

## 2. LEXICOSTATISTICAL SURVEY

### 2.1. Preliminaries

It is generally acknowledged by field workers that lexicostatistics can provide preliminary classifications which can serve as a point of departure for further research (McElhanon 1971). It is further a widely shared experience of field workers in New Guinea that the test list of Swadesh cannot be applied in its entirety. Some of its items are simply not found in New Guinea, others involve the repetition of the same vernacular item, still others may prove to be extremely difficult to elicit. Thus McElhanon found that of Swadesh's 100 item list only 87 were usable in the non-Austronesian languages of the Huon Peninsula and that of Swadesh's 215 list only 150-160 might be usable (1971:127). My own experience is similar to McElhanon's. In an attempt to make the results comparable to those obtained by the use of the full lists, some investigators have adjusted the percentages, others have supplemented the shorter lists with items not found in the Swadesh list but more suited to the local circumstances or with such items as allowed optimal use of the available lexical data<sup>7</sup>.

Subjecting the Asmat data to a lexicostatistical analysis met with an additional difficulty: the fact that the collected word lists were of a widely varying length and composition. I have approached this problem in the following way: first, I devised a test list of 204 items counting 141 items of Swadesh's 215 item list and 63 supplementary items. This list (List I) I used to calculate the cognation percentages shared by KW, KP, KN, CC, and CI, using the fullest lists on hand, viz. Drabbe's lists of Ayam (KW), Namen (KP), Komor (KN), Senggo (CI) and Pirimapun (CC)<sup>8</sup> and my own list of Suru (KW). Since

<sup>7</sup> Adjusting the percentages is based on the experience that shorter versions of the Swadesh list usually give somewhat higher percentages than the full list. Such lists usually contain the more easily obtainable words, mainly nouns and a few basic verbs which perhaps have a higher retention rate than other word classes. In an earlier publication I subtracted one percent for every ten items the list was short of the full (adopted) Swadesh list. However, when the number of supplemented items becomes quite high - as was the case with the lists used here - the assumption does not hold any longer, see below p. 7.

Replacing missing items by others moreover can have a significant effect on the results. McElhanon found that different classifications resulted when he supplemented his Swadesh list with different sets of items and consequently he refrained from supplementing the Swadesh list at all (McElhanon 1971:128). In the present case the paucity of data made it necessary to use whatever non-cultural items were available to compile lists sufficiently long for lexicostatistical use.

<sup>8</sup> Drabbe's lists number 377 items. Still, they do not contain quite a few items of the Swadesh lists which normally would not have to be omitted.

31 items of test list I were not in any of the shorter word lists I set up a second test list (list II) of 193 items. Of these it shared 173 items with list I; 20 items not found in I were added<sup>9</sup>. List II contained 143 items of Swadesh's 215 item list.

However, the actual number of items of list II that could be used for the calculation of cognation percentages varied from one case to another due to the uneven length of the collected word lists and the randomness of the gaps in them. Thus almost every pair of word lists shared a different number and combination of items of list II.

Under such circumstances it is impossible to adjust the percentages. There is no certainty that use of the full test list would in any case have produced a lowering of the percentage according to a determinable scale. In fact, in the few cases in which I was able to compare results yielded by a shorter and longer version of test list II, the cognation percentage remained approximately the same or even rose slightly. A valid question is whether given these facts the results yielded by list II have any value as indicators of genetic relationship. The answer in my opinion is that they indeed have such value provided the objective is no more than to get a broad outline of genetic relationships. The picture is not as black as it seems. Firstly, for all the differences between the collected word lists there still is a considerable overlap between them. Secondly, we are here dealing with fairly high cognation percentages giving less room for wide variation in the results than low percentages would - provided the results are based on lists of approximately the same size<sup>10</sup>.

<sup>9</sup>In order to make optimal use of the available data, cf. note 7.

