Interactions of Semantics and Phonology: Evidence from an Austronesian Language[1]

Adrian Clynes
Universiti Brunei Darussalam

0 Introduction

The still-widely-accepted view of the relationship between semantics and phonology is, simply put, that there is none. Some even claim that there 'can be' no interaction between the two (Archangeli & Pulleyblank 1994:5, 433). For proponents of this view, often called the Double Articulation principle, phonological structure can never bear meaning.

Over the years a significant minority of linguists, including Trubetzkoy (1969 [1939]), Hymes (1974), Bolinger (1977), Jakobson and Waugh (1979), Woodbury (1987), and the contributors to Hinton et al (1994), have argued in various ways against this position. Woodbury, for example, presents strong evidence that postlexical phonological processes are used to signal 'linguistically significant expressive or other pragmatic meaning' (1987:685).

In this paper I aim to show that in Balinese, just as in English (Fudge 1970), semantics also interacts with phonology at a 'prelexical' level. Just as with post-lexical processes, expressive or pragmatic meaning is signalled by this interaction. My findings also support the views of those who have claimed that in a given language there will be a continuum of phonotactic structures, from those which are maximally well-formed through to those which are increasingly ill-formed, but tolerated.[2] The latter, I argue, are often tolerated *because* they occur in expressives: semantics 'licenses' marginal phonological structures (Clynes 1995).

The phenomena discussed in this paper cannot be dismissed as peripheral or marginal. Balinese makes extensive, productive, use of relatively 'ill-formed' phonological structures, to create expressive vocabulary, which moreover constitutes a major element in the lexicon. The latter fact is true for western Indonesian languages in general, having been often commented upon for Austronesian languages like Javanese (Gonda 1947,

Uhlenbeck 1949) and Malay (Gonda 1947, Carr 1966). Eisman (1990) and Scherzer (1993) both write on the large number of Balinese expressives.

The paper has the following structure. The claims made for Balinese are briefly exemplified in §1. I then discuss in §2 the only previous large-scale statistical study of the phonotactics-semantics correlation that I am aware of, Fudge 1970, showing how my approach differs from his. The statistical procedure I use to test the correlation is described in §3. Further synchronic evidence is presented in §4.1, and §4.2. I draw conclusions in §5.

1 Phonology and semantics: first exemplification

Where morphemes contain phonologically marginal phonotactic structures in Balinese, they regularly belong to one of three broad categories. Either they are members of one of the 'expressive' semantic classes (see below), or else they are loanwords (Chomsky & Halle 1968), or they are grammatical morphemes. For example, exceptions to the following (informally expressed) morpheme structure constraint are often associated with expressiveness:[3]

(1) Prefer morphemes where cooccurring [+ATR] vowels share the same value for the feature [HI]

This constraint 'disprefers' the cooccurrence in a morpheme of either of the mid-vowels /o/ and /e/ with either of the high vowels /i/ and /u/. Phonological evidence that it holds as a synchronic constraint in the language comes from loanwords, which are regularized in North Bali varieties of Balinese.[4] Except for **éling** (Javanese), the examples in (2) come probably via Malay, from a variety of languages:

(2)	Source	>	Balinese	
	roti		**ruti**	'bread'
	polisi		**pulisi**	'police'
	kopi		**kupi**	'coffee'
	éling		**éling ~ iling**	'remember'

Other evidence that the harmonizing structure is preferred comes from productivity. A count of morphemes containing

(...)$V_{[+ATR]}CV_{[+ATR]}$(...) sequences in Warna et al (1990) shows that 'regular' sequences outnumber 'irregular' by 10 to 1:[5]

Sequence	Tokens	Totals
hi-hi mid-mid	1305 917	2222 regular
mid-hi hi-mid	102 119	221 irregular

