Lund University, Dept. of Linguistics
Working Papers 45 (1996), 5–24

Deep linguistic prehistory with particular reference to Andamanese

Niclas Burenhult

Introduction

In 1992, American linguist Johanna Nichols introduced a new method of detecting typological patterns at great time depths, based on the morphological analysis and cross-linguistic comparisons of several structural types and grammatical categories (Nichols 1992). She claimed that her method reveals patterns that may go back as far as the initial modern human colonization of the globe, and she set up a preliminary model of early linguistic spread. Has Nichols taken a ground-breaking step towards a greater understanding of our distant linguistic past? And how can we test this?

Towards the end of her book, Nichols 1992:263-65 calls for an analysis of 'critical' languages which are in a unique position to fill the gaps in her study and thus essential to our understanding of global linguistic prehistory. Using Nichols' method as a testing model, this article highlights one such critical language group – the Andamanese language family, spoken by the indigenous Negrito population on the Andaman Islands, in the Bay of Bengal – in an effort to shed further light on the distant linguistic past of our species.

Johanna Nichols: Linguistic diversity in space and time

Nichols' 1992 study involves a statistical survey of four salient morphological patterns labelled structural types (head/dependent marking, morphological complexity, clause alignment and word order) and a number of other phenomena called grammatical categories (inclusive/exclusive opposition, noun classes, numeral classifiers, alienable/inalienable possession, number and valence-affecting operations).

Nichols notes several geographical discrepancies, notably between east and west, or, more specifically, between the Old World and 'colonized areas' (the Pacific and the New World), and she offers a far-reaching

historical interpretation of her findings. On the basis of the east-west discrepancies, she posits an ancient typological split between the linguistic populations of the Old World and those of the Pacific, with the Pacific then functioning as a secondary center of spread and source of circum-Pacific linguistic colonization (Nichols 1992:228-229). She suggests that these patterns could only go back to the initial global spread of our species, starting in Africa some 100,000 years ago (Nichols 1992:258-259, 274-275).

The Andamanese, their islands, and their language

Situated in the Bay of Bengal between Sumatra and Burma, the Andaman Islands is a group of some 200 islands with a total area of 6,340 square kilometres. The main group – North, Middle and South Andaman – are separated only by narrow straits and are collectively known as Great Andaman. Other islands include, for instance, North Sentinel Island and Little Andaman. The islands are mountainous and covered by dense tropical forest. Politically, the Andaman Islands and the neighboring Nicobar Islands constitute a union territory of the Republic of India, but geographically they form part of Insular Southeast Asia.

The indigenous population of the Andamans belongs to the Negrito stock, a group of dark-skinned and short-statured peoples also found in the Malay Peninsula and the Philippines, who are often considered to represent the original population of Southeast Asia. At the time of European contact, the Andamanese were divided into 13 separate tribes of hunter-gatherers. Following European colonization of the Andamans in the late 1700s, the indigenous population dwindled rapidly as a result of infectious diseases, and today only three tribes remain – the Onge, the Jarawa and the Sentinelese – together numbering only a few hundred individuals. The Onge and some Jarawa groups have friendly relations with outsiders, but most Jarawa and the hostile Sentinelese still aggressively oppose any foreign encroachment. (For descriptions of the Andamanese, see e.g. Man 1883, Radcliffe-Brown 1933 and Singh 1975).

The Andamanese language family has been classified into two subgroups, Great and Little Andamanese. The Great Andamanese subgroup, spoken in the Great Andaman archipelago, consists of ten extinct languages: Bea, Bale, Puchikwar, Juwoi, Kol, Bo, Cari, Kede, Jeru and Kora (spelling and classification follow Manoharan 1983, 1986). The Great Andamanese languages became extinct in the 19th and 20th centuries, but a creolized

form of Great Andamanese (labelled Present Great Andamanese by Manoharan 1986:27, 1989, and here referred to as PGA), predominantly based on Jeru, is still spoken by a handful of individuals resettled on Strait Island.

The Little Andamanese subgroup includes three languages: Onge, Jarawa and Sentinelese. Onge is spoken by less than 100 individuals on Little Andaman, and Jarawa speakers (approx. 300) still inhabit the western parts of Great Andaman. Sentinelese is spoken by some 150 individuals on isolated North Sentinel Island (Singh 1975:72). The classification of Sentinelese among the Little Andamanese languages is, however, pure guesswork and has not been established scientifically, as no linguistic studies have been carried out on the island (Manoharan 1983:83).

The relationship between the Great and Little Andamanese subgroups has been the subject of some discussion. Few cognates link the two subgroups together, and Greenberg 1971:810 doubts their genetic relationship. However, it was pointed out already by Radcliffe-Brown 1914:40, 1933:497 and later by Manoharan 1983:86 that great morphological resemblances point to a common genetic origin.

Although the Andamanese language family is generally regarded as an isolate, attempts have been made to classify it into broader groupings. The most serious of these is Greenberg's 1971 Indo-Pacific hypothesis, which links Andamanese to the non-Austronesian languages of Melanesia and to Tasmanian on the basis of 35 cognates and a few grammatical features, but this hypothesis has not gained widespread support.

