There are some 7 000 languages in the world (Gordon, 2005). A language family is defined as a (possibly one-member) set of languages that which have been shown, in publication, using orthodox comparative methodology (Campbell, 2004) to stem from a common ancestor. One classification which aims to consistently follow this definition recognizes the ca 7 000 languages to fall into some 400 families (Hammarström, 2007). As is also the case for alternative classifications, the ca 400 families are of very unequal size. In fact, we can observe the family sizes¹ to follow a power-law distribution (Wichmann, 2005). In other words, there are very many tiny families and a few very large families, as shown on the next page.

Explaining Language Family Sizes

There are two main trends in explaining why some families are big and some are small; a) one that observes that some big families can be associated with a technological advantage, most often agriculture, and thus hypothesize that such families are big and the rest are small (Bellwood and Renfrew, 2002), and b) one that observes that stochastic branching processes (which seem like reasonable simplifications of how real-world language families come about through human migration), are expected to yield just the kind of power-law distributions that we observe with language family sizes (Wichmann, 2005; Arnold and Bauer, 2006).

In this talk we will not explore the merits of the a)-theories since it is too big a question for a talk. Instead, we will focus on the stochastic models (and their parameters) that are likely to give rise to the observed language family sizes, in the following novel way. The world can be divided into

six fair size geographical areas – Eurasia, N America, S America, Africa, New Guinea and Australia – all relatively separated from each other. A plausible stochastic branching model for the language family sizes in the world as a whole should also fit each of the six disjoint areas individually.

References

- R. Arnold and L. Bauer. 2006. A note regarding 'on the power-law distribution of language family sizes'. *Journal of Linguistics*, 42:373–376.
- Peter Bellwood and Colin Renfrew, editors. 2002. *Examining the farming/language dispersal hypothesis*. McDonald Institute Monographs. McDonald Institute for Archaeological Research, Oxford.
- L. Campbell. 2004. *Historical Linguistics: An Introduction*. MIT Press, 2 edition.
- Raymond G. Gordon, Jr., editor. 2005. *Ethnologue: Languages of the World*. SIL International, Dallas, 15 edition.
- Harald Hammarström. 2007. The language families of world: A critical synopsis. Manuscript available at http://www.cs.chalmers.se/~harald2/language_families_full.pdf accessed 25 Sept 2007.
- S. Wichmann. 2005. On the power-law distribution of language family sizes. *Journal of Linguistics*, 41:117–131.

¹The size of a family is defined as the number of its member languages.