There are 66 non-loan morphemes containing the 'disfavoured' sequences iCé, éCi, iCo, oCi, uCo, and oCu listed in Warna et al. Of these 42, or 64% belonged to just two of the expressive semantic classes (defined in §3) BAD, and PLANT/ANIMAL.[6] Examples of morphemes containing these sequences, with semantic/functional types commonly found with them, include:

(3) a. BAD:
ingé 'careless' **bebéki** 'mischievous'
céti 'pimp' **bisén** 'attack from behind'
b. PLANT AND ANIMAL names:
bekicot 'snail' **bligo** 'kind of gourd'
croring 'k.o. fruit' **wéni** 'k.o. fruit'
c. Personal NAMES:
Séli, Kédi, Kéni, Maséni, Réli, Wéli
d. High style/honorific terms and other LOANWORDS
icén 'grant' Jav **biséka** 'royal title.HON' Skt
e. GRAMMATICAL words:
tidong 'not'

This then is a first, so far impressionistic, example of the correlation between phonology and semantics in word-formation processes: there is independent evidence for the marginality of the phonological structure in question, as well as an apparently higher than usual proportion of 'expressives' among the morphemes having that structure. The statistical procedure used to test the latter claim is described in §3.

Similar recurrent patterns in the related language Javanese led Uhlenbeck (1949, 1950) to distinguish between 'central' and 'peripheral' root morpheme classes. His central morphemes are phonologically totally regular, while peripheral root

morphemes are both phonologically unusual in some way, and belong to recurrent semantic or functional types similar to those above: loanwords, 'affective-expressive', onomatopoeia, adhortatives, names of plants and animals, sometimes deictic and *krama* [High/Honorific] morphemes (1950:32).

2 Fudge's approach, and the present one

Uhlenbeck's claims remain essentially anecdotal, however, in that they are not supported by a statistical analysis. One of the few attempts at a systematic study of the relationship between phonotactic constraints and semantics is the important work by Fudge (1970). Fudge makes a strong case for a relationship between phonological markedness and expressiveness in English, while pointing out that then current theories of phonology had no means of dealing with it. At the same time, Fudge's paper has shortcomings, most of which are shared with other, less rigorous studies of what many refer to as 'sound symbolism'.

Fudge assigns the label 'expressive' to a given word if it belongs to any of the following types (1970:162):

(i) Onomatopoeic words
(ii) Movement words
(iii) Affective words [including 'words denoting intense reactions (*horrible, marvellous*), words with a pejorative connotation (*grumble, sly* etc) and words with a jocular or endearing connotation (*plump*)']
(iv) Hypocoristic [ie 'pet'] names eg *Bob, Tom*
(v) Nursery words eg *teddy, bib*
(vi) Colloquial words and taboo words
(vii) Abbreviations eg *perk, rep*

Fudge provides strong evidence that a variety of structures in English, including complex syllable onsets, are regularly associated with morphemes with expressive meanings (1970:168). However, while he gives abundant statistics to support his case, there is virtually nil exemplification. A good example to illustrate is the sequence /tw/. Three types of phonological evidence suggest that syllable-onset /tw/ is a peripheral structure in English: (i) it occurs before a much smaller set of possible nucleus vowels than other clusters

(Gimson (1983:241)); (ii) unlike most other initial CC clusters, it has no /sCC/ counterpart, and (iii) unlike most other initial CC clusters, it has no syllable-coda 'mirror image' counterpart, /wt/. The following lists all monomorphemic (non-loan) words beginning with the sequence /tw/ listed in the Collins Shorter dictionary (McLoed & Makins eds, 1993), and grouped by Fudge's semantic types:[7]

(4)

onomatopoeic:	twang, tweet, twitter
movement:	twill, twiddle, twinkle, twirl, twine, twist; twinge, twitch, tweak, tweezers
bad/pejorative:	twaddle, twat, twee, twerp, twit
colloquial:	twig (understand); tweedledum & tweedledee
unclassified:	twain, twin, twice, twilight, twenty; twig, twitch [grass], twayblade [plant]; twite [bird]; tweed [orig. placename]