<sup>10</sup>This point may be illustrated by the chart below which has to be read as follows: 'If a 100-item list would yield a cognation percentage of A percent, then a defective list of B items could be expected on mathematical grounds to give results varying from C to D percent':

A	B	C - D	Variation
100	0 - 100	100	0
90	80	87-100	13
	70	85-100	15
	60	83-100	17
	80	50-75	25
	70	57-86	29
60	60	67-100	33

In the next section I shall present and discuss the cognation percentages yielded by test list I. These percentages are mutually fully comparable, being based on practically identical lists. They will serve as a point of reference for the discussion of the percentages produced by the shortened versions of list II.

## 2.2. Test List I

Chart I: cognation percentages shared by Suru, Ayam (KW), Komor (KN), Namen (KP), Senggo (CI) and Pirimapun (CC).

Chart I

	KW		KN	KP	<u>CI</u>	<u>CC</u>	Number of items counted
	Suru	Ayam	Komor	Namen	Senggo	Pirimapun	
Su		204	202	203	198	201	
Ay	98		202	203	198	201	
Ko	87	86		200	195	199	
Na	89	90	85		196	200	
Se	75	76	76	75		195	
Pi	79	78	74	77	68		

Cognation percentages

The cognation percentages appear to fall into four distinct levels: 1. 98%, 2. 85-90%, 3. 74-79%, and 4. 68%. The minimum difference between 1. and 2., and 2. and 3. is 8%; between 3. and 4. it is 6%. The 68% shared by Senggo and Pirimapun falls below the criterion percentages used by most investigators to distinguish between language and dialect<sup>11</sup>, and well below the 81% adopted by Swadesh. On the basis of this percentage CI and CC can be classified as separate languages. The percentages of level 2 all fall within the range expected of dialects and show that KN, KP, and KW definitely are dialects of one language. I shall call this language Central Asmat (CA). The 98% shared by Suru and Ayam is indicative of the percentages we may expect to be shared by the subdialects of KW.

<sup>11</sup> There is a wide range of opinion concerning which percentage represents the language-dialect boundary. McElhanon (1971) lists nine investigators of New Guinea languages who used between 80% and 86%, twelve who used between 70% and 80% and four using even lower percentages (54%, 60%, 62%, 65%) as the criterion percentage.

The percentages of level 3 are those shared by the three dialects of CA on the one hand and Senggo and Pirimapun on the other. Some investigators have allotted language samples sharing percentages in this range (74%-79%) to the same language (see note 11). I prefer to classify CA, CI, and CC as separate but closely related languages for the following reasons:

firstly, there is a clear gap between the percentages of level 3 and those of level 2;

secondly, CA geographically occupies the middle ground between CC and CI, and one can expect the percentages shared by CA and CI, and by CA and CC to be higher than the one shared by CC and CI.

Summarising: the percentages of chart I allow us to distinguish three separate but closely related languages in the Asmat area, one of which is divided into three dialects.

### 2.3. Test List II

Now we come to the percentages produced by the various shortened versions of test list II. First I shall deal with the cognation percentages shared by some of the Central Asmat lists in the following order: KW lists, KP lists, KN lists, the Sokoni list, the Momogo, Pupis, and Irogo lists. Then the CC lists will be discussed shortly. The only CI list on hand has already been dealt with in the previous section.

#### 2.3.1. The KW Lists

The percentages shared by some of the KW lists - only those counting 70 items or more - have been set out in chart II. They represent the four subdialects postulated by Drabbe. Kainak is represented by the Atat list, Simai by the Ayam list, Mismam by the Suru list, and Mecemup by the lists of Amisu and Omanesep. The percentage shared by Ayam and Suru has been taken from chart I to serve as a reference point for the other percentages.

Chart II

	(Ka) Atat	(S1) Ayam	(Mm) Suru	(Me)	
				Amisu	Omanesep
Atat		127	127	84	95
Ayam	99		<u>204</u>	98	114
Suru	95	<u>98</u>		98	114
A'su	97	100			71
O'sp	96	97	98	98	

Number of items

Cognation percentages

The underlined figures have been taken from chart I.