The earliest account of Andamanese language is that of Colebrooke 1795. A great deal of research was carried out in the late 1800s by British government servants, predominantly by Man, Temple and Portman. Man 1883:49-56, for instance, gives a fair amount of data on Andamanese in general, and Portman 1898 makes a detailed description of five languages on southern Great Andaman. Useful linguistic notes were also made by anthropologists working in the Andamans in the early 1900s, notably Radcliffe-Brown 1933:495-504.

Since 1950, linguistic research in the Andaman Islands has been carried out by linguists from the Anthropological Survey of India (Zide & Pandya 1989:640). Field investigations on Onge have been conducted by Ganguly 1972, Nigam 1969 and Dasgupta & Sharma 1982. A brief description of Jarawa has been collocated by Nair 1979. Attempts at subgrouping and analyzing the Andamanese language family as a whole have been made by

Manoharan 1983, 1986, who has also conducted extensive field work among the few remaining speakers of Present Great Andamanese (Manoharan 1989).

Structural types and grammatical categories

In this section, attempts are made to identify the structural types and grammatical categories present in the Andamanese languages. However, before doing this, it is necessary to make some introductory clarifications.

Andamanese nouns and verbs can be classified as either 'dependent' or 'independent'. This terminology is unfortunate for our purposes, since 'dependent' is used in quite another sense in this discussion, and for the sake of clarity the terms 'inalienable' and 'alienable' will be used here. Inalienable nouns, which usually denote body parts, kins and parts of a whole, cannot appear without a possessive prefix or a cliticized noun. Similarly, inalienable verbs always incorporate a pronoun or noun indicating subject or object relationship. Alienable nouns and verbs, on the other hand, are free forms. Although systems vary, the basic idea appears to be the same in all Andamanese languages. As will be noted, the split pattern results in differences in the morphological marking of syntactic relations. (See Radcliffe-Brown 1933:497-98; Ganguly 1972:3-4; Nair 1979:22-23; and Dasgupta & Sharma 1982:10-12).

Head or dependent marking?

The following sections deal with the morphological marking of syntactic relations in NPs, PPs and clauses. In the examples given, the head of the constituent is boldfaced and markers are italicized.

Morphological marking in NPs. When combined with alienable nouns, nominal possessors in Onge take a genitive suffix, as illustrated by the following example (Dasgupta & Sharma 1982:15):

œŋe-g-*a*	**kwelabo**	'the Onge's cloth'
Onge-ART-GEN	cloth	

In Jeru and some other languages of North Andaman the genitive marker of nominal possessors can also stand by itself (Radcliffe-Brown 1933:503, 504):

Buio *iʧo*	**roa**	'Buio's canoe'
Buio GEN	canoe	

In Onge, pronominal possessors in combination with alienable forms take the same genitive suffixes as the nominal possessors (Dasgupta & Sharma 1982:18). The languages of southern Great Andaman (including Bea, Bale, Puchikwar, Juwoi and Kol), on the other hand, appear to have possessive pronouns without a distinguishable genitive suffix (Portman 1898:131). Pronominal possessor in Onge (Dasgupta & Sharma 1982:18):

m-*a*	**kwelabo**	'my cloth'
1sg-GEN	cloth	

In combination with inalienable forms, possessors take on a different character. Nominal possessors are cliticized to the head noun, and pronominal possessors turn into possessive prefixes, as shown by the following examples from Onge (Ganguly 1972:3-4; Dasgupta & Sharma 1982:10-13):

uemeg-**otiʧu**	'the dog's head'
dog.ART-head	
et-**ejalle**	'our faces'
1pl-face.PL	

In Onge, modifying adjectives are incorporated into the head noun, either directly after it or following an article (Dasgupta & Sharma 1982:38-39). This type of NP may therefore be regarded as head-marked.

kɔoŋ-*ue*-ra	'a big snake'
snake-big-SING	

Information on modifying adjectives in the Great Andamanese languages, including PGA, is scanty, but examples from Bea (Portman 1898:118, 120) may indicate that adjectives stand by themselves and that neither head nor dependent exhibit any form of marking.

Morphological marking in PPs. In all Andamanese languages, direction, location, instrument etc. is expressed by a wide range of suffixes on nouns and pronouns. In the literature, these are generally called postpositions, but presumably they are more correctly referred to as case (or case-like) suffixes (as recognized by Basu 1952:64 and Manoharan 1989:77-80). Case suffixes in Onge (reanalyzed example from Dasgupta & Sharma 1982:20-22, 53-54):

gaiborale-kata	'from the forest'
forest-ABL	

and Bea (reanalyzed example from Radcliffe-Brown 1933:503):

ŋ-ik 'with you'
2sg-SOC

So, instead of being an independent word and head of the constituent, the adposition is suffixed onto its noun. Therefore, the Andamanese languages cannot be said to have a phrase which can be counted as a PP. Nichols 1992:59 refers to these constructions as dependent-marked Ns rather than PPs.