Note that the items listed in (4) as 'unclassified' by Fudge's categories fall into semantic classes said in other studies to be regularly associated with peripheral structure: grammatical words, place names, and the names of plants and animals (Uhlenbeck 1950, this study). Fudge's methodology suffers then in that it establishes in advance the semantic types being looked for, rather than seeing what correlations emerge from the data. Fudge's, like most previous analyses, also leaves itself open to charges of subjectivity. Fudge concedes, probably over-cautiously, that 'we are investigating a single idiolect (168)'. The results could therefore be argued to have that degree of validity only: they may, or may not, apply in other idiolects, and to that extent are not falsifiable.

While I agree that the fundamental correlation does involve an 'expressive/subjective' semantic element,[8] an analysis should aim to be objective, and reproducible by others working with the same data. I have therefore avoided classes such as 'colloquial', or the potentially circular 'expressive'/'affective'.[9] This is particularly needed when the language concerned is not the analyst's native tongue.

To avoid a proliferation of semantic classes in statistical analyses, I first test for the numbers of morphemes belonging to three classes only: SENSES, PLANT/ANIMAL, or BAD (defined in §3). These three classes recur in sufficiently high numbers in

morphemes containing peripheral phonological structures for them to be given special diagnostic status. The SENSES and PLANT/ANIMAL classes have the added advantage of not requiring a subjective judgement: anyone using the same dictionary would agree for example that in (4) 'twitch' is a type of plant, or that 'twiddle' and 'twitter' are classifiable in the SENSE class (defined in §3). These three classes, with other semantic classes recurring in phonologically peripheral morphemes, I refer to as the 'expressive' semantic classes.[10]

In subsequent analysis, I do include a wider variety of semantic and functional categories (below), including grammatical morphemes and loanwords. It is common cross-linguistically for both grammatical morphemes, and loanwords (Chomsky and Halle 1968:373), to behave as 'marked' subclasses in the phonology. Fudge counts them in the non-expressives in his totals: as a result his conclusions are expressed as strong tendencies, rather than as more categorical generalizations.

Fudge's study, while not entirely blemish-free, remains an important work, containing a range of impressive findings. It was novel (for the topic studied) in its insistence on methodological rigour, in its demonstration of strong statistical trends pervading a large data base, and in erring on the side of caution in its methodological decisions. Any study of the semantics-phonology nexus should aim to replicate those features.

3 Testing the claim statistically

The claim that morphemes which contain marginal phonological structures belong in unusually large numbers to a limited set of recurrent semantic (and other) classes is meaningless unless the levels of these classes in the phonologically regular portion of the lexicon are known. As a control, a sample of the semantic types of morphemes predicted to have the most 'neutral/optimal' possible structure in Balinese was made, again using Warna et al (1990) as database. This involved a sample of 209 morphemes of shape $C_1V(N)CVC_3$, where C_1 was one of the three commonest initial consonants, /p/, /t/ and /s/, and C_3 was one of the two commonest final consonants, /ŋ/ and /h/.[11] Any morpheme containing a dispreferred phonotactic combination, or one suspected as such,

was excluded. The numbers of morphemes belonging to the three 'criterion' semantic classes, the PLANTS & ANIMALS, SENSE and BAD classes, was then calculated, for this sample.

The 'peripherality' or otherwise of any other data set can then be established by comparing the numbers of morphemes in it belonging to the three 'criterion' semantic classes, with their levels in the control sample. I define these classes as follows:

- the PLANTS & ANIMALS class refers to names of both plants and animals, and of parts of plants and animals.

- the SENSE class includes (a) any morpheme which lexicalizes vivid sense impressions in some way, be these impressions of sound (onomatopoeics), sight, touch, taste or smell.[12] (b) morphemes lexicalizing human feelings and emotions.[13]

- the BAD class includes morphemes with either pejorative value, or else which refer to undesirable things (*eg* evil spirits) and states (*eg* mental, bodily, atmospheric).

