

NOMINAL COMBINING FORMS IN SORA AND GORUM

Arlene R. K. Zide

Munda Languages Project, University of Chicago

1. *Introduction*

Sora and Gorum¹ are sister languages spoken in various parts of southern Orissa and Andhra Pradesh, India. Together with Juray they form a subbranch of the Koraput Munda branch of the South Munda family.

As in Gta?, another Koraput Munda language,² we find in Sora and Gorum an old set of syntactic-semantic structures, presumably reconstructable for South Munda, which use, in a variety of constructions, combining forms (CFs) or compositional "short forms" of longer full form (FF) nouns. Truncated or otherwise shortened bound morphemes, these CFs are derived from underlying FF nominals. They may be attached to verb roots as their object forming complex verb stems which can be inflected for tense, aspect, and so on. Also found are complex nominal forms consisting of a series of CFs or created by addition of a CF to a FF or by addition of one or more CFs to a verb root. In Sora, which preserves these phenomena in a relatively active state, we find a considerable variety of such derivationally related nominals. In Gorum, which has not preserved this derivational process, we nonetheless find forms which can be shown to have been derived

a similar fashion at some earlier period. The following examples illustrate these details:

verbal derivatives

. + CF:

Sora *doŋ-gəl* 'to shoot forth' (of an ear of corn), < FF *gəle* 'ear of corn' > CF *-gəl*.

Gorum *gal-baʔa* 'to tie a turban' (literally 'tie-head'), < FF *bab* 'head' > CF *-baʔa*. Cf. Sora FF *bɔʔɔb* > CF *-bəb*, probably < Sora-Gorum **bɔʔɔb*.

nominal derivatives

. + CF + CF:

Gorum *kid-suŋ-kui* 'widow' (literally 'die-house-woman'), < FF (*aHsuŋ*) 'house' > CF *-suŋ* and FF *ku'j=ol* 'woman' > CF *-kui*.

' + CF:

Sora *əɾiɓ-da* 'buttermilk' (literally 'milk-water'), < FF *əɾiɓ* 'milk' and FF *daʔa* 'water' > CF *-da*.

' + CF + CF:

Sora *asɔŋ-bəb-mər* 'insolent man' (literally 'feces-head-man', one who defecates on one's head), < FF *bɔʔɔb* 'head' > CF *-bəb* and FF *mənra* 'man' > CF *-mər*.

The present paper is concerned with the phonological problems involved in deriving CFs from FF nominals and seeks to restate possible derivational rules in Sora and Gorum. In addition, it will be suggested that CFs in Sora and Sora-Gorum are derived by rules differing little from those operative in Gtaʔ.

FF and CF Data

Monosyllabic CFs of FF nominals compounded with verb roots are exemplified by the following sentence:³

kərib-ñen mal-jom-jel-mər-te

2 1 3 5 4

"My sword longs to eat human flesh."

1 2 3 4 5

Here -mər is the CF of mənra 'man' while -jel is the CF of jelu 'meat, flesh'. In the following sentence -bəŋ, the CF of bəŋtel 'buffalo', is used in the same way:

gəd-bəŋ-te-ji

3 4 2 1

"They are cutting (= slaughtering) the buffalo."

1 2 3 2 4

CF nominals may also be attached to other nominals to form compounds of a descriptive or particularizing type such as sidtri-kul 'gruel made from *Eleusine coracana*' (< kuɾu 'gruel, porridge' > CF -kul) or verbal-nominal combinations such as jom-mər-kid 'man-eater' (literally 'eat-man-tiger') (< kina 'tiger' > CF -kid). A representative list, phonetically accurate for the Sora dialect of Tabono Savra of Serango, of FFs and their corresponding CFs is given in the Appendix.

3. Previous Treatments⁴

No exhaustive statement of CF derivational rules has been made other than the fairly complete listing of forms by Ramamurti (1931) and an attempt at derivational treatment by Biligiri (1965).

Biligiri's exposition accounts for most CFs in the language

his approach is narrowly descriptive, with its main emphasis on mechanical "classes" rather than on rules which might account for CFs and their relationship both to their parent FFs and to other derivational rules in the language. The six descriptive classes he has set up appear to have no basis in the morphology or phonology either of synchronic Sora or of diachronic, internally reconstructed pre-Sora. Somewhat arbitrarily, these classes are based on the behavior of final consonants and final vowels in FFs and are held to show "similar contractions." Thus Class I derives its CFs by dropping the first vowel and following consonant of the FF, and includes such divergent forms as *jiʔi* 'tooth' > CF *-ji*, *ləbo* 'earth' > CF *-lo*, and *rəneŋ* 'wing' > CF *-reŋ*. Class II CFs are derived by dropping all but the initial (C)VC of the FF, as in *bəŋtəl* 'buffalo' > CF *-bŋ*, *dango* 'stick' > CF *-daŋ*, and *ɔla* 'leaf' > CF *-ɔl*. CFs of Class III are derived by dropping all except the final (C)VC or (C) of the FF: *omod* 'smoke' > CF *-mod*, *əntid* 'bird' > CF *-tid*, *su* 'fever' > CF *-su*, but also *məmə* 'breast' > CF *-mə* and *ɔab* 'sago leaf' > CF *-ab*. In Class IV all but the first CV and last vowel are dropped: *kondəm* 'grain' > CF *-kom*, *kɪnam* 'husk' > CF *-kɪm*, *lur* 'blacksmith' > CF *-lur*, *gunlu* 'snail' > CF *-gul*. Members of Class V are said to derive their CFs in two steps. First they are truncated in the manner distinctive of Class I; they then undergo further truncation of the Class II type: *mərɪdsa* 'chilli pepper' > **mɪdsa* > CF *mid*, *sənuru* 'rain-hat' > **suru* > CF *-sur*. Class VI is made up of "irregular" (presumably suppletive) forms:

bora 'jackal' > CF -bub, tɔʔɔd 'mouth' > CF -tam, kətəɾa 'castor oil' > CF -tel.