Austronesian	1275	Tupi	76	Otomanguean	179	Indo-European	449
Trans New Guinea	338	Arawak	62	Mayan	69	Sino-Tibetan	402
Torriceili	53	Quechuan	46	Uto-Aztecan	61	Austroasiatic	168
Sepik	48	Carib	32	Eyak-Athapaskan-Tlingit	45	Tai-Kadai	76
Lower Sepik-Ramu	33	Panoan	28	Algic	44	Dravidian	73
Lakes Plain	20	Tucanoan	25	Salishan	27	Turkic	40
Morehead and Upper Maro Rivers	17	Chibchan	21	Mixe-Zoque	17	Uralic	39
North Halmahera	16	Je	16	Siouan	17	Miao-Yao	35
West Timor-Alor-Pantar	16	Chocoan	12	Eskimo-Aleut	11	Nakh-Dagestanian	29
Border	15	Arawa	8	Iroquoian	11	Mongolian	14
Angan	13	Matacoan	7	Totonacan	11	Japanese	12
Tor-Orya	13	Barbacoan	6	Miwok-Costanoan	10	Tungusic	12
Binanderean	11	Bora-Huitoto	6	Cochimi-Yuman	9	Great Andamanese	10
Cenderawasih Bay	11	Tacanan	6	Chumashan	7	Abkhaz-Adyge	5
Bosavi	10	Zaparoan	6	Pomoan	7	Chukotko-Kamchatkan	5
South Bougainville	9	Chapacura-Wanham	5	Kiowa-Tanoan	6	Kartvelian	5
Kwerba	8	Guahibo	5	Muskogean	6	Basque	3
Kojarian	7	Guaicuruan	5	Caddoan	5	Hurro-Urartian	2
Sko	7	Lengua-Mascoy	5	Sahaptian	5	Jarawa-Onge	2
Baining	6	Jivaro	4	Wakashan	5	Yeniseian	2
Eleman	6	Nadahup	4	Huavean	4	Yukaghir	2
Kiwaian	6	Yanomam	4	Maiduan	4	Ainu	1
Left May	6	Aymara	3	Misumalpan	4	Burushaski	1
Marind	6	Bororo	3	Tsimshian	3	Elamite	1
South Bird's Head Proper	6	Charrua	3	Chimakuan	2	Etruscan	1
Yuat-Maramba	6	Katukina	3	Chinook	2	Hattic	1
Goilalan	5	Nambiquaran	3	Haida	2	Iberian	1
Nimboran	5	Saliban	3	Keresan	2	Korean	1
Tirio	5	Araucanian	2	Palaihnihan	2	Kusunda	1
West Bird's Head	5	Cahuapanan	2	Tarascan	2	Nihali	1
Eastern Trans-Fly	4	Chonan	2	Tequistlatecan	2	Nivkh	1
Kaure-Kapori	4	Harakmbut	2	Adai	1	Shom Pen	1
North Bougainville	4	Hibito-Cholon	2	Alsea	1	Sumerian	1
Pauwasi	4	Huarpean	2	Atakapan	1	Atlantic-Congo	1400
Sentani	4	Jabuti	2	Beothuk	1	Afro-Asiatic	346
Suki-Gogodala	4	Kakua-Nukak	2	Cayuse	1	East Sudanic	92
Walio	4	Kawesqar	2	Chimariko	1	Mande	71
Awin-Pa	3	Peba-Yagua	2	Chitimacha	1	Central Sudanic	66
East Bird's Head	3	Uru-Chipaya	2	Coahuilteco	1	Omoti	24
East Timor	3	Zamucoan	2	Comecrudan	1	Khoe-Kwadi	13
Kayagar	3	Aikana	1	Coosan	1	Heiban	10
Kolana-Tanglapui	3	Aimore	1	Cotoname	1	Ijoid	10
Kolopom	3	Andaqui	1	Cuitlatec	1	Dogon	9
Mairasi	3	Andoque	1	Esselen	1	Maban	9
Mongol-Langam	3	Arara do Rio Branco	1	Guaicurian	1	Saharan	9
Turama-Kikori	3	Atacame	1	Jicaquean	1	Narrow Talodi	8
West Bomberai	3	Atacameno	1	Kalapuyan	1	Songhay	8
Amto-Musan	2	Awake	1	Karakawa	1	Ju	6
Baibai	2	Awshiri	1	Karuk	1	Kadugli-Krongo	6
Bayono-Awbono	2	Betoi-Jirara	1	Klamath-Modoc	1	Koman	5
Biksi	2	Candoshi-Shapra	1	Kutenai	1	Tuu	5
Bulaka River	2	Canichana	1	Lencan	1	Mao	4
Duna-Bogaya	2	Cayuvava	1	Maratino	1	Furan	3
East Kutubu	2	Chiquitano	1	Molala	1	Kuliak	3
Inanwatan	2	Chono	1	Natchez	1	Rashad	3
Konda-Yahadian	2	Cofan	1	Salinan	1	Katla-Tima	2
Kwomtari	2	Culli	1	Seri	1	Bangi Me	1
Lower Mamberamo	2	Fulnio	1	Shasta	1	Berta	1
Maybrat	2	Guamo	1	Siuslaw	1	Gumuz	1
Mombum	2	Guato	1	Takelma	1	Hadza	1
Pahoturi	2	Iranxe	1	Timucua	1	Ho	1
Piawi	2	Itonama	1	Tonkawa	1	Jalaa	1
Senagi	2	Jirajaran	1	Tunica	1	Kujarge	1
Teberan	2	Jodi	1	Wappo	1	Kunama	1
Yawa	2	Kamakan	1	Washo	1	Laal	1
Abinomn	1	Kamsa	1	Wintuan	1	Meroitic	1
Abun	1	Kanoe	1	Xincan	1	Ongota	1
Anem	1	Karaja	1	Yana	1	Sandawe	1
Arafundi	1	Kariri	1	Yokutsan	1	Shabo	1
Ata	1	Kwaza	1	Yuchi	1	Tegem	1
Bilua	1	Leko	1	Yuki	1		
Burmeso	1	Lule	1	Zuni	1		
Busa	1	Maku	1	Pama-Nyungan	175		
Dem	1	Matanawi	1	Western Daly	10		
Duranmin	1	Maxakali	1	Gunwinyguan	9		
Elseng	1	Mochica	1	Nyulnyulan	9		
Fasu	1	Moseten-Chimane	1	Worrorran	7		
Guriaso	1	Movima	1	Iwaidjan Proper	5		
Hatam	1	Muniche	1	Mirdi	5		
Kaki Ae	1	Mura-Piraha	1	Mangarrayi-Maran	4		
Karkar	1	Ofaie	1	Maningrida	4		
Kehu	1	Omurano	1	Minkin-Tangkic	4		
Kembra	1	Oti	1	Giimbiyu	3		
Kol	1	Otomaco	1	Jarrakan	3		
Kuot	1	Paez	1	Tasmanian 1	3		
Lavukaleve	1	Pankararu	1	Tasmanian 2	3		
Lepki	1	Puelche	1	Yangmanic	3		
Masep	1	Puinave	1	Anson Bay	2		
Molof	1	Puquina	1	Bunaban	2		
Mor	1	Puri	1	Eastern Daly	2		
Moraori	1	Rikbaktsa	1	Limilngan	2		
Mpur	1	Sape	1	Marrku-Wurrugu	2		
Murkim	1	Securan	1	Northern Daly	2		
Oksapmin	1	Tallan	1	Southern Daly	2		
Papi	1	Taruma	1	Umbugarla-Ngurmbur	2		
Pawaia	1	Taushiro	1	Anindilyakwa	1		
Porome	1	Ticuna	1	Gaagudju	1		
Pyu	1	Timote-Cuica	1	Garwan	1		
Savosavo	1	Tinigua	1	Kungarakany	1		
Sulka	1	Trumai	1	Larrakiyan	1		
Taiap	1	Tuxa	1	Tiwi	1		
Tambora	1	Urarina	1	Wagiman	1		
Tanahmerah	1	Vilela	1				
Taulil-Butam	1	Waorani	1				
Tofanma	1	Warao	1				
Touo	1	Xukuru	1				
Uhunduni	1	Yamana	1				
Usku	1	Yaruro	1				
Waia	1	Yurakare	1				
Yale	1	Yuri	1				
Yeli Dnye	1	Yurumangui	1				