Universal grammar, language evolution, and documenting an ancient language

Anvita Abbi

Proceedings of Conference on
Language Documentation & Linguistic Theory 3

Edited by Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan

19-20 November 2011 School of Oriental and African Studies, University of London

Hans Rausing Endangered Languages Project
Department of Linguistics
School of Oriental and African Studies
Thornhaugh Street, Russell Square
London WC1H 0XG
United Kingdom

Department of Linguistics:
Tel: +44-20-7898-4640
Fax: +44-20-7898-4679
linguistics@soas.ac.uk
http://www.soas.ac.uk/academics/departments/linguistics

Hans Rausing Endangered Languages Project:
Tel: +44-20-7898-4640
Fax. +44-20-7898-4349
elap@soas.ac.uk
http://www.hrelp.org

ISBN: 978-0-7286-0398-1

This publication can be cited as:

Anvita Abbi. 2011. Universal grammar, language evolution, and documenting an ancient language. In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*, 27-38. London: SOAS.

or:

Anvita Abbi. 2011. Universal grammar, language evolution, and documenting an ancient language. In Peter K. Austin, Oliver Bond, Lutz Marten & David Nathan (eds) *Proceedings of Conference on Language Documentation and Linguistic Theory 3*. London: SOAS. www.hrelp.org/eprints/ldlt3_04.pdf

Universal grammar, language evolution, and documenting an ancient language

ANVITA ABBI
School of Oriental and African Studies, University of London

1. INTRODUCTION

Working on and documenting an ancient language, and one which has been isolated from other language communities for thousands of years, is not only a challenge in documentation but also a gift to linguists in search of the parameters of language evolution. The structures of such an 'early' language could shed light on to what must have been the basic or prime linguistic features involved in language evolution and language universals. Present-day Great Andamanese (PGA) is such a language.

PGA belongs to the sixth language family of India (Abbi 2006, 2009, Blevins 2007) and is spoken in the Andaman Islands, the Union territory of India. Refer to Map 1. It is a koiné which derives its lexicon from four north Great Andamanese languages: Khora, Bo, Jeru and Sare. Sporadic interferences from the central variety, Aka Pucikwar, have been observed in its present form. It is a moribund language and has only five speakers (mainly of Jeru and Sare) remaining.

The data presented in this paper were drawn from nine fluent speakers as speakers of Khora and Bo were still alive when I conducted fieldwork from 2005 to 2009[1]. The indigenous populations of the Andaman Islands seem to have remained in isolation for a much longer period than any known ancient population of the world (Kashyap et al. 2003) and thus the language has retained some very ancient and unusual linguistic structures not shared by any other language known to the author.