It is worthwhile to consider Biligiri's six classes from the point of view of whatever explanatory solution they provide. In his own Class V, requiring two steps identical with those attributed to two other classes, there is certainly a hint that the derivation of CFs from FFs may not be as arbitrary as the classes themselves would imply. It is entirely true, in a mechanical sort of way, that $jiʔi > -ji$, $lebo > -lɔ$ and $rɛnɛŋ > -rɛŋ$ (all of Class I) follow one and the same type of change. But to construe this as a "rule" ignores the similarity of the change to that undergone by $məridsa$ (Class V) and identical items such as $omod > -mod$ and $ɔntid > -tid$ (Class III) as well as the similarity of $məridsa$ and $sənuru$ (both of Class V) with $gunlu$ (Class IV). One can indeed account for all types of contraction in Sora simply by recognizing a sufficient number of classes, in this case six. But, those classes are unsatisfactory in terms of linguistic motivation, or explanation, for what are otherwise merely arbitrary classifications. In terms of linguistic motivation, however, such classes are unsatisfactory, for they are arbitrary groupings which fail to show the rationale of the processes in question. In reality, CFs and their derivational relationships are not as unrelated to other areas of Sora morphology as they appear to be from Biligiri's *treatment*.

The handling of *Gta?* contraction by Mahapatra and Zide (1972), while offering an imperfect solution because of the nature of the materials, is more linguistically satisfying even though it fails to fit all details into neat categories or to account for all contractions with equal ease. It is more satisfying because it is based on morphological patterns exemplified and supported elsewhere in the grammar. It is therefore externally as well as internally motivated and provides an integrated picture of the derivational history of such forms.

. *Derivation of Sora CFs*

It has already been suggested that certain of the morphological properties of FF nominals must be assessed before prediction of derived CFs can be undertaken with any hope of accuracy. If one works exclusively with synchronic Sora data the required information is in many instances not retrievable from surface forms. At the same time, information can be retrieved in some cases thanks to the existence in the language of a related form or forms, such as a verb, from which the FF nominal is itself derived. Thus *gətasi* 'play' > *-gəsi* may be compared with the verb *gəsi-* 'to play', *kənoŋ* 'razor' > *-koŋ* may be compared with *koŋ-* 'to shave', and *gən(d)ɔj* 'shame' > *-gəɔj* may be compared with *gəɔj-* 'to redden, blush, be ashamed'. Elsewhere in the language is found ample evidence of the active derivation of various types of nominals from verbs by infixation of *-(V)ɾ-* and *(V)n-*. For example,

tɪy 'to give', > t-ən-ɪy 'that which is given; giving'
and t-ər-ɪy-tɪy 'vessel (= place in which a thing is
given)'.
120

koŋ 'to shave', > k-ən-oŋ 'razor' (> CF -koŋ and
k-ər-oŋ-koŋ 'shaving', as in kəroŋkoŋ-bəb-sɪŋ 'barber-
shop' (literally 'shaving-head-house').

gɔb 'to sit', > g-ər-ɔb 'seat' (> CF -gɔb).

bɔj 'to sew', > b-ər-ɔj-bɔj 'needle and thread', b-ən-
ɔj-bɔj 'sewing', and am-b-ən-ɔj 'embroidery (= end-
product of the act of sewing)', from underlying
*Vn-b-Vn-ɔj.

ga 'to eat', > g-ər-a-ga 'plate' and g-ən-a-ga 'food'.

əb-ga 'to feed', > ər-əb-ga 'spoon, plate, feeding-
chair' and ən-əb-ga 'food being fed (to an infant)'.

Other types of bisyllabic and polysyllabic FFs are found
which are not necessarily derived from verbs or which are no
longer recognized as having been derived from earlier verbal
forms. These include nominals with prefixed bound morphemes which
may originally have been nothing more than the very common nom-
inalizing prefix (V)n-, such as anur 'scent' > CF -nur, əñum
'urine' > CF -ñum, areneŋ 'wings' > CF -reŋ, əsaj 'charcoal' >
CF -saj, əsoŋ 'excreta' > CF -soŋ, əsu 'illness' > CF -su, əndɪ
'play, sport' > CF -did, əyɔ 'fish' > CF -yɔ, abay 'seed' > CF
-bay, and ajed 'root' > CF -jed.

Also found are bisyllabic and polysyllabic stems consisting
of conjoined monosyllabic morphemes; these include such items as
bɔŋtel 'buffalo' > CF -bɔŋ, benta 'hunting' > CF -ben, bɔŋseŋ
'trough' > CF -bɔŋ, daŋkəra 'basket of ten measures' > CF -daŋ,
and buroy 'millet' > CF -bur. The morpheme -bur is also seen in

e FF kəmbur, both buroy and kəmbur designating varieties of
llet. The CF -eŋ is similarly found in eŋlud and eŋra, both
aning 'cucumber'.

From this much it can be seen that the first step in
counting for the derivation of CFs must be the identification
FFs which are themselves derived from monosyllabic morphemic
minimal stems. For nominals derived from verbs, reduction is
effected by deinfixation, *i.e.*, removal of -(V)n- or -(V)r- in
own instances. For nominals formed by concatenation of sev-
eral potentially independent morphemes, truncation is effected
deleting the second syllable in most cases. Cases do occur,
however, in which the first syllable must be deleted; this is
a problem involving accentual patterns, to which we shall return
later in this paper. For nominals with prefixed ə/a or ən/an,
it is presumed that deletion of prefix *V(n) yields the CF.