Morphological marking in clauses. It has proved almost impossible to find suitable examples of clauses in the material available, but the following conclusions can be drawn from the data on Onge (Dasgupta & Sharma 1982:40-69): in combination with alienable verbs, nominal subjects and direct objects appear to be unmarked, whereas indirect objects presumably take a case suffix. The following clause (involving an instrument instead of an indirect object) will have to illustrate this pattern (Dasgupta & Sharma 1982:53-54):

œɳal-le-i wagili-*a* ʧendalu **bene** ati
woman-PL-ART iron hoe-INSTR tuber dig ASP
'the women dig tubers with iron hoes'

Pronominal subjects are unmarked, but both direct and indirect objects, if pronouns, receive an object suffix (Dasgupta & Sharma 1982:24, 64):

ŋi m-*a* aŋgibete **belebe**
2sg 1sg-OBJ match-box give.PRET 'you gave me a match-box'

weg-*a* gi m-*a* **uebe**
clay-INSTR 3sg 1sg-OBJ paint.PRET 'she painted me with clay'

If the verb is inalienable, it is prefixed by a nominal or pronominal subject in intransitive clauses, and by a nominal or pronominal object in transitive clauses. Intransitive clauses with inalienable verb (Dasgupta & Sharma 1982:25, 63):

uem-**egatekkebe** 'the dog barked'
dog-bark.PRET

g-**etabetebe** 'he drowned'
3sg-drown.PRET

Transitive clauses with inalienable verb (Dasgupta & Sharma 1982:24, 63):

mi *cɔgeg*-**antitebe** 'I pierced the fish'
1sg fish.ART-pierce.PRET

mi *g*-**alabukea** 'I scold him'
1sg 3sg-scold.PRES

I have found no examples of indirect objects in connection with inalienable verbs. However, phrases designating location, instrument etc. are not cliticized to the verb and take the usual case suffixes, as in the following example (Dasgupta & Sharma 1982:63):

inene-gi *kue*-**ilowabegi** kubarane-g-*a*
foreigner-ART pig-shoot.PRET.IND gun-ART-INSTR
'the foreigner shot the pig there with the gun'

Judging from examples provided by Manoharan 1989:105-8, a similar system exists in PGA. In combination with alienable verbs, nominal and pronominal subjects and nominal direct objects occur independently and are unmarked (it is not clear whether pronominal direct objects take an object suffix, as in Onge). Inalienable verbs must have a pronoun prefix or a noun clitic.

Portman's 1898 and Radcliffe-Brown's 1933:501, 504 data, although meagre, indicate a similar morphosyntactic structure in the extinct Great Andamanese languages, but no definite conclusions should be drawn from this material.

Conclusions. We have to admit that not all of the examples given above live up to the constituent requirements stipulated by Nichols 1992:46-47. We also have to face the fact that data is lacking from the majority of the Andamanese languages, and that the material on Onge is by far the most exhaustive. Indeed, there is a risk that our heavy dependence on Onge may result in a skewed picture of the morphosyntactic structure in Andamanese in general. As noted above, however, we can be fairly sure that the basic pattern is the same in all Andamanese languages, and Onge will be treated below as a typical representative.

Two important conclusions can be drawn. First, the Andamanese languages exhibit a very clear split in head and dependent marking in NPs and clauses. NPs with alienable nouns, and clauses with alienable verbs, are dependent-marked, whereas NPs with inalienable nouns, and clauses with inalienable verbs, are principally head-marked. The head marking pattern is thus restricted to a bound set of nouns and verbs, but, following Nichols' 1992:60-61 criteria for counting markers, such a marking pattern is still

major and salient and should be counted. Second, the Andamanese languages lack true PPs, since they have a set of case-like suffixes instead of adpositions. Therefore they have dependent-marked Ns instead of PPs (Nichols 1992:59).

Counting markers and determining type. The tabulation of head and dependent marking involves a count of the number of 'points' (i.e. affix, clitic or particle slots) which are head-marking (H), dependent-marking (D) or detached (F) for each of the three constituents (NP, PP and S) described above. Constituents with noun dependents and constituents with pronoun dependents are counted separately and totaled for each constituent type (see Nichols 1992:56-62, 97-98).

The Onge noun phrase thus contributes two dependent-marking points (the genitive suffix on noun and pronoun dependents in combination with alienable heads) and three head-marking points (the pronoun prefix and noun clitic on inalienable heads, and the cliticized modifying adjective) to the Onge total. True adpositional phrases do not exist, so the PP constituent contributes no points. The clause presumably contributes three dependent-marking points (suffixes on noun direct objects and pronoun direct and indirect objects) and two head-marking points (the direct objects in the form of pronoun prefix and noun clitic on inalienable verbs). The NP+S total is thus five dependent-marking points and five head-marking points.

In determining the head-dependent 'type' of the language, the proportion of dependent-marking points in NP+S is computed as D/(D+H+F) (Nichols 1992:59-60). According to Nichols' 1992:97-98 specifications, languages can be counted as head-marking if they have a proportion of 0.0–0.3, as double/split-marking if they have a proportion of 0.4–0.6, and as dependent-marking if they have a proportion of 0.7–1.0. In Onge, the proportion of D-points is 0.5 and hence the language falls neatly into place within the double/split-marking section of the scale, as would be expected.