From the figures in the chart it can be concluded that the cognation percentages within KW lie in the high nineties. I do not expect that filling out the lists would result in significantly lower percentages. The results tally well with the one percentage based on test list I. Further it is clear that the results shown here do not allow a subdivision of KW, although Atat seems to be somewhat closer related to Ayam than to Suru. More data are needed to establish whether lists of the same 'subdialect' indeed share somewhat higher percentages than lists belonging to different 'subdialects'. This is true at least for the Mismam villages of Suru and Yepem whose vocabularies share 100% cognates in the full test lists I and II. (For this reason Yepem has not been entered separately on the chart.) The Amisu-Omanesep figure is not conclusive as it is based on a fairly short list.

### 2.3.2. The KP Lists

Chart III shows the cognation percentages shared by three KP lists. To contrast these with the percentages shared with a list in another dialect I have added those shared with Suru (KW). As before, the underlined figures have been taken from chart I. They show that the short lists probably give somewhat deflated results. Two very short lists, one from Yaosakor and one from Awok have not been included. Given the shortness of the Fos and Miwar lists the figures cannot be conclusive although they tend to confirm that in Namen, Fos, and Miwar the same dialect is spoken. The most significant feature of the chart is the 'jump' between the figures in the first and those in the second column.

Chart III

	KW	KP		
	Suru	Namen	Fos	Miwar
Su		<u>203</u>	51	63
Na	<u>89</u>		47	60
Fo	84	94		27
Mi	86	93	100	

Number of items

Percentages

### 2.3.3. The KN Lists

Of the five lists collected only the MU-Akani list is too short to be usable. The remaining four share between 93%-97% cognates. With Suru of the KW dialect they share between 79%-87% cognates, 6%-16% lower than the previous figures, see chart IV. There are good grounds then to assign the four KN lists, Komor, Sawa, Yipaer, and Manep, to the same dialect.

Note that the percentage shared by Suru and Komor is considerably higher than the other percentages in the first column. It is possible that the figure is somewhat inflated because Drabbe's list appears to have been collected from a man of Ayam (KW) who had been living for many years in Komor<sup>12</sup>. The figure shared by Suru and Yipaer is very low but it may reflect the relative shortness of the Yipaer list. If the 82% shared by Manep and Suru is a reliable indication of the level of relationship between KW and KN, then KW and KN are less closely related than KW and KP (cf. chart III). Manep and Namen (KP) share 82% cognates in a 111 item list. It seems that KW and KP group together as against KN<sup>13</sup>.

<sup>12</sup>At the time Drabbe collected his data he seems not to have been aware of the fact that his Komor informant originally came from Ayam. This came to light much later; I myself learned about it only during my second visit to the Asmat in 1970. It explains the remark made by Drabbe on p. 3 of his dialect study that 'we find in Keenok a fair number of synonyms; of each pair one always seems to belong to the Kawenak dialect' (translation mine). A list collected by myself in Komor shows that in so far as Drabbe's list and my list overlap, Drabbe's list is substantially correct. Still, it is not impossible that some Kawenak words found their way into the remainder of his list.

<sup>13</sup>Supporting evidence is found in the traditions of the coastal Asmat people, see 3.3.2.