In addition to morphological affixation we find in Sora, as
Gta?, a generalization of truncation and deaffixation rules
include "phonological" as well as morphological infixation,
so that CVCVC and even CVLLV forms are subject to deaffixation
rules. In Gta? bVle? 'feather' > -be? exemplifies what Mahapatra
and Zide (1972:2) have termed a "phonological" infix. The rule
for truncation by deinfixation of verbally derived nominals is
extended to include similarly shaped syllables in other kinds of
nominals. Sora has the counterpart of the Gta? example just
given in such cases as bæled 'feather' > -bed, bənīm 'anthill'
> -bīm, deraŋ 'house lizard' > -daŋ, gerij 'tile, potsherd' >

-gɨj, and even kəɾib 'sword' > -kib, where -ər- and -ən- are extended to include other "phonological infixes" of the vowel-plus-continuant shape. The removal of the phonological "infix" yields a CF from, roughly, the C₁V₂C₃ of the FF.

Two other types of truncation by deaffixation of real affixes occur. One of these is the reduction of reduplicated forms to their root, as in the case of saŋsaŋ 'tumeric' > -saŋ, kaka 'crow' -ka, and məmə 'breast' > -mə; similar to the deletion of the prefixes ə/a-, ən/an- is the deletion of prefixes such as kən/kɨn/kin- especially common, with animal names and presumably reflecting *kVn- 'large, primary, generic', and also the deletion of what is probably an old infix *V? which, like VL,⁵ becomes generalized to include "phonological" infixes under the same rule. Thus, ɔʔɔn 'child' > -ɔn, uʔu 'hair' > -u, raʔa 'elephant' > -ra, siʔi 'hand' > -si, taʔal 'spleen' > -tal, and aneb ~ neʔeb 'tree' > -neb.⁶

There are polysyllabic forms requiring application of more than one rule: first, a rule of morphological deaffixation and, secondly, a rule of phonological deinfixation and/or truncation. For example, the CF of əndaraj 'brinjal, eggplant' > -daj is derivable either as əndaraj > *daraj > -daj or alternatively as əndaraj > *əndaj > -daj; arenəŋ 'wings' > -rəŋ is derivable either as arenəŋ > *renəŋ > -rəŋ or as arenəŋ > *arəŋ > -rəŋ. Similarly, two rules or steps are required for the derivation of such polysyllables as kəɾob'pi 'earring' > -kob. Deinfixation--phonological in this instance--would first yield

ob'pi, which would then become subject to the rule governing the deletion of the second morpheme-syllable, yielding the appropriate CF -kob.

By these means a series of ordered rules is discovered which account for the derivation of most CFs in the language. Thus:

A. *Deaffixation:*

a. prefixes $\tilde{a}/\tilde{a}/\tilde{e}n/\tilde{a}n/k\tilde{V}n$ -----> \emptyset / \lfloor CF

b. infixes $\tilde{V}L/(\tilde{V})?$ -----> \emptyset / \lfloor CF

e., real and "phonological" infixes and by extension $\tilde{V}C[-obs]$ -----> \emptyset / \lfloor CF in order to include such forms as basid / sid 'salt'.⁷

Generalization of rule *b* to include any phonological "infix" consisting of an unstressed vowel plus a nonobstruent consonant in FFs of the shape CVCVC seems to reflect the situation in present-day Sora, at least for all forms except those in which preserved verbal relations serve as effective interference.

c. reduplication⁸ R -----> \emptyset / $- [(C)V(C)]_{\text{syll}}$

B. *[CVC] Preference Rules:*

a. loss of an unstressed syllable in the environment of stressed syllable: $[(C)V(C)]_{\text{syll}}$ ----> \emptyset / $([C\acute{V}C]_{\text{syll}})^9$

b. $\left\{ \begin{array}{l} [(C)\acute{V}]_{\text{syll}} \\ [V(C)]_{\text{syll}} \end{array} \right\}$ -----> \emptyset / $([C\acute{V}C]_{\text{syll}})$

Thus tárba 'flower' > -tar, lakíj 'sand' > -kij, pìsán 'bitter' > -saŋ, rəgál 'sweet' > -gal, gəmol 'insect' > -mol èlén 'salty' > -leŋ, gúrdá 'pus' > -gur, and so on.

There seems to be a general preference or "conspiracy" for CFs of the permissible CVC shape, a first-syllable CVC shape so that we end up with a series of little subrules all seemingly designed to produce CFs of a form which is [CVC]_{syll}. Under B we require, for example, a subrule to the effect that, given a FF of shape CVLV where V₁ is not unstressed, the preferred CF form is CVL:

c. V -----> ∅ / CVL-#¹⁰

Examples are numerous: bárá 'gun barrel' > -bal,¹¹ bòrá 'jackal' > -bor, bóró 'hoe' > -bor, márá 'peacock' > -mar, márá 'loft, shelf' > -mal, múré 'boil' > -mur, kóró 'red-faced monkey' > kor, kórá 'oven' > -kol, búlú 'thigh' > -bul, kúrú 'gruel' > -kul, sèró 'paddy' > -sər, jórá 'stream' > -jəl, gélé 'ear of grain' > -gəl, bèró 'lungs' > -bər, pèrí 'cluster of fruit' > -pəl, and so on.

We also require a CVC-preference subrule to reduce FFs of the shape CVVC to CV₁C for derivation of the CF:

d. V₂ -----> ∅ / #CV₁C#

Thus boab 'wildcat' > -bob, luad 'rope' > -lud, ruaŋ 'sky' > -ruŋ, raŋ 'yoke' > -raŋ.¹²

C. [$\nabla(C)CV'$] FF Preference Rules:

In addition to the foregoing subrules accounting for the C-preference "conspiracy," we require other rules governing s deriving from FFs of the shapes $\nabla C C V'$ and $\nabla C V'$, which are not counted for by the rules of deaffixation. Deaffixation would yield derivations such as əyɔ 'fish' > -yɔ and ɪʔɪ 'louse' > -ɪ, t would not yield olda 'bone marrow' > -ol, əd'rɛ 'egg' > d or uɾa 'mango' > -ul. In such cases, where a CVC shape is possible, the preference is for CFs derived from the first syllable wherever this is permissible. Thus:

$$a. \left\{ \begin{array}{c} V(C) \\ CV \end{array} \right\} \text{ ----- } \rightarrow \emptyset / \#VC\text{-}\#$$

The rules formulated above would leave as exceptions, or require formulation of other rules to account for, forms such as əɾáŋ 'kite' > -əl, əláŋ ~ əláŋ 'thatching straw, jungle grass' > -əl, əréŋ 'stone' > -ər, and əŋəl 'firewood' > -əŋ as opposed to such forms as ədəŋ 'honeycomb' > -dəŋ, áláŋ 'tongue' > -laŋ, and áŋál 'clearing' > -al.