As noted above, the information on the Great Andamanese languages is far too limited for a similar analysis. It could be argued, however, that some Great Andamanese languages would contribute fewer points overall (note the lack of marking in some NPs in Bea, for example), but this would probably not result in any dramatic shifts in the type scale. It will be assumed here that all Andamanese languages belong to the split-marking type.

Morphological complexity

Morphological complexity, in Nichols' terms (Nichols 1992:64-65, 87-88, 98), is calculated simply by adding up the D, H and F points for NP and S. Languages totalling between 1 and 5 points show low complexity, those totalling between 6 and 10 points show moderate complexity, and those totalling between 11 and 15 points show high complexity. Onge totals 10 points and consequently shows moderate, although near-high, complexity. It is possible that some of the Great Andamanese languages show less complexity than Onge (due to the lack of marking in some NPs mentioned above), but they would probably still be moderately complex.

Alignment

Nichols 1992:65-66 lists six alignment categories, based on the morphological distinction or nondistinction of subject of transitive (A), direct object (O) and subject of intransitive (S): neutral, accusative, ergative, three-way, stative-active and hierarchical. In determining the alignment type of a language it is necessary to tabulate an alignment category for each part of speech (Nichols 1992:88-91). Hence we need to identify morphological alignment marking in pronouns, nouns and verbs.

Judging from the examples above, nouns and verbs in Onge clauses with alienable verbs have no inflectional oppositions identifying A, O and S and thus exhibit a neutral alignment pattern. However, pronouns show an accusative pattern, because pronoun O appears to receive a distinct marking, while pronoun A and S are unmarked.

Clauses with inalienable verbs, on the other hand, present a different pattern. Here, S and O (both nouns and pronouns) are incorporated into the verb whereas A is independent and distinct. Clearly, such a pattern could be referred to as ergative-like, but it does not involve inflectional marking and should therefore not be considered here (Nichols 1992:65). Instead, A, O and S should be regarded as unmarked and neutral.

In determining the dominant alignment type of a language, Nichols 1992:92 excludes the neutral pattern unless the language "has absolutely no relevant morphology". The sole non-neutral type in Onge – accusative – is therefore the dominant type. It is not clear whether PGA and the extinct Great Andamanese languages have object inflection in pronouns. Their dominant alignment type is accusative if they do, and neutral if they don't.

Word order

The basic word order is verb-final in all Andamanese languages. Onge (Dasgupta & Sharma 1982:40) and Jarawa (Nair 1979:23-24) have SOV, and so does PGA (Manoharan 1989:105). Judging from Radcliffe-Brown 1933:504 and examples from e.g. Portman 1887:7, 1898:126-31, the same was true for the now extinct Great Andamanese languages.

Inclusive/exclusive distinction

Manoharan 1986:28-29, 1989:67-68 distinguishes an inclusive/exclusive opposition in the first person plural pronoun of PGA. However, no inclusive/exclusive distinction has been observed in any other Andamanese language, including the extinct Great Andamanese languages, on which PGA is based, and one might suggest that the inclusive/exclusive distinction in PGA is a recent phenomenon and not a feature typical of Andamanese in general.

Noun classes

Opinions differ as to the existence of noun classes in Andamanese. In the early years of research, much attention was directed to a set of formative prefixes which were added to the nouns denoting body parts and which could be extended to certain other nouns in the form of "ordinary prefixes" on verbs and adjectives (Portman 1898:34-45, 60, 79-83). Portman listed at least seven such prefixes related to various parts of the human body and concluded that they indicated gender, or, rather, different genera. Radcliffe-Brown 1933:498-501 states that "they give expression to a number of rather indefinite categories." However, Manoharan 1983:30; 1989:61-64, who lists 11 formative prefixes in PGA, is doubtful about these interpretations and states that the formative prefixes change "the meaning of the primary concept into the specific meaning" and therefore do "not organize different words into one group."

It is extremely difficult to get a clear idea of this system of formative prefixes, and the lack of raw data prevents us from taking the analysis any further, but let us outline the system and see if we can draw conclusions about its status as noun categorizer.

Human body parts are subdivided into a number of categories, each category having a distinct prefix. Thus, body parts like head, brain and heart form one category; hand, wrist, knuckle, nail, foot and ankle form a second, etc. (The number of prefix categories ranges from five to eleven or more, but the system is basically the same within the whole language family). The prefixes are obligatory and intimately associated with inalienability and hence the use of pronominal prefixes. The body part prefixes can then be applied to adjectives and possibly also verbs that refer to any noun (not necessarily inalienables), and the choice of prefix category depends on the properties of the noun in question, such as shape or position. The noun itself (if alienable) appears not to receive a prefix.

Evidently, the human body serves as a foundation for a wider classification of nouns, but the exact criteria for placing a noun within a particular category are not clear. It is not even clear whether agreement is obligatory or not. Furthermore, Onge word lists reveal that inalienable verbs display constant formative prefixes. This would mean that verbs do not show agreement with the noun they refer to, but carry their formative prefix for a different reason. Clearly, many questions remain unanswered and more material is necessary if we are to get a more complete picture of the formative prefixes. However, it will be concluded here that the Andamanese languages do display a noun class system, albeit unclear and extremely fluid, but no definite conclusions will be drawn about agreement.