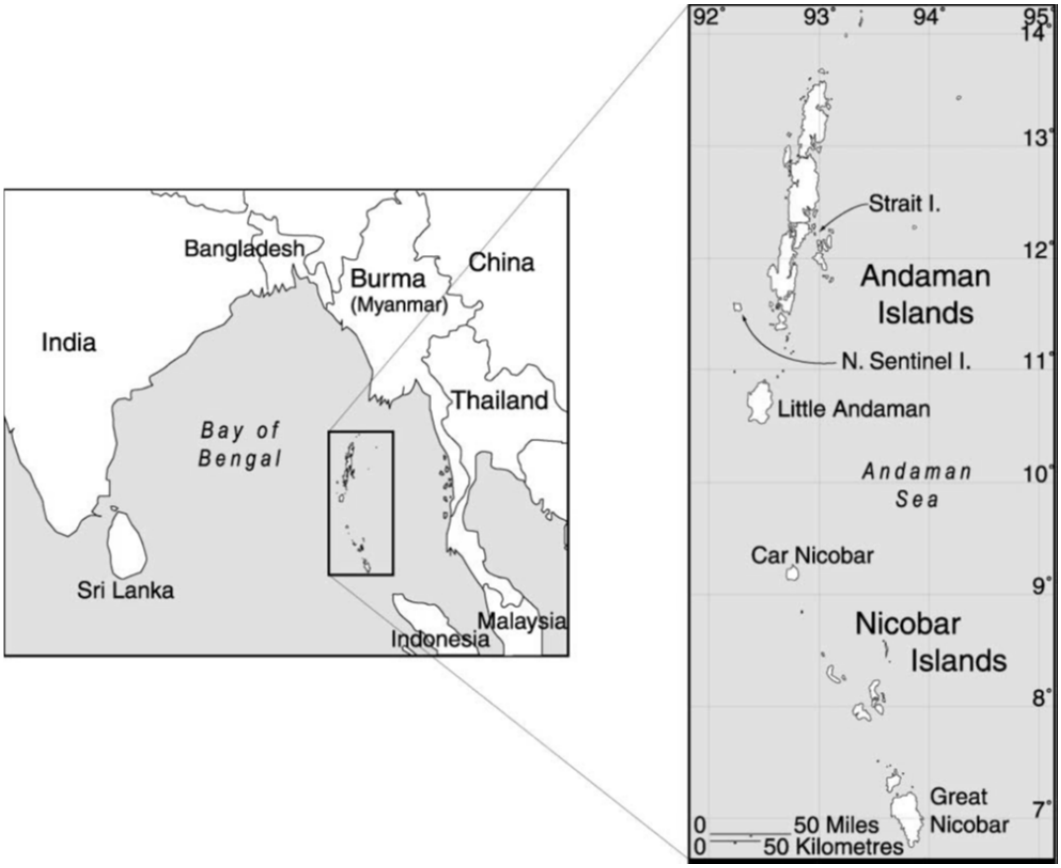
The Andamanese are considered to be the last survivors of the first migration out of Africa that took place 70,000 years ago (Thangaraj et al. 2005). Until about three hundred years ago, the Great Andamanese tribes were spread all over the Andaman Islands, divided into ten different sub-groups. What remains is a mixed group of people comprising fifty members, who are descendants of different sub-groups. The community uses a variety of Hindi, called Andamani Hindi, for intra-group communication as well for communicating with non-indigenous people.

I will first give a brief summary of various theoretical issues concerning innateness and Universal Grammar (UG) and then move on to describe, in brief, the salient structures of PGA which do not conform to the prototypical language universals.

---

[1] My fieldwork on Great Andamanese was supported by the Hans Rausing Endangered Languages Project, SOAS, University of London, under the Endangered Languages Documentation Programme for the project *Vanishing Voices of the Great Andamanese (VOGA)*, 2005-2009. I am thankful to Andrej Malchukov, Andrew Spencer, Balthasar Bickel and Tania Kuteva for their constructive criticism and suggestions made on various issues raised in this paper.

Map 1
Map of Southeast Asia and the location of the Andaman and Nicobar Islands, Republic of India.



1.1. Innate cognitive modules and universal grammar

Linguists have proposed several lists of innate cognitive modules, and an equal number of methodologies to test them (Tomasello 2004:642). Although innate cognitive modules might be expected to constitute a defined set, every experimenter and linguist working on the identification of innateness in human language has a different list of innate cognitive modules and there is no consensus on the methods for deciding on them. O'Grady (1997) proposes a division between lexical and grammatical categories; Jackendoff (2002) linking rules, movement rules and grammatical morphology; Crain and Lillo-martin (1999) propose some syntactic rules including Wh-movement and C-command. Hauser, Chomsky and Fitch (2002) suggest that recursion is fundamental, while Hauser, Chomsky and Fitch (2004) posit the syntactic operation of 'merge' as innate. Baker (2001) and Wunderlich (2004) cover a large number of grammatical features found across languages. However, a large number of lesser-known languages are neither represented in these lists of parameters nor show any overt or covert movement rules.

The recent debate on UG has highlighted the fact that there is no fixed set of 'universals' nor do the world's languages conform to any 'common pattern'. The

most debated issue has been the question of fixed universals in human language against the new discoveries of language varieties explored by linguists and typologists in last 30 years (Wunderlich 2004, Evans and Levinson 2009, 2010). The number of 'exceptions' to proposals for UG increases as more languages are researched. This leads us to raise the fundamental question as to how human language evolved or more precisely, what could be the structure of human language in its early stage of evolution? This paper aims to answer this question by presenting very unusual structures of Great Andamanese, a moribund language of the Andaman Islands.