In addition, forms must be accounted for which seem to follow the rules but show unexpected vowel alternations. This problem deserves full discussion and will not be gone into here. Suffice it to note in passing that forms such as bèrú 'hill, mountain' > -bír, mìñám 'blood' > -məñ, mɔʔɔd 'eye' > -məd, ʔɔb 'head' > -bəb, and réŋgé 'wind' > -rɪŋ must be accounted

for in terms of vowel reconstruction rather than of deviation from more general derivational rules.

5. *Summary: Reinterpretation of Data*

The mechanical categorization of all forms into arbitrary classes which fail to take the rest of Sora morphology into account, or thinking solely in terms of what rules are needed to cover every CF derivation, yields results which are not particularly interesting. A far more satisfactory picture emerges if the data are examined from the point of view of the *effect* of rules on their ultimate goal.

It would appear that Sora prefers CFs of a stressed CVC (C here represents a nonvowel) shape.¹³ When this is impossible for one reason or another, (including lack of a viable CVC in the original FF), the first syllable of the FF is the "second choice," as it were, for the CF.

In general terms, the rules can be summarized as:

- A. *Deaffixation*: morphological and, by extension, phonological.
- B. *Preference I*: CVC $\lfloor_{\$1}$ when possible; otherwise, CVC $\lfloor_{\$2}$ or CV₁C₃ (where $\$$ = syllable)
- C. *Preference II*: first syllable is $\overset{\dagger}{CV}$ or ∇C .

The exceptions then prove to be far more interesting than forms which follow all the rules.

Thus while a case such as gorjaŋ 'village' > -jaŋ is totally inexplicable from the data we possess (one unequivocally

pects *-gɔr), forms such as bísíŋ 'headman' > -bíŋ, áŋáí 'clearing' > -al, bèsíḏ 'salt' > -bíḏ, and sàtár 'vein, sinew, nerve' > -tar are explainable by a variety of interconnected reasons; specifically, *-bis, *-ŋal, *-bəs and *-sat all have impermissible consonants either in initial or in final position.

In the case of *-ŋal we have to assume ŋ > Ø by virtue of a general rule prohibiting initial ŋ in surface forms; the identical rule is seen in operation in the case of óŋóí 'darkness, evening' > -ol (i.e., not *-ŋol). Final voiceless consonants such as s and t are not allowable in any context in Sora; so that where there is a CVCVC structure on the model of other structures containing infixes, the next permissible C is taken, by analogy with -CVC₃ structures elsewhere. In the case of sàtár 'vein' the derived CF is not *-sar probably because of contamination or association with the CF -tar < táréí 'thread, string, wire' with its close meaning elsewhere in the language.

This being the case, why do we not get *-əl < éŋéí 'firewood', on the analogy of -al < áŋáí 'clearing', or, for that matter, *-aŋ < èŋáŋ 'kite', *-reŋ < əréŋ 'stone', and -laŋ < èláŋ 'thatch'? We find rVŋ and lVŋ elsewhere in Sora in permissible sequences: əréŋ 'sour' > -reŋ, áláŋ 'tongue' > -laŋ, əláŋ 'fodder, trampled straw' > -lam, èléŋ 'salty' > -leŋ, and so on. It is clear from these examples that it is not stress¹⁴ that determines the possible CFs, since all possibilities appear to be represented in the permitted sequences. Working synchronically without recourse to comparative data,

we can only set such exceptions aside without explanation.

If, however, we bring in comparative material from Gorum and genetically related languages further afield, some of the apparent exceptions are explainable. Thus:

<i>SORA</i>	<i>GORUM</i>	<i>GRG</i>	<i>KHARIA</i>	<i>ELSEWHERE</i>
əŋəl/-əŋ	aŋal/-suŋ	*sVŋol	soŋol	PM*sVŋaXl
ərəŋ/-ər	ərəŋ/-reŋ, sar=		soreŋ	
əlam/-lam	lam			
əlaŋ/-əl 'thatch'	alaŋ	*ə-loŋ		
alaŋ ~				
laʔaŋ/-laŋ	laʔaŋ ~	*ə-loŋ		
	lag/-laŋ 'tongue'			
əraŋ/-əl 'kite'	uraŋ	*l/naʔlaŋ (but Ga. hnalia)		
kənsim/-im	----	*siŋ		PM *sim
'chicken'				

From this it can be seen that some CFs of Sora-Gorum, or at least of pre-Sora, must have been derived at a time when either an initial s was present (and then later lost) or when the effects of an initial s (secondary stress, possibly) were in force. In the case of əraŋ 'kite' either an initial consonant (in the shape of n) or similarly its residual effect on the stress pattern was probably present. On the other hand, əlaŋ 'thatch' does not appear to be susceptible to this type of explanation, and hence must remain an exception.