Numeral classifiers

The formative prefixes described above have been likened to numeral classifiers (Manoharan 1986:30). It is true that the formative prefixes, like classifiers, involve shape categories and that classification is fluid, but they are fewer in number (classifiers typically range between 20 and 200) and, more importantly, they do not appear to be associated with numerals (see Nichols 1992:132). Moreover, unlike the formative prefixes, classifiers are always free forms and never form a morphological unit with the noun (Dixon 1986:106). Taken together, these criteria indicate that we are dealing with noun classes and not with numeral classifiers.

Alienable and inalienable possession

The Little Andamanese languages display a very clear alienable/inalienable distinction. In Onge, for instance, kinship terms, names of body parts and words denoting parts of a whole cannot appear without a possessive prefix (Dasgupta & Sharma 1982:10-13; Ganguly 1972:3-4). Judging from the limited data, the same pattern appears to exist in Jarawa (Nair 1979:22-23), and a similar system is evident in PGA (Manoharan 1989:64-65, 78-79).

The information available on the extinct Great Andamanese languages is rather confusing but appears to reveal a similar picture. In his survey of the

languages of southern Great Andaman, Portman 1898:37, 60-69 lists both "simple pronouns" and "abbreviated forms of simple pronouns" and states that the formative prefixes used with names of body parts and words referring to "the human race" in general are "capable of combination with the abbreviated forms of the Pronouns". Although Portman probably did not recognize their significance as markers of inalienable possession, the "abbreviated forms of pronouns" are clearly identical to the possessive prefixes found in Onge, Jarawa and PGA. As to the languages of northern Great Andaman, Radcliffe-Brown 1933:501 states that personal pronouns may be either words or prefixes, but he does not relate the distribution of the pronominal prefixes to words denoting body parts or other inalienables. Still, it seems safe to conclude that all Andamanese languages display an alienable/inalienable distinction, in which nouns denoting body parts, and probably also some kinship terms, cannot be expressed without a possessive prefix.

Number

In determining whether a language exhibits plurality neutralization or not, Nichols 1992:146 limits her survey only to those languages in her sample that have nonzero marking of both dependent and head at the clause level. This is to avoid the problem of distinguishing between underdevelopment of number and more general lack of inflection. Onge clauses with inalienable heads do display both head and dependent marking, and presumably Andamanese is then qualified for inclusion in the survey of plurality neutralization.

Onge has three numbers – singular, dual and plural – marked by suffixes on the noun (Ganguly 1972:4; Dasgupta & Sharma 1982:13-14). These number suffixes are optional. No data is available on Jarawa. As for the Great Andamanese languages, Portman 1887:4 claims that number is absent from nouns, and Radcliffe-Brown 1933:503 makes no mention of number in his account of nominal suffixes in Bea. Moreover, Manoharan 1989:61 states that PGA nouns do not have number inflection, which would seem to confirm the notion that the Great Andamanese languages did not exhibit number distinction in nouns.

Number distinction (singular and plural) exists in the pronominal systems of all Andamanese languages studied. No plural pronouns have been observed in Jarawa, but this is probably due to the difficulties involved in collecting material (Nair 1979:23). In Onge (Ganguly 1972:5; Dasgupta &

Sharma 1982:17), the singular and plural pronouns appear to be expressed by unrelated roots. In the Great Andamanese languages, however, personal plural pronouns sometimes appear to be derived from their singular counterparts (see examples in Portman 1898:60-61; Radcliffe-Brown 1933:501; Manoharan 1989:68).

Some Onge verbs sometimes take a plural suffix, identical to the nominal plural suffix *-le*, when the subject is in the plural. This is illustrated by the following example (Ganguly 1972:6):

ekw-akobele-te-**lle**-be-gi 'they came running'
3pl-run-DIR-PL-PRET-IND

In the Great Andamanese languages, however, verbal number markers appear not to exist, and the same is true for PGA (Manoharan 1989:83).

Conclusions. Although very common, noun number suffixes in Onge are optional. Judging from Nichols' 1992:145, 296-97 treatment of Djingili, such a pattern should be considered an example of plurality neutralization. Similarly, only some Onge verbs take a plural suffix. The distribution of this suffix is unclear, but it should be evident that verbs too display a certain amount of plurality neutralization. Onge pronouns, on the other hand, do not suffer any plurality neutralization, since singular and plural forms appear to be expressed by unrelated roots. In the Great Andamanese languages, plurality neutralization is very apparent in nouns and verbs, which lack number markers altogether, and to some extent in pronouns as well, but this is less clear (for a definition of plurality neutralization in pronouns, see Nichols 1992:151-52). Hence all Andamanese languages exhibit plurality neutralization of some kind, but it is also evident that there are considerable differences between the Great and Little Andamanese subgroups, the former being radically more neutralizing than the latter.

Analysis

Correlations of types and categories

The data presented here does not seem to contradict the correlations made between structural types in Nichols' 1992:97-115 survey. For instance, her claim that double/split marking favors moderate or high complexity and verb-final word order, and that accusative alignment, moderate complexity and SOV order are associated universally, is supported by the data from Onge.