The analysis casts doubts on several proposed features of UG, but more significantly draws our attention to the phenomena of dependency and inalienability in grammar. PGA is a head-marking polysynthetic and agglutinative language with an SOV clausal pattern. It has a very elaborate system for marking inalienability (Abbi 2006, 2010) which is based on seven possessive markers designating different body-divisions. These markers are further grammaticalized in the language and appear as proclitics which classify a large number of lexical items as dependent categories. The author proposes that the Great Andamanese conceptualize their world through these interdependencies. The grammar of the language encodes this important phenomenon in every grammatical category expressing referential, attributive and predicative meaning. In addition, the language offers several structures which refute some of the proposed parameters of UG.

2. ABSENCE OF SOME UG FEATURES

PGA lacks the following features which are often considered to be part of UG.

2.1 Strict dichotomy between form classes

PGA does not maintain a strict distinction between noun, adjective and verb classes. Any content word can occupy the predicate position and take appropriate tense, mood and aspect inflections. Consider the examples in (1) which can all occupy the predicate position, as in (2) to (6).

(1) (a) *pʰɔŋ* 'hole'[2] N
(b) *kʰuro* 'big' ADJ
(c) *ut=ɲo* 'live' V
(d) *lɔto* 'much'

[2] The following abbreviations are used. ABS = absolutive, ACC = accusative, Adj.= adjective, Adv = adverb, ARG = argument, CL = class, COP = copula, D = possessed, DIR = directional, FA = formative affix, GEN = genitive, N = noun, NEG = negative, NPST = non past, OBJ = object, PST = past, R = possessor, POSS = possessive, V = verb, 1SG = first person singular, 2SG = second person singular, 3SG = third person singular.

(2) *ʈʰu* *e=pʰoŋ-om*
1SG CL 5=dig-NPST
'I am digging.'
N > V

(3) *ʈʰu* *e=pʰoŋ-e* *kʰuro-k-om*
1SG CL 5=dig-ACC big-FA-NPST
'I am making a big hole.'
Adj > V; V > N

(4) *eremla* *ʈʰ* i *(u)t=ɲyo-ø*
alone here/earth cl 4=live-pst
'(He) lived here all by himself.'
N >V

(5) *cai* *ono* *lɔtɔ-k-o*
why oil much-FA-PST
'Why have you put so much oil?'
Adj/Adv > V

(6) *ʈɔliocɔŋ* *ot=ɲo* *ʈhimikʰu* *bi*
deers CL 4=house jungle COP
'Deer dwelling is forest/ deer live in forest.'

2.2. Absence of logical connectives
PGA does not offer logical connectives, *e.g. 'that', 'despite', 'in spite of', 'or'* to form subordinate clauses. Clauses are juxtaposed and the sense is conveyed by the context of the speech event. Thus:

(7) *a=karkʰa-ø* *tɛkʰamo* *u* *port blair* *et=ok-om*
CL1=tell-PST soon 3SG Port Blair 3SG.OBJ=leave-NPST
'(He) said that he will leave Port Blair soon.'

(8) *jicer-bi* *cer-om* *ʈh-ut=* *cone-pʰo-be*
rain -ABS rain- NPST 1SG- CL4= go-NEG-NPST
'Since it is raining I will not leave.'

2.3 Absence of open temporal adverb class
The major feature of temporal adverbs designating morning and evening is that they are host to a pronominal proclitic co-referring to the subject, a rare phenomenon in the linguistic literature. Thus:

ʈh-ambikihir
1SG-tomorrow
(spoken by the speaker referring to self as a participant of the action)

ŋ-ambikhir
2SG-tomorrow
(spoken by the speaker to the addressee)

ak-ambikhir
3SG-tomorrow
(spoken by the speaker for the person other than the addressee)

(9) *a-sulu* *aka-ambik^{h}ir* *stret-ak* *ot=cone-b-ɔ*
ARG-Sulu 3SG-morning srtait-DIR CL 4=go-FA-PST
'Sulu left for Strait yesterday.'

The word for 'yesterday' and 'today' is the same.

3. BODY DIVISION CLASSES

There is another factor which forces us to rethink the features of UG: there pervasie are grammatical categories that are obligatorily attached to a large variety of nouns, adjectives and verbs. These are BODY DIVISION CLASSES that indicate seven divisions of the human body. They not only classify body part terms but also individuate noun reference and event semantics. These classes have realizations in bound morphemes and are termed BODY CLASS markers in the PGA grammar. Morpho-syntactically they appear as proclitics attached to the left of the host (for details see Abbi 2010). Thus, the class marker for protruding parts of the body, class no. 4, is *ut-* or *ot-* as in [ot=bec]$_{N}$ 'hair'. However, it is obligatorily attached to other form classes as with the verb [ot=cone]$_{V}$ 'to go', [ot=le]$_{deixis}$ 'seaward' and with the modifier [ot=belo]$_{ADJ}$ 'wide'. The semantic congruence between class markers and root lexemes cannot be established without some speculation. Although transparent in some respects, they have become highly grammaticalized. Let us consider them in some detail.