The derivation, by the rules proposed here, of various sample forms is shown in the accompanying charts. It can be seen that there are not many cases which cannot be explained by our general rules; the chief exceptions to the latter are forms subject to interference from a related verb, which seems

FF: ád'ré ólá R-bód múj'jà R-pó pèrí ràòŋ bòáb kosáííí benfm gamól

[bcdbod'] [popo]

Rule

A. Deaffixation:

- a. Infix (ád'ré) (ólá) (R-bód)(múj'jà) (R-pó) (pèrí) (ràòŋ) (bòáb) (kosáííí) bim gamól
- b. Prefix " " bod " po " " (kV)sáííí " "
- c. † V-n-LV# " " " " " " " "

B. CVC-Preferences:

- a. CVC " " muj " " " " " " " " " "
- b. CVL/_V " " " " *peŋ " " " " " " " "
- c. 1st syllable CVC with 2 primaries " " " " " " " " " " " "
- d. 1st syllable other ed (e.g. VC||CV, VC][V) ol " " " " " " " " " "
- e. V₂ + Ø / CV_C " " ran bob " " " " " "

C. Readjustment (because of surface constraints)

pel

FF Type: I II III IV V VI

CF	<u>ánál</u> <u>ónól</u>	<u>tánál</u> <u>kómáb</u> <u>mórón</u> <u>túdum</u> <u>tóród</u> <u>kútór</u> <u>tábár</u>	kíná > -kíd rená > -réd	<u>éjúlú</u> <u>gorján</u> <u>regém</u> <u>(Enra)</u>	<u>bísíŋ</u> <u>réméŋ</u> <u>tábáŋ</u>	terá > -tal <u>téríb</u> > -tír <u>berú</u> > -bitr
----	----------------------------	--	----------------------------	--	--	---

Type I. is derived in the usual way but is subject to a surface constraint on initial ŋ, which → ø.

Type II. From tánál we should expect *-tañ, which is in fact the form given by Ramamurti; my -tal may be merely a dialect variant reflecting a rule *ñ → l. I have no explanation for torod > -tod rather than the expected *-təl, the derivation may have something to do with the sequence (C)VrVd, which is neither usual nor found in other FFs. In C₁V₂C₃ forms, C₂ may well be ambisyllabic; the are probably from single bisyllabic morphemes rather than conjoined morphemes; and this may entail a separate rule for retention of elements of both syllables of the single morpheme in the CF. Contrast this with remeŋ Type V, where rVm is attested everywhere (e GRG) as a separate morpheme. bísíŋ > -bíŋ, with two primary stresses, does not belong here.

Type III. With no cognates elsewhere in Munda, rəna > -rəd 'gold' is inexplicable. It may have been formed in a manner analogous kina > -kíd 'tiger', which is widely attested in Munda (cf. Gorum kula > -kul). One may postulate a complex vowel or a vowel plus laryngeal followed by /l/, i.e. *kuXla > -kuXl, which might give rise to such forms via *ku/ɦl > *-kɪ'd in Sora. Examples are elsewhere in the language to confirm this hypothesis. l/d correspondences between Gorum and Sora are common.

Type IV. It is probable that eŋ takes precedence over *ɪud for a lexical reason such as a generic meaning (e.g., 'cucumber'); such instances are found in some Gta? derivations. Similarly, gorjan may be subject to a lexical requirement; the only possible cognate Gorum is jā(ŋ)-kor 'sacred place, village shrine'. Open to speculation is regem > -re; its correspondences invariably include a final ŋ or ñ (Gorum in-ran). However, it may be that g is nonpermitted final, though if this were so one would expect -rēm on the analogo bísíŋ > -bíŋ.

Type V. The CF of bísíŋ is probably derived in the expected manner as *bis, which is an impossible sequence because of the nonpermitted final s. As CV is not a preferred shape for CFs, the only other consonant in the word is taken over: *-bis > *-bi > -bíŋ.

Type VI. The CFs of téríb and beru can be explained by vowel reconstruction rather than derivation. In the case of térá > -tal, the expected *-ta is probably ruled out by the absence of final C, although vowel reconstruction may be relevant here also.

Pre-Sora (SG?) FF: *ku'riŋ
 (Sora FF: urŋ) *se'ŋel eŋel *se'reŋ ereŋ *Nə"raŋ e'raŋ *kVnsim kensim kəŋib [kəŋiŋ] kəŋiŋ

Rules

A. Deaffixation

- a. Infix " "
- b. Prefix " "

B. CVC-Preference

- a. 1st syl. *ku'r *se'o
- b. CVC " "

Later General

Rules for Sora-Gorum (? and pre-Sora)

- 1. k → ø / _VL ur
- 2. s → ø / _ə'L
- 3. n → ø / _ə"L *e'

Late Surface Constraint Rule for Sora

ɾ → | / _# e|

deter further reduction (this is unlike Gta?, where such reduction would be permitted and where the monosyllabic shape of the CF takes precedence over all other considerations).

The subrules or specific rules which make up these more general rules are fairly numerous--perhaps more numerous than Ligiri's six classes. But the explanatory power of our general rules far outweighs their numerical "inelegance." We must cope not only with surface constraints which have definite effects on the output of rules and consequently entail further subrules; we must also deal with what has been termed a "conspiracy" (though this may be too strong a word to use for such similar-looking rules) in the form of different kinds of rules which are not easily collapsible (or even relatable in some cases) but which all seek a single end or goal.

It is inevitable that there will be a small residue of forms which cannot be accounted for as thoroughly as might be wished. In several such cases (e.g., tərʃb 'cloud' > -t̄r, r̄ú 'hill, mountain' > -b̄r), however, we are liable to be involved in problems of phonology or of the reconstruction of levels which are outside the immediate concerns of CF derivation *per se*. The existence of such forms is not something to dismay the comparative and historical linguist, but rather should be welcomed as reflecting an earlier stage of the language awaiting discovery. It is definitely preferable to obscuring such irregularities by mechanical categorization which would set aside such forms as unusual in any way.