The marking of NPs in the Andamanese languages is consistent with Nichols' claim that inalienable possession is most often head-marked while alienable possession is dependent-marked (Nichols 1992:117). The Andamanese pattern is also consistent with the claim that noun classes are associated with double/split marking and, more generally, with accusative alignment (Nichols 1992:138). However, in this respect Andamanese behaves more like a noun class hotbed language than an outlier, despite the fact that it does not seem to belong to a noun class hotbed (see below).

Incorporating Andamanese into Nichols' model of language spread

Before placing Andamanese in a wider perspective, it is necessary to make some basic assumptions about its origin and prehistory. As was pointed out at the beginning of this article, the Andamanese languages are spoken by Negritos, dark-skinned and short-statured hunter-gatherers who are considered by most experts to be descendants of the aboriginal population, and perhaps the first modern human settlers, of Southeast Asia (Bellwood 1985). The languages show no apparent genetic affinity to other languages of Southeast Asia or, indeed, the rest of the world, and there is no evidence of outside influence in the form of borrowing or precolonial linguistic colonization. In sum, it seems reasonably safe to assume that Andamanese is the sole remaining linguistic representative of pre-Neolithic Southeast Asia, its roots perhaps going back as far as the initial colonization of Southeast Asia by modern humans – an isolate that has remained largely unaffected by the vast linguistic spreads that have occurred elsewhere in Southeast Asia in Neolithic and post-Neolithic times, i.e., the expansion of the Austroasiatic, Austronesian, Sino-Tibetan and Tai stocks. Andamanese might therefore give us a glimpse of what pre-Neolithic Southeast Asia may have looked like.

So, following Nichols 1992, what could we expect a pre-Neolithic language at the southeasternmost extreme of the Old World to be like? Should we expect it to behave like other languages of the Old World? Or, considering the fact that Southeast Asia was a stepping-stone for the settling of the Pacific, will it have more in common with the colonized areas? Let us see how the Andamanese languages behave in relation to other parts of the world (Nichols 1992:184-208).

Structural types

The scale of distribution of head/dependent marking is continental in size, and the most important patterns to be noticed are the high frequency of dependent marking in Africa and Eurasia, the high frequency of double/split marking in Australia, and the high frequency of head marking in the New World. The Andamanese languages, being split marking, thus exhibit the kind of marking that is predominant in Australia.

The distribution of alignment involves a worldwide preference for the accusative, and questions of the distributional scale of accusative are therefore irrelevant. Ergative alignment is a cluster phenomenon, mainly in parts of the Old World and Australia, and stative-active is a macroareal phenomenon that is common in the New World and New Guinea. Since the accusative pattern is the universal favorite, it is not surprising to find it in the Andamanese languages.

The mean complexity level is highest in the Old World, and moderate and high complexity levels are about equally common in this macroarea. In the colonized areas, however, moderate complexity levels are by far the most common. Onge, and presumably Andamanese in general, thus falls within the complexity range that predominates in the colonized areas.

The scale of geographical distribution of word order is smaller than continental, and the universally favored order is SOV. It is the most common order in all three macroareas, and its presence in Andamanese is no surprise.

Grammatical categories

There is a clear global cline in the distribution of inclusive/exclusive opposition and a considerable discrepancy between the Old World, which has low frequencies (21%), and the colonized areas, which have high frequencies (62% in the Pacific and 54% in the New World). Andamanese, which presumably lacks inclusive/exclusive distinction, conforms to the pattern that predominates in the Old World.

Noun classes and numeral classifiers are so-called hotbed phenomena which are smaller than continental in scale. Noun classes are frequent in the Old World and the Pacific (typical hotbeds include Africa, Europe and northern Australia). Numeral classifiers are rare overall but cluster along the Pacific rim. There are no examples of Southeast Asian languages with noun classes in Nichols' sample. Andamanese therefore does not appear to be part of a hotbed but forms a clear outlier. However, Nichols' 1992:131-32 claim that outliers typically give evidence of distant or former connection with hotbeds may be relevant to Andamanese (see below).

Plurality neutralization frequencies form a clear global cline. Frequencies are low in the Old World (18%) and high in colonized areas (79% in the Pacific and 72% in the New World). Andamanese conforms to the pattern that predominates in the colonized areas.

Inalienable possession also forms a global cline. Frequencies are low in the Old World (23%) and high in colonized areas (65% in the Pacific and 51% in the New World). Andamanese conforms to the pattern that predominates in the colonized areas.

The adpositional phrase

Yet another global cline. The Old World is highly consistent in having adpositional phrases (76%), whereas the colonized areas show less tendency to have them (33% in the Pacific, Austronesian languages excluded, and about 50% in the New World). Again, Andamanese conforms to the pattern that predominates in the colonized areas, particularly to that of the Pacific.

Summary

Two of the features above, alignment and word order, display universal preferences. Accusative alignment and SOV word order are the most common in all parts of the world. Consequently, their presence in Andamanese is not unexpected and they are therefore more or less irrelevant to this areal comparison.