3.1 The seven divisions of the human body

There are seven distinct divisions or areas that are recognized within the human body and each is symbolized by a monosyllabic or disyllabic body class marker serving as a possessive class marker, which is preceded by the appropriate term for the body part, i.e., the head noun. The typical structure of a noun phrase with body part terminology is:

(S 1) R CLASS$_{n}$= D

Here R is a possessor which, in this case, is a pronominal clitic or a noun followed by an appropriate possessive class marker attached to the left of the dependent noun D. Refer to Table 1 for the basic seven divisions of human body in PGA.

Table 1
Seven basic zones in the partonomy of the body

Body division classes	**Partonomy of human body**	**BODY CLASS MARKERS**
1	mouth and its semantic extension	*a-*
2	major external body parts and face-related	*ɛr-*
3	extreme ends of the body like toes and fingernails	*oŋ-*
4	bodily products and part-whole relationship	*ut-*
5	organs inside the body	*e-*
6	parts designating round shape/sexual organs	*ara-*
7.	parts for legs and related terms	*o-* ~ *ɔ-*

Some examples will clarify the structure. A pronominal proclitic attracts the class marker towards itself giving rise to clitic sequencing (for details refer to Abbi 2010).

(10) *ʈh=a=* *pʰoŋ*
1SG=CL1.POSS= cavity
'My mouth cavity'

(11) *ʈh=ɛr=* *co*
1SG=CL 2.POSS= head
'My head'

(12) *nao* *ut=ʈʰi*
Nao CL 4.POSS=breath
'Nao's breath'

(13) *ŋ=e=* *teɖu*
2SG=CL 5.POSS= pancreas
'Your pancreas.'

Of the seven classes, five are used to classify various kinship relations. These are: *a-, ɛr-, ut-, ara-,* and *o-*.

(14) *lico* *ut=t^{h}ire*
Licho CL 4.POSS =child
'Licho's child'

(15) *lico* *ɛr=boi*
Licho CL 2.POSS =spouse
'Licho's husband'

3.2 The inherency and inalienability factors

The conceptualization by the Great Andamanese is anthropocentric. They use human categorization to describe and understand non-human concepts. The human body provides the most important model for expressing concepts not only of spatial orientation (§ 6), but also of relational nouns, attributive categories, inherently related objects of actions and events, or any two objects and two events which are conceptually dependent upon each other. The semantics of 'conceptual dependency' engulfs the concepts of 'inalienability' and 'inherency'. Between the varying degrees of inalienability lie various kinds of inherent relations.

Inherency has a realization in bound morphemes. All content words, i.e., nouns, verbs, adjectives and adverbs have the option of occurring as bound or free. The former occur with body class markers.

3.2.1 Separated body parts and part-to-whole relationship

The separated parts of the body of an animal are obligatorily attached with a body class marker, indicating its inanimate nature. Hence all such terms are preceded by a dental *t-*. Thus:

(16) (a) *ra ɛr=co* 'Pig's head'
(b) *ra* ***tɛr****-co* 'Pig's head' [cut]
(c) *k^{h}eŋe* ***ra****-uli* 'Cat's tail'
(d) *k^{h}eŋe* ***tara-****uli* 'Cat's tail' [cut]

Thus an intact body part belongs to one particular class and a detached one is treated differently but belongs to the same class. This strategy is also applied to describe the part-to-whole concept. Thus:

(17) (a) *bun* ***tɛr=****p^{h}ɪr* 'The sharp edge of a shell'
(b) *fɛc* ***ta=****p^{h}oŋ* 'Mouth of a vessel'
(c) *k^{h}idɛr* ***tɛr=****ʈɔŋ* 'Branch of a coconut'

The fact that separated body parts and parts of an object are expressed similarly indicates the inherent relationship between a whole and its parts. PGA is a rare language where the choice of possessive marking is decided by both the possessor and the possessed. Although we have considered only eight types of body class possessive markers in the paper, there are in fact twelve different varieties of

possessive markers in the language. The large variety of possessive class markers observed in PGA is unusual cross-linguistically.