Appendix

LIST OF SORA COMBINING FORMS¹⁷

<i>FF</i>	<i>CF</i>	
ədəŋ	-dəŋ	honeycomb
ədí	-dí	cotton
əd'ré	-əd	egg
əŋáŋ	-əl	kite (bird)
əgáb	-gab	grass
əyś	-yc	fish
əléŋ	-əl	thatching straw
əlám	-lam	trampled straw
ələb	-leb	mouse deer, chevrotain
əlí	-sal (<i>suppl.</i>)	liquor
əlś	-lc	river
əndàráj	-daj	brinjal, eggplant
əndíd	-did	sport, play
anéb ~ neʔeb	-neb	tree
ənrśd	-rcd	bee
ənselś	-bcj (<i>suppl.</i>)	woman
əñúm	-ñum	urine
əŋəl	-əŋ	firewood
ərəŋ	-rəŋ	sour
ərəŋ	-ər	stone, rock
ərgal	-gal	thirst

óy	-roy	fly
éd	-red	parrot
j	-saj	charcoal
ón	-son	dung
i	-su	illness, fever
áy	-bay	seed
úr	-gur	ripe fruit
éd	-jed	root
án	-lan	tongue
úr	-nur	scent
ál	-al	clearing on hill
gáj	-gaj	moon
á ~ erá	x	tree, wood
enén	-ren	wings
él	-rel	hail
sí	-ar	monkey
síd	-sid	class of (untouchable) Soras
léd	-bed	feather
nfm	-bim	anthill
ónráj	-ben	flour
eró	-ber	lungs
erú	-bir	hill
esfd	-bid	salt
éssér	-baj (suppl.)	large lizard

FF

CF

bat í	-píd (<i>suppl.</i>)	mushroom
bárá	-bal	gun barrel
bénta	-ben	hunting
bísín	-bíñ	headman
bɔj	-bɔj	woman
bomán	-man ~ -bom	bloodsucker (reptile)
bóñtéí	-bɔñ	buffalo
bòró	-bor	hoe
bɔʔóbb	-bəb	head
boáb	-bob	wildcat
bódbód	-bod	worm
bóyá	-boy	priest
bónsén	-bɔñ	trough
borá	-bor	jackal
búlú	-bul	thigh
búróy	-bur	<i>Setaria italica</i> , millet
déddeṛéy	-dəd	snail
dərán	-dɔñ	house lizard, gecko
dángo	-dɔñ	stick
dáñkəra	-dɔñ ~ -tod	basket of ten measures
dáñkí	-dɔñ	pot
dáréj	-dar	rice
déréñ	-der	horn
díá	-dɪb (<i>suppl.</i>)	waist
dòláj	-dol	hunger

bá	-ed	thorn
éŋ	-leŋ	salty
júm	-jum	axe
l úd	-eŋ	cucumber
rá	-eŋ	cucumber
lé	-gəl	ear of grain
mán	-gəm	rich man
mól	-mol	small insect that flies in face at night
náy	-gay ~ -god	tuber (<i>cf.</i> gay 'to dig')
ndrój	(gəɾɔj)	shame (<i>cf.</i> gəɾɔj 'to put to shame')
núr	-gɪm (<i>suppl.</i>) ~ (-gur <i>Rm.</i>)	rain (<i>cf.</i> gur 'to rain')
r f j	-gɪj	tile, sherd
tasí	-gəsi	play (<i>cf.</i> gəsí 'to play')
ngá	-gaŋ	<i>Panicum miliare</i> , millet
nd f j	-gɔn	squirrel
rján	-jaŋ	village
d'dán	-gud	field
nj f	-gun	a kind of creeper
ntúr	-gun ~ (-tur <i>Rm.</i>)	rat
rdá	-gur	pus
i	-i	louse
bmól	-jəb	seed

FF

CF

jánnón	-jən	field
jəno	-jɔ	broom (<i>cf.</i> jɔ 'to sweep')
játé	-ja	sesame
jeʔén	-jeŋ	leg
jélù	-jɛl	meat
jiʔi	-ji	tooth
jób'ba	-jɔb	mud
jórá	-jɔl	stream
jɔʔɔ	-jɔ	fruit
jumpe	-jum	bunch (<i>Oriya</i>)
kəʔétti	-kəd (<i>suppl.</i>)	sickle
kəʔíb	-kib	sword
kəʔín	-diŋ	drum
kəʔób'pí	-kob	earring
kəmbón ~ kámbon	-jɛl (<i>suppl.</i>) ~ -bon	pig
kəmbúd	-bud	bear
kəmbúr	-buj (<i>suppl.</i>)	<i>Sorghum Roxburghii</i> , millet
kəmpón	-pɔŋ	stomach
kəndód	-dod	frog
kəndón	-dɔŋ	back
kənjín	-jiŋ	porcupine
kənlá	-kəl	small cup
kənón	-kɔŋ	razor

nreŋ	-reŋ	forest
nróm	-rom	large black gram (probably <i>Vigna</i>)
nsím	-im	chicken
ntúj	-tuj	owl
náñjá	-kəñ	tree name (<i>Oriya</i>)
nrí	-kər	fat
nró	-kər	red-faced monkey
nb'para	-kab	duck
nd'du	-kad	bangle
nká	-ka	crow
nméd	-med	goat
nná	-kɪd	tiger
nnád	-kad	crab (< *kɪn + kad)
nnám	-kɪm	husk
ndál	-dal	basket-maker
njær-pón	-poŋ	person with a big belly
nján	-jaŋ	backbone
nláy	-lay	pumpkin, gourd
nrəpéd	-ped	large mouse
nsér	-sər	sambar (deer)
nsód	-sɔd	dog
ntəŋ	-təŋ	brain
ntál	-tal	wall
ntəra ~ kəntəra	-tel	castor oil