In five of the features, which all exhibit considerable discrepancies between the Old World and the colonized areas, the Andamanese languages uniformly correspond with the pattern that predominates in the colonized areas. Only one feature, inclusive/exclusive opposition, connects Andamanese to an Old World pattern.

Noun-classifying Andamanese cannot be connected to a noun class hotbed, geographically or historically, and thus appears to form a clear noun class outlier. However, Andamanese behaves like a typical hotbed language in having double/split marking and accusative alignment (Nichols 1992:138), and one might speculate that it was once part of a now-vanished Southeast Asian noun class hotbed.

Discussion

The patterns summarized above tell us two things. First, the assumptions made at the beginning of this section appear to find support in the data presented. On the whole, Andamanese behaves very differently from its

host area and macroarea. Hence there is no evidence of precolonial linguistic influence or colonization from surrounding Old World areas. The isolated status of Andamanese is thereby further attested.

Second, Andamanese shows clear connections to the colonized areas. In three of the four features that display a global cline, as well as in continental features like head/dependent marking, Andamanese behaves like a typical 'eastern' or 'colony' language. The only exception to this rule is the inclusive/exclusive distinction, and the reasons for this discrepancy can only be hinted at. Andamanese languages may never have had an inclusive/exclusive opposition, or they may have lost it at some stage in history (due to internal forces, if we exclude the possibility of Old World influence). Alternatively, the Andamanese languages do have the opposition, but it has not been noticed by researchers. This would explain its presence in PGA (indeed, Nichols 1992:209 claims that inclusive/exclusive opposition is the genetically most stable of the features surveyed). Only future research can shed light on this. So, if we view the interpretation of inclusive/exclusive distinction made above as uncertain and disregard it, all relevant grammatical features point to a striking uniformity between Andamanese and the colonized areas.

What bearing does this uniformity have on Nichols' view on early linguistic spread? If we posit that the Andamanese languages are typical representatives of the languages spoken in Southeast Asia at the time of the human crossing to Australia, New Guinea and adjacent insular areas, between 50,000 and 60,000 years ago, it could be argued that the ancient typological split Nichols 1992:228, 275 suggests to have occurred when Old World people set out to colonize new worlds actually took place long before that. In other words, the first major typological split took place somewhere west of Southeast Asia and not at the departure point of the first colonizers as indicated by Nichols. The colony patterns visible today would thus originate in the Old World, not in the Pacific. This view presupposes that Southeast Asia was typologically relatively homogeneous at the time of Pacific colonization, perhaps due to a random founder effect or an early intra-Old World stabilization of frequencies (for a discussion on stabilization of frequencies, see Nichols 1992:213-15). Subsequent Old World interaction and language succession would explain the near total obliteration of these extremely early Southeast Asian patterns.

This interpretation may have far-reaching implications on Nichols' macroareal division, because it suggests that parts of the Old World once

behaved like the colonized areas and that the human leap to Australia and the Pacific was not a typologically very dramatic one. Her division of the world into an Old World homeland vs. colonized areas would thereby appear artificial. The scenario presented here is perhaps more in line with prehistoric reality, since the 'Eve' theory of modern human origin, to which most prehistorians adhere, treats sub-Saharan Africa as the true homeland of our species and all other parts of the world as colonized areas.

However, there is a possible alternative explanation to the similarities between Andamanese and the colonized areas. If Nichols is correct in suggesting a Pacific origin of the colony patterns, one might speculate that Southeast Asia has been subject to secondary, pre-Neolithic influence from the Pacific area, Andamanese then being a witness to extensive east-to-west impulses. Clearly, such an interpretation would be more in line with Nichols' view of the Pacific as a secondary center of spread and source of circum-Pacific colonization.

Conclusion

Incorporating a single language or language group into a model based on statistical data from a huge number of languages is a venturesome undertaking. The strength of Nichols' survey lies in its sheer size, and any attempt to verify or falsify her model (or parts of it) on the basis of data from a single language can easily be brushed aside as statistically uninteresting. Similarities and dissimilarities to patterns in the model may simply be due to chance. Still, the isolated Andamanese languages, spoken by people who are believed to be descendants of the aboriginal population of Southeast Asia, can by all means be regarded as 'critical' and may, if added to Nichols' model, have a significant bearing upon her interpretation of linguistic prehistory.

As noted above, Andamanese exhibits striking correspondences to typical 'colony' patterns, that is, to typological patterns that predominate in the Pacific and the New World, and two different scenarios explaining these similarities were presented. Clearly, the first scenario is the one that is most compatible with the archaeological record. It is a well-established fact that Australia-New Guinea was populated from Southeast Asia (White & O'Connell 1982:42-46), but there is no prehistoric evidence of secondary impulses going in the opposite direction. However, it would not be unreasonable to assume that such impulses have taken place at some stage in prehistory. It will therefore not be concluded here which scenario is the

most likely one, only that the first scenario is more compatible with the archaeological record, and that the second is more in line with Nichols' linguistic model.

One thing is certain, whichever view we take on the directionality of pre-Neolithic linguistic spread or influence in Australasia: the Andamanese languages, little known and largely neglected, do have an interesting story to tell. And, as Nichols 1992:230 points out, the most striking aspect is the fact that modern linguistic evidence can have anything at all to say about early human prehistory.