3.2.2 Other nouns

Nouns other than those described above have the potentiality of being attached by the body class markers suggesting an inherent relationship between the object and its location or its associations.

1. Associatives such as language, words, names, clan, community, e.g. *ɛr=liu* 'his name'
2. Ailments both physical and mental, e.g. *er=eʈɛŋe* 'measles'
3. Spatial terms, e.g. *e=julue* 'in front of'
4. Seascape and landscape terms, e.g. *buruin ter=tekʰ-il* 'in the middle of mountain'

3.2.3 Alienable possession

This is designated by a GENITIVE morpheme *-iʃo ~ -ico* suffixed to the possessor noun/pronominal clitic. Hence *ŋu-iʃo ko* 2SG-GEN bow 'your bow'. This is another unique feature of the language: it is head-marked in inalienable possession but dependent marked in alienable possession constructions.

4. MODIFIERS

The body part semantics individuate attribution of an object (adjective) or of an action (adverb). Inherently relational elements are conceptually dependent and thus also define attributes. The inherent attribute may include inherent personal attributes such as propensity, nature, weight, height, size, shape, state of health, temperature, blood pressure, energy, bodily functions, consciousness, courage, fear, name, and others. Table 2 lists all seven classes and their role in deciding the semantics of the adjectival construction.

Table 2
Classificatory functions of the body division classes in adjectives

Class No.	Body class markers	Semantics	Examples
1	*a-*	mouth-related attribute	*a=mu* 'dumb', *a=tutlup* 'greedy'
2	*ɛr- ~ er-*	external attribute	*ɛr=buŋoi* 'beautiful', *ɛr=achil* 'surprised'
3	*oŋ- ~ on-*	attributes related to limbs	*oŋ=karacay* 'lame', 'handicapped', *on=toplo* 'alone'
4	*ut- ~ ot-*	negative attribute	*ot=lile* 'decay', *ot=lɔk^{h}o*/ nude'
5	*e- ~ i- ~ ɛ-*	inherent attribute	*e=sare* 'salty', *ɛ=bɛn* 'soft'
6	*ara-*	belly-related attribute	*ara=p^{h}eʈk^{h}etɔ* 'big bellied', *ara=kaʈa* 'stout/dwarf'
7	*o- ~ ɔ-*	attribute of shape and texture	*o=baloŋ* 'round', *o=p^{h}elaɲa* 'slippery'

5. VERBS

The body part semantics shift into event type semantic categories of various kinds. The body class markers combine with transitive and intransitive verbal roots. Hence, verbs with body class marker 4, for emission *ot-,* would refer to an action of motion away from the speaker, such as 'go', 'exit'; thus *ʈh=ut=cone-bom* 'I am going', or, where something is being (not necessarily tangibly) generated as in experiential verbs 'feeling sad/happy/hungry/thirsty/' etc. as in *thire ut=ʈheʈe-bom* 'the child is hungry'. These elements of experience, namely 'hunger', 'thirst' etc, are an inherent part of the experience (hence inalienable) and emerge involuntarily in a person. They are seen as products of the body or 'self'. Similarly, verbs like 'shake' and 'kiss' have the class marker 2, i.e., *er-* for major external body parts and face-related concepts, while verbs like 'pound' or 'beat to a pulp' uses class marker 5, i.e., *e-* for internalized objects. Manner of an action is represented by the use of different class markers. Thus *ut=ʃile* 'aim from above', while *e=ʃile* 'aim to pierce'.

6. LOCATIONS

Various body class markers can attach to the same nominal modifying it further to indicate various locations of the object noun. Thus:

(18) (a) *ot=cala* (CL 4=scar) 'scar left by arrow-head'

(b) *er=cala* (CL 2=scar) 'scar on the head'

(c) *oŋ=cala* (CL 3=scar) 'scar on the limbs'

(19) (a) *e=tei* (CL 5=blood) 'blood inside the body'

(b) *ot=tei* (CL 4=blood) 'blood outside the body' [when bleeding]

(c) *oŋ=tei* (CL 3-blood) 'blood on finger or from finger'

6.1 Spatial Deixis

Various spatial deictic references are individuated by body division classes. Refer to Table 3 which is self-explanatory. We intentionally avoid giving sentential examples due to lack of space.