FF

CF

kɪnté	-tɛ	banana
kitún	-tuŋ	god, spirit
kórǎ	-kol	oven
kòlbá	-kol	ghost
kòmáb	-kob	ashes
kóndém	-kon	hill paddy (grain)
kóndí	-kon	knife
kórgéd	-kor	blackberry tree
kórpál	-kor	<i>loki</i> , gourd
kórráb	-kod	twigs
kórrán	-kor	tree bark
kosá lí	-sal	cucumber-like vegetable
koʔo	-ko	<i>Lagenaria</i> , gourd, ladle
kúrú	-kul	porridge
kukkúr	-kur	dove
kùllú	-kul	turtle
kúmbɸl	-kum	rat
kúrtá	-kur	horse
kutám	-kum ~ (-təm ?)	<i>mohwa</i> stick
lebó	-lɔ	earth
lakfj	-kɸj	sand
lɔʔɔj	-lɔj	testicles
lɔʔór	-lor	a kind of snake
lónɔr	-lon	cave
luád	-lud	rope

láj	-lan	iron
lájúd	-lud	ear
lándórá	-min	small pot
lónrá	-mər	man
lóríd'sa	-mid	chilli
lórá	-mal	loft shelf
lórǫñ	-mal	large measure
lórýón	-mañ	chest
lórá	-mar	peacock
lórémé	-mε	breast
lórǫñám	-mǫñ	blood
lórǫñfm	-nim	year
lórǫñól	-mε (suppl.)	oil
lórá	-mól	fog
lórǫñ	-mǫñ	big basket
lórǫd	-mǫd	eye
lórǫj'já	-muj	ant
lórǫntá	-mun	small <i>loṭa</i>
lórǫnú	-mun	black gram (<i>Telugu</i>)
lórǫré	-mur	boil
lórǫú	-mu	nose
lórǫnd	-ñod	ghost
lórǫá	-ṅəṅ	bow
lórǫlím	-ñil	leech

FF

CF

òrǎj	-ɔl	<i>Dolichos biflorus</i> , horsegram
òlá	-ɔl	leaf
òlǒj	-lɔj	fruit with poisonous seed (e.g., cashew)
ònrɸj	-rey	pestle
ònréŋ	-rɛŋ	mouse
òntɸb	-tib	fruit (<i>Calatropis alba</i> ?)
ɔʔɔn	-ɔn	son
òáb	-əb	sag, green vegetable, greens
òyóŋ	-yɔŋ	sun
òldá	-ol	bone marrow
omód	-mod	smoke
ònjǎr	-jər	peepul tree
òntíd	-tid	bird
òŋól	-ol	darkness, evening
pǎrǐ	-pəl	cluster of fruit
pənàrá	-para	hurt, wound (cf. para 'to hurt, ache')
pənéd	-pəd	lock, latch (cf. Gorum pad)
pəndér	-pən	rabbit
pəppér	-pər	grasshopper
pəɸj	-sɸj	child
páb'páŋù	-pab	top stem of <i>salop</i> tree where tapped
páŋǐ	-pal	ridges in a field

isán, əsán	-sən	bitter
ogá	-pəd (<i>suppl.</i>)	tobacco
ópó	-pə	cake
əgál	-gəl	sweet
əméŋ	-rəm	cat
əná	-rəd	gold
ənáj	-rəj	stringed instrument
ənán	-rən	cold
əŋgé	-rɪŋ	wind
àón	-rən	yoke
aʔa	-ra	elephant
égəm (-surtəm)	-re	medicine
óŋkó	-rən	raw rice
ùán	-ruŋ	sky
əráj	-sən (<i>suppl.</i>)	red gram
ənəŋ	-sən	door
ənár	-sar	comb
əndí	-sən	cot
ənùrù	-sur	umbrella
əŋká	-sən	neck
ərəmóñ	-sem	kind of long <i>dal</i> , (gram)
əró	-sər	paddy
ərsán	-sər	vegetable
álpəm	-sal	<i>Caryota urens</i> , <i>solop</i>
àmbí	-sam	buttocks

FF

CF

sárdá	-sar	<i>jati</i> (no intermarriage earlier)
sàtár	-tar	vein, nerve
sèndá	-sen	well
sénré	-sen ~ -kab (<i>suppl.</i>)	cloth
seʔeŋ	-seŋ	side
síndí	-sin	date palm
síngér	-siŋ	ginger
sìterí ~ síd'tari ~ sítèri	-sid	<i>Eleusine coracana</i> , <i>ragi</i> , millet
siʔí	-si	hand
sóldá	-sol	earth
sóllá	-sol	small measure
sórà	-sor	Sora
sònfm	-sim	God
sòndán	-son	post
sórbán	-sor	wasp
súndém	-sun	<i>chunam</i> (L.), lime
súttí	-sud	small pot
suʔuŋ	-siŋ	house
tərá	-tal	tank
tənfɓ	-tib	share
tənoŋbá	-təb (<i>suppl.</i>)	husband (<i>cf.</i> <i>tonbá</i> 'to have sexual intercourse')
təngám	-təŋ	beads
təŋkél	-təŋ ~ -kel	shallow basket

əŋóŋr	-god (<i>suppl.</i>)	way, road
ərɸb	-tír	cloud
ábəŋ ~ tábəŋ	-tab	bamboo shoot
ábár	-tar	banyan
áñál	-tal (-tañ <i>Rm.</i>)	crocodile
áŋkó	-taŋ	fruit stone
áŋlɸy	-taŋ	cow, cattle
ár bá	-tar	flower
árél	-tar	thread
á?ál	-tal	spleen
əyǵəl	-tod (<i>suppl.</i>)	large basket of two measures
əmbér	-tem	(fruit eating) jungle rat
érém	-ter	small bee
əttín	-tin	tamarind
ə?əd	-təm (<i>suppl.</i>)	mouth
ə?əǵe	-tod (<i>suppl.</i>)	fire
ə?rá	-tol	<i>mohwa</i> seed
ə?ród	-tod	basket of eight measures
əbléd	-tol	gum
əttóm	-tom	iguana
ədúm	-tum	drum (<i>onom. ?</i>)
əjtúj	-tuj	star
əmbá	-tum	calabash shell (<i>Oriya</i>)