Results of the control group analysis are given in Table 1 above, with a breakdown by expressive and other semantic/functional classes. The background level of morphemes belonging to the SENSE, PLANT/ANIMAL, and BAD classes is thus 33%:

/C_1...C_3/	SENSE	PL/AN	BAD	LOANS	Other PERIPH.	UNCLASS	Total
/p...ŋ/	3	4	3	5	4	21	40
/p...ɦ/	5	4	3	6	5	14	37
/t...ŋ/	4	3	3	3	2	20	35
/t...ɦ/	5	4	7	3	5	11	35
/s...ŋ/	4	2	9	4	3	11	33
/s...ɦ/	2	0	5	7	3	12	29
totals	23	17	30	28	22	89	209
=%'age	11%	8%	14%	13%	11%	43%	100%

total PLANT/ANIMAL, BAD, SENSE	70	(=33%)
total Expressives & loans	120	(=57%)
total Unclassified	89	(=43%)

Table 1 Semantic classes of morphemes in control sample

Compare these figures with those in Table 2, which gives the semantic classes of morphemes breaking, in different ways, the constraint preferring no more than one consonant of a given place of articulation per morpheme, underlyingly:[14]

structure	SENSE	PL/AN	BAD	LOANS	OTHER	UNCLS	Total
/k... ...k/	71	18	17	7	6	6	125
/#gVg.../	7	2	3	2	4	2	20
/...tVt#/	15	11	8	12	6	6	58
/#tVd.../	5	-	7	3	4	4	23
Totals:	98	31	35	24	20	18	226

total SENSE, PLN/AN, BAD	164 (=73%) p <.0000
total Expressives & loans	208 (=92%) p <.0000
total Unclassified	18 (= 8%)

Table 2 Semantic classes of some morpheme-types breaking the Constraint 'one $C_{\alpha\ place}$ per morpheme' (data from Warna et al 1990).[15]

The 'diagnostic' classes BAD, SENSES, PLANT/ANIMAL now constitute 73% of the total sample, versus only 33% in the control. A standard statistical test indicates that there is less than one chance in ten thousand that the difference between these figures could be due to random effects. The procedure used to determine this is described in the following paragraphs.[16]

I take the data samples analysed in Tables 1 and 2 to derive from distinctly different statistical populations, since from the outset the samples have been selected as phonologically 'optimal' and phonologically 'marginal' respectively. The proportion of morphemes belonging to the classes BAD, SENSES, PLANT/ANIMAL in the control sample is 0.33. Given this, the probability of obtaining a proportion of 0.73 of morphemes belonging to those same classes in the second sample can be calculated using the Median Test (Spiegel 1972:171, Hatch & Lazaraton 1991:271):

$$p(Z > t) = \Phi^{-1}(Z)^{17}$$

Using this formula, we obtain the result $Z = 9.105$, $p(Z > 9.105) < .0000$. That is, the probability of getting this difference between the two proportions in a purely random way

is less than one in ten thousand. If we redo the analysis, including *all* the potentially 'peripheral' elements (57% of the control sample classifiable as 'expressives' or loans, versus 92% in the second sample), we obtain the result: $Z = 9.042$, $p(Z > 9.042) < .0000$. Again, a probability of less than one in ten thousand.

In Clynes 1995 a variety of marginal phonological structures are analysed in this way. Repeatedly, I find that a high proportion of morphemes containing these structures belongs to the three criterion 'expressive' classes - a much higher proportion than is found with phonologically regular morphemes. The statistical test repeatedly shows that the disparity cannot be plausibly attributed to random effects.

4 Other evidence for the productive interaction of phonology and semantics

Independent evidence that violable constraints are active synchronically, and deliberately and productively broken to form expressives, is given in §4.1 and §4.2.