Table 3
Body division classes designating spatial relations

Class No.	**BODY CLASS MARKERS**	**Body division**	**Spatial relations**	**Reference points**
1	*a-*	mouth cavity	surface	'front'
2	*ɛr-*	face	anterior, exterior	'front', 'out'
4	*ut-*	body products	posterior, superior	'up'
5	*e-*	internal parts	interior, centre	'in'
6	*ara-*	sides	periphery	'edge'
7	*ɔ-*	lower parts	inferior	'down'

The basic division is between up/down; in/out, and periphery/centre. Class 3 for extremities does not refer to any deictic concepts. The role of extremities to indicate spatial concepts in other languages have been observed as "virtually insignificant" (Heine 1997).

7. CONCLUSION

The choice of the BODY DIVISION CLASSES in clustering body parts into divisions/areas of the body, and the consideration of each division/area as inalienable and inherent, is a culture-specific phenomenon.

The dependency feature of the verbal root, modifier or noun on the preceding body division class marker may be understood as an 'inherency factor'. The relationship between two nominal categories, an action and its results, an object and its attribute, an action and its mode of operation or resultant state, is seen as inherent and inextricable. Visualizing the world through divisions of the body and conceptualizing various objects, events and actions as interdependent on body-related phenomena appears to be one of the prime and initial stages of language evolution. It is not surprising to find such an anthropocentric conceptualization of the world by an ancient civilization such as that of the Great Andamanese.

REFERENCES

Abbi, Anvita. 2006. Vanishing voices: A typological sketch of Great Andamanese. In Anju Saxena, & Lars Borin, (eds.) *Lesser-known Languages of South Asia. Status and Policies, Case Studies and Applications of Information Technology*. 107–123. Berlin, New York: Mouton de Gruyter.

Abbi, Anvita. 2009. Is Great Andamanese genealogically and typologically distinct from Onge and Jarawa? *Language Sciences* 31, 791–812.

Abbi, Anvita. 2010. Body divisions in Great Andamanese: possessive classification, the semantics of inherency and grammaticalization. *Forthcoming. Studies in Language.*

Baker, Mark. 2001. *The atoms of language.* New York: Basic Books.

Blevins, Juliette. 2007. A long lost sister of Proto-Austronesian? Proto-Ongan mother of Jarawa and Onge of the Andaman Islands. *Oceanic Linguistics* 46: 1, 155–198.

Chomsky, Noam. 2004. "Three factors in language design: background and prospects". Invited address at the 78th annual meeting of the Linguistic Society of America, Boston.

Crain, Stephen. & D. Lillo-Martin. 1999. *An introduction to linguistic theory and language acquisition.* Oxford: Blackwell.

Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioural and Brain Sciences* 32, 429-492.

Evans, Nicholas & Stephen C. Levinson, 2010. Time for sea-change in linguistics: Response to comments on 'The Myth of Language Universals'. *Lingua* 120, 2733-2758.

Hauser, M.D., Noam. Chomsky & W. T. Fitch. 2002. The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569-1579

Heine, Bernd. 1997. *Cognitive foundations of grammar.* Oxford: Oxford University Press.

Jackendoff, Ray. 2002. *Foundations of language: brain, meaning, grammar, evolution.* New York: Oxford University Press.

Kashyap, V.K., T. Sitalaximi, B.N. Sarkar, & T. Trivedi 2003. Molecular relatedness of the aboriginal groups of Andaman and Nicobar Islands with similar ethnic populations. *International Journal of Human Genetics* 3(1), 5-11.

O'Grady, William. 1997. *Syntactic Development.* Chicago: Chicago University Press.

Thangaraj, K. Kumarasamy, Lalji Singh, Alla G. Reddy, V. Raghavendra Rao, Subhash C. Sehgal, Peter A. Underhill, Melanie Pierson, Ian G. Frame & Erika Hagelberg. 2005. Reconstructing the origin of Andaman Islanders. *Science* 308, 996.

Tomasello, Michael. 2004. What kind of evidence could refute the UG hypothesis? Commentary on Wunderlich. *Studies in Language* 28:1, 642-645.

Wunderlich, Dieter. 2004. Is there any evidence that refutes the UG hypothesis? *Studies in Language* 28:1, 646-647.