References

Basu, D.N. 1952. 'A linguistic introduction to Andamanese'. *Bulletin of the Department of Anthropology* 1:2, 55-70. Anthropological Survey of India.

Bellwood, Peter S. 1985. *Prehistory of the Indo-Malaysian archipelago*. Sydney: Academic Press.

Colebrooke, R. 1795. 'On the Andaman Islands'. *Asiatick Researches* 4, 385-94 [reprinted by Cosmo Publications, New Delhi, 1979.]

Dasgupta, D. & S.R. Sharma. 1982. *A hand book of Onge language*. Calcutta: Anthropological Survey of India.

Dixon, R.M.W. 1986. 'Noun classes and noun classification in typological perspective'. In Colette Craig (ed.), *Noun classes and categorization*, 105-12. Amsterdam: John Benjamins.

Ganguly, Pranab. 1972 (periodical dated 1966). 'Vocabulary of the Negritos of Little Andaman with grammatical notes and materials'. *Bulletin of the Anthropological Survey of India* 15:1-4, 1-30.

Greenberg, Joseph H. 1971. 'The Indo-Pacific Hypothesis'. In Thomas A. Sebeok (ed.), *Current trends in linguistics: Linguistics in Oceania* 8:1, 807-71. The Hague: Mouton.

Man, Edward Horace. 1883. *Aboriginal inhabitants of the Andaman Islands*, reprinted 1975. Delhi: Sanskaran Prakashak.

Manoharan, S. 1983. 'Subgrouping Andamanese group of languages'. *International Journal of Dravidian Linguistics* 12:1, 82-95.

Manoharan, S. 1986. 'Linguistic peculiarities of Andamanese family of languages'. *Indian Linguistics* 47:1-4, 25-32. Pune: The Linguistic Society of India.

Manoharan, S. 1989. *A descriptive and comparative study of Andamanese language*. Calcutta: Anthropological Survey of India.

Nair, V.S. 1979. 'A note on the language of Jarawas'. *Bulletin of the Anthropological Survey of India* 28:3-4, 17-38.
Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.
Nigam, R.C. 1969. 'Report on field investigations on 'Onge speech' (1955-56)'. *Bulletin of the Anthropological Survey of India* 12:3-4, 127-45.
Portman, M.V. 1887. *A manual of the Andamanese languages*. London: W.H. Allen & Co.
Portman, M.V. 1898. *Notes on the languages of the South Andaman group of tribes*. Calcutta: Office of the Superintendant of Government Printing.
Radcliffe-Brown, A.R. 1914. 'Notes on the languages of the Andaman Islands'. *Anthropos* 9, 36-52.
Radcliffe-Brown, A.R. 1933. *The Andaman islanders* (reprint). Cambridge: Cambridge University Press.
Singh, Raghubir. 1975. 'The last Andaman islanders'. *National Geographic* 148:1, 66-91.
White, J. Peter & J.F. O'Connell. 1982. *A prehistory of Australia, New Guinea and Sahul*. Sydney: Academic Press.
Zide, Norman & V. Pandya. 1989. 'A bibliographic introduction to Andamanese linguistics'. *Journal of the American Oriental Society* 109, 639-51.

Lund University, Dept. of Linguistics
Working Papers 45 (1996), 25–44

Noun incorporation in Hopi*

Claire Gronemeyer

Introduction

This paper examines Noun Incorporation (NI) in Hopi and presents evidence that Hopi has Type IV noun incorporation according to the typological classification presented in Mithun 1984. The incorporated noun (IN) is visible to discourse reference, NI can strand modifiers, and the IN can have a classificatory function. Hopi thus fulfills the criteria for a syntactic analysis as proposed in Baker 1988, 1995. However, the claim that word formation actually occurs in the syntax has been controversial and is challenged by a number of lexicalists (e.g. Rosen, Mithun). The arguments crucially weigh on the Lexicalist Hypothesis and the division of labour between the morphology and the syntax. Both the syntactic and the lexical approaches seem to account for the basic facts of NI, but significant differences arise on closer examination. The goal of this paper is to contribute some previously unknown data to the current discussion of this rather unusual morphological process and to consider possible analyses of the data. Furthermore, a brief overview of the polysynthetic properties of Hopi is presented to see whether these tendencies may account for NI in Hopi. Except when indicated, all the examples used in this paper are taken directly from *The Hopi dictionary* (Hill et al. in press).

General properties of Hopi

Hopi is a Uto-Aztecan language spoken in Northern Arizona. Word order is predominantly SOV, although some scrambling may occur for discourse reasons. The Hopi case system includes the unmarked or nominative case for subjects and the marked or accusative case in all other positions; the oblique cases are marked by postpositions. The grammar distinguishes

*This article developed out of a course paper written during the LSA Linguistic Institute 1995. I wish to thank Mark Baker, Ken Hale, and LaVerne Jeanne for their comments on and discussions of this topic during those courses. My gratitude also extends to the IDO fund which supported my participation in the Linguistic Institute and to the research project Parametric Typology (funded by Riksbankens Jubileumsfond) which has provided the financial support necessary to continue this research.