Chart IV

	KW	KN			
	Suru	Komor	Sawa	Manep	Yipaer
Su	<div></div>	<u>202</u>	151	114	86
Ko		<u>87</u>	149	110	85
Sa		81	97	109	84
Ma		82	93	94	76
Yi		79	94	94	96

Number of items

Percentages

## 2.3.4. The Sokoni List

One of the villages not classified by Drabbe but included in the Mecemup group by the Asmat Census of Cultural Groups (Van Amelsvoort 1964:193) is Sokoni on the As River. A list collected in this village shares the following percentages with Suru, Ayam, Amisu, and Atat (KW), Namen (KP) and Komor (KN):

Chart V

	KW				KP	KN
	Suru	Ayam	Atat	Amisu	Namen	Komor
So	97	97	66	54	87	86
	84	85	85	85	87	86

No. of items

Percentages

The percentages Sokoni shares with the KW lists are well below the figures one would expect of a subdialect of KW. They are of the same order as those shared with KP and KN and suggest that Sokoni (SO) represents a separate dialect of CA. This is supported by the percentages it shares with Senggo (CI): 74% of 88 items, and with Pirimapun (CC): 73% of 83 items.

## 2.3.5. The Momogo, Pupis, and Irogo Lists

The Asmat Census of Cultural Groups (Van Amelsvoort 1964:193) includes the northern village of Momogo in the Yopmagau group which is the group speaking the Keenok dialect. Linguistically however Momogo does not seem to be intimately related to this group at all. Like the

CI and CC lists the Momogo list seems to represent a separate language which is closely related to Central Asmat. I have labelled it North Asmat (NA). A selection of the percentages shared by Momogo and the CA, CI, and CC lists is given in chart VI.

Chart VI

		<u>CA</u>				
KW		KP	KW	SO	<u>CI</u>	<u>CC</u>
Komor Sawa- Erma Yipaer Manep						
Namen						
Momogo	73	87	94	38	81	84
	68					
	71					
	60					
	69					
	61	75	72	73	74	72
	71					
	75					
Percentages						

No.  
of  
items

The Pupis and Irogo lists are very short but share enough cognates with the Momogo list to suggest that they belong to the same language. The cognation percentages are: Pupis/Momogo 90% (40 items), Irogo/Momogo 93% (32 items), Irogo/Pupis 100% (26 items).

### 2.3.6. The Casuarina Coast Lists

Of the eight lists collected three are too short to be used here. Two of the longer lists, viz. those of the northern villages Masim and Otenep are virtually identical, and for that reason only the Otenep list has been included in chart VII below. The remaining three are those of Pirimapun and Aorket in the centre of the CC area and of Tareo in the south. All except the Pirimapun list are fairly short and the percentages do no more than show that one and the same language is spoken in the four villages. We shall see later in chapter 6 that the CC Asmat distinguish two regional groups and that the lexical data point towards the existence of two dialects coinciding with those groups. Otenep would belong to one of the dialects, the other lists to the other. The difference however is too slight to show up in the present figures.



Chart VII

	Otenep	Pirimapun	Aorket	Tareo
Ot		79	38	40
Pir	99		58	63
Aor	94	96		43
Tar	97	97	100	

Percentages

Number of items

#### 2.4. Summary

The lexicostatistical figures presented above allow us to distinguish within the Asmat area four closely related languages whose mutual relationships lie between 69%-79%. They are: Central Asmat (CA), Casuarina Coast Asmat (CC), Citak Asmat (CI), and North Asmat (NA). Central Asmat is spoken in four dialects: Kawenak (KW), Keenok (KN), Keenakap (KP) and Sokoni (SO). They share between 84% and 90% cognates. No further division into subdialects as claimed by Drabbe is apparent in KW. Lists representing one and the same dialect share 93%-100% cognates.

The data in CC are still insufficient to signal the presence of dialects. The same applies to CI and NA.

### 3. CENTRAL ASMAT

I shall begin this chapter by presenting some general observations on the phonology and grammar of Central Asmat. We shall need this information when discussing the CA dialects and when we compare CA with CC, CI, and NA.

The sections following those observations will be devoted to a more detailed discussion of each of the four dialects. They will be further defined in terms of their phonological, lexical and grammatical characteristics and where the data allow it something will be said about subdialectal differences. Also the dialect boundaries will be tentatively established. The element of uncertainty is due to the fact that we still need to rely partly on local opinion, that is on what people in one village say about the language spoken in other villages known to them but not represented in the language samples.