FF

CF

úrǎ	-ul	mango
úrǎŋ	-ur	bamboo
úsál	-sal	skin
u?u	-u	hair

References

- BILIGIRI, H.S., "The Sora Verb, A Restricted Study," in *Lingua*, 15 (1965): 231-50.
- MAHAPATRA, Bijoy, and Ranganayaki MAHAPATRA, *Computer Printout of Sora Vocabulary* (1967).
- MAHAPATRA, Khageshwar, and Norman H. ZIDE, "Gta? Nominal Combining Forms," in *Indian Linguistics*, 33 (1972) 3: 179-202.
- PINNOW, Heinz-Jürgen, "A Comparative Study of the Verb in the Munda Languages," in *Studies in Comparative Austroasiatic Linguistics*, edited by Norman H. Zide (The Hague: Mouton, 1966), 96-195.
- RAMAMURTI, Rao Sahib G.V., "Sora Contracted Forms," excerpted from *A Manual of the Sora (or Savara) Language*. Madras, 1931.
- RAMAMURTI, Rao Sahib G.V., *English-Sora Dictionary*. Madras, 1931.
- RAMAMURTI, Rao Sahib G.V., *Sora-English Dictionary*. Madras, 1938.
- STAMPE, David L., *Proto-Sora-Parengi Phonology*. Unpublished M.A. thesis, Department of Linguistics, University of Chicago, 1963.
- STAROSTA, Stanley, *Sora Syntax*. Ph.D. dissertation, University of Wisconsin, 1967.
- ZIDE, Arlene R.K., *A Comparative Lexicon of Sora-Gorum*. University of Chicago, 1968. Mimeographed.

¹Also known as *Parengi*, *Pareng*, and *Parenga*. Materials on Gorum are taken from the writer's field notes, informants coming from the villages of Semla and Khāśra, Korāput District, Orissa. Sora data are largely taken from the writer's notes, and were elicited mainly from Tabono Savra.

²See Mahapatra and Zide (1972).

³Ramamurti (1931: 154).

⁴The syntactic implications of CFs, while interesting, lie outside the scope of this paper; see Starosta (1967).

⁵The symbol V represents an unspecified, automatic vowel derivable by application of late phonological rules relating to the other vowels of a given form. In Sora it is usually realized as ə but sometimes appears as ɨ or e. The symbol L stands for a continuant, and encompasses n, r, l, ɾ and, rarely, ŋ or ʃ.

⁶If Biligiri's classification is followed, derivation of neʔeb from aneb ~ neʔeb would require the two FFs to be relegated to different classes. Interpreting deaffixation as one type of general rule (with various subrules for the derivation of CFs) would not entail problems of classification in such cases.

⁷The expected form is either *sɨd or perhaps *bəs, in which case the first syllable may have had secondary stress (e.g., in pre-Sora. It is unstressed in present-day Sora.) This suggests a first vowel other than ə for Sora-Gorum. Gorum has bosud 'salt' which seems to bear out such a hypothesis.) If the hypothesis is correct it might explain bɨd as a substitution for *bəs which has an impermissible final s, giving *bəd, which later on was perhaps reinterpreted as derived by de-infixation yielding the form bɨd.

⁸Reduplication in underlying FFs is represented by the symbol R plus the morpheme reduplicated; thus saŋsaŋ 'turmeric' is the surface form of /R-saŋ/. Later phonological rules map the appropriate realizations of such reduplicated forms as tika 'crow', tittin (< *tintin) 'tamarind', pəppər (< *pərpər) 'grasshopper'. Reduplication can be interpreted as operating like prefixation, and its behavior in CF derivation seems to be in no way deviant.

⁹Parentheses around bracketed forms, ([.....]), are used to mean "in the general environment of" or "next to," *i.e.* before or after.

¹⁰The symbol ⁺ over a vowel indicates neutral (*i.e.*, either secondary or primary) stress, as opposed to an unstressed vowel. It is worth noting here that my corpus includes no forms with front V₁.

¹¹These examples exhibit the morphophonemic relationship between intervocalic ɾ and final l. From this and other relationships we can internally reconstruct for Sora (a) an underlying *ɾ, with morphophonemic realizations VɾV and Vl#, and (b) an underlying *d, with morphophonemic realizations VɾV, Vd#, and #dV ~ = dV. These are exemplified above, and in the contrasting CFs of kɔɾa 'oven' > -kol, kɔɖa ~ kɔɾa 'crane' > kod. Cf. also kəɾɪb 'sword' > -kib as opposed to kəɾɪŋ 'drum' > -dɪŋ where the underlying forms are */kəɾɪb/ and */kəɾɪŋ/ respectively.

¹²Perhaps to be included here is oab 'greens' > -əb, although the vowel alternation raises problems we cannot deal with here.

¹³We are presumably dealing with some type of surface constraints here.

¹⁴Stress in the Munda languages appears to be a complex problem. It is nevertheless clear that Sora stress is subphonemic, being determined (as elsewhere in Munda) on the basis of a variety of factors including the inherent unstressedness of such vowels as ə and ɨ in Sora as well as on the basis, in Munda generally, of syllable shape, position in relation to other syllables within the word, and morphology. For discussion of stress in other Munda languages see N. Zide, *Korku Phonology and Morphology*, Ph.D. dissertation, University of Pennsylvania, 1960, and Arlene R.K. Zide, *Some Problems in the Generation of Gorum Obstruents and Related Glides*, University of Chicago, mimeographed 1966. Ramamurti treats Sora stress as distinctive, but often the stress he records seems to be incorrect. Stampe, Starosta, Biligiri and Mahapatra all regard stress as subphonemic in Sora

¹⁵No CFs of the shape -CV are found except as an output of (obligatory) deaffixation and in jate 'sesame' > -ja, where further surface constraints would preclude a CF of any other form.

¹⁶Irregular deinsertion and blocking of further reduction because of interference from existing verbal relations.

¹⁷The forms listed were elicited from Tabono Savra of Serango. The symbols t and d represent the dental and retroflex

ops respectively. As has been said, stress in nonphonemic;
is used here only as suggestive raw data yet to be assessed.
Primary stress is indicated by the acute accent (´) over a
vowel, secondary stress by the grave (`), unmarked vowels being
unstressed. The following abbreviations are used: *onom.* =
onomatopoeia, *Rm.* = Ramamurti, *suppl.* = suppletive, *L.* = loan
word.