4.1 Nicknames

It could be argued that the correlations between expressiveness and phonological structure described in §3 are not the results of synchronic processes, but rather reflect constraints that are no longer part of the synchronic grammar of the language. Naming practices in Bali provide one important source of evidence that these processes *do* function synchronically.

Personal names are one of the semantic types regularly associated with marginal phonological structures (cf footnote 10). Most Balinese as children are given a deliberately inelegant, often meaningless, nickname, which for some is the only name they are known by throughout their lives. There is no doubt that these names have inelegant connotations. The generic term for them is **adan jelé**, 'bad name'. Where nicknames *are* meaningful, they often poke fun at physical appearance, are onomatopoeic, or are in other ways undesirable. Meanings in my fieldwork village included 'pot' [of a fat child], 'squashed flat', 'acrid', 'rowdy', 'turtle', 'thump!',

'squeak!', 'pilferer', and 'scar'. Not surprisingly, these days most people also have an alternative, usually Sanskrit-derived, 'good name' (**adan melah**), which others use when referring politely to them.

These nicknames invariably contain dispreferred phonological structures. Where they are otherwise 'meaningless' they provide good evidence of the synchronic nature of violable constraints, since their inelegant connotations must derive directly from their irregular form. Examples of meaningless, one-off, and definitely inelegant names, are:[18]

(5) **Klemug, Cidaku, Namprut, Cengkug, Gomblos, Toti, Latep, Ut, Jruit, Mahmuh, Joét**

Other (apparently) unique names, while not inelegant, are still phonologically marginal, such as these:

(6) **Kédi, Kéni, Maséni, Toti**

The names in (5) and (6) variously 'break' one or more of the following violable constraints:

1. Prefer morphemes in which two consonants of the same place of articulation do not cooccur: **Klemug, Cidaku, Namprut, Cengkug, Toti, Latep, Jruit, Mahmuh, Joét**
2. Prefer morphemes which are disyllabic: **Ut, Kicadu**
3. Prefer morphemes which do not contain nasal consonants in syllable-onsets: **Mahmuh, Namprut, Klemug, Kéni, Maséni**
4. Prefer morphemes where cooccurring [+ATR] vowels agree in height: **Kédi, Kéni, Maséni, Toti**
5. Prefer morphemes which do not contain CL clusters: **Klemug, Namprut, Gomblos, Jruit**
6. Prefer morphemes which are not complex: **Mahmuh, Toti**
7. Prefer morphemes which do not end in labials: **Latep**

I take this productive coining of 'inelegant' nicknames, using otherwise meaningless morphs containing dispreferred phonological structures, to constitute strong evidence for a synchronic interaction between semantics and phonology in Balinese. Based on evidence like this I argue in Clynes 1995

that marginal phonological structures generate 'expressive' meaning components directly.

4.2 Dorsal-dorsal interactions

One particularly striking example of the productive way in which phonologically marginal structures are used to create expressive vocabulary was touched upon in §3: the figures in Table 2 suggested that the constraint preferring no more than one consonant of a given place of articulation per morpheme is regularly broken to generate expressives. This is especially clear when morphemes containing more than one dorsal consonant are examined. The raw figures for morphemes of structure /#$C_{Initial}$··· ···C_{Final}#/ in Warna et al 1990 are given in Table 3:

	p	b	m	t	d	n	l	r	s	k	g	ŋ	h
p	1	1	14	120	36	65	64	69	76	131	22	144	100
b	3	0	14	105	21	57	89	64	77	124	58	145	102
m	0	1	9	31	2	22	27	25	23	37	18	58	40
w	0	0	0	9	4	25	12	6	16	17	3	37	33
t	21	16	14	17	19	78	59	41	110	138	37	125	113
d	12	0	7	26	1	33	31	14	15	48	9	55	29
n	0	0	2	2	0	6	4	2	3	2	0	13	3
l	15	15	3	78	47	38	0	100	65	80	48	116	112
r	11	10	13	59	14	18	4	0	18	58	26	63	56
c	9	7	10	65	10	8	34	38	17	135	22	122	61
j	4	0	5	45	5	21	15	26	13	29	27	59	30
ñ	2	1	2	9	5	3	3	1	0	2	9	18	18
s	51	28	20	187	52	66	63	118	7	97	52	176	139
y	0	0	0	2	0	5	0	0	0	0	0	1	1
k	33	27	28	157	55	86	68	100	115	193	42	243	114
g	22	3	24	101	6	18	64	89	66	93	80	188	87
ŋ	1	0	0	6	2	4	0	4	5	9	4	10	9

Table 3 Numbers of morphemes of structure /#$C_{Initial}$··· ···C_{Final}#/ in Warna 1990 (column = $C_{Initial}$, row = C_{Final})

The raw data alone suggest that there are restrictions dispreferring cooccurrence of more than one labial, apical stop, liquid (/l/ or /r/, which here apparently behave as a separate class), or laminal.[19] However, these figures also suggest that there is no constraint dispreferring cooccurrence of consonants belonging to the remaining place category, dorsal-dorsal. In fact there are *more* morphemes with dorsals in both initial and final positions, than with any other place combination.

It turns out, however, that morphemes with both initial and final dorsal consonants are almost exclusively expressive in meaning. This is shown in Table 4 below, which gives the results of an analysis of 153 morphemes of this structure. A random sample for each sub-type was taken. Clearly there is a very high correlation with expressiveness: 80% of the sample belongs to the three criterion semantic classes, with a big clustering in the SENSE category. The skew is again highly statistically significant, $Z = 10.247$, $p < 0.000$, using the methodology outlined in §3.

structure	SENSE	PL/AN	BAD	LOANS	OTHER	UNCLS	Total
/k... ...g/	27	-	2	-	-	1	30
/k... ...ŋ/	14	7	3	3	-	3	30
/g... ...k/	18	4	2	-	4	5	33
/g... ...g/	25	-	1	-	-	4	30
/g... ...ŋ/	8	5	7	-	2	8	30
	92	16	15	3	6	21	153
total SENSE, PLN/AN, BAD						123 (=80%)	p <.0000
total Expressives & loans						132 (=86%)	p <.0000

Table 4 Semantic types of /#C_{dorsal}... ...C_{dorsal}#/ morphemes

Table 4 does not include data for morphemes of structure /#k... ...k#/: this was given in Table 2, and is repeated below. For this combination the skew is particularly striking, with only 5% of morphemes not immediately classifiable into one of the expressive or other peripheral categories (see Appendix).

structure	SENSE	PL/AN	BAD	LOANS	OTHER	UNCLS	Total
/k... ...k/	71	18	17	7	6	6	125
total SENSE, PLN/AN, BAD						106 (=85%)	p <.0000
total Expressives & loans						119 (=95%)	p <.0000

Table 5 Analysis of a random sample of morphemes of shape /k... ...k/ (data from Warna et al 1990).

In short, and despite their occurrence in large numbers, the fact that morphemes with dorsal consonants in both initial and final position contain a very high proportion of expressives (and particularly so in the case of morphemes of shape /#k... ...k#/) leads to two conclusions:

1. The constraint dispreferring cooccurrence of consonants of the same place of articulation still holds, though in the non-expressive component of the lexicon only.
2. In the coining of expressive lexis this constraint is productively 'broken', using the dorsal place of articulation in particular.

5 Conclusions

I have given evidence that in Balinese there is a strong statistical correlation between morphemes containing certain less-than-well-formed phonological structures and a small set of expressive semantic classes (§3). (A more extensive analysis, with evidence from processes of language change, and explicit formulations of the semantic elements hypothesised to be associated with expressiveness, is found in Clynes 1995.)

Evidence from semantics lends support then to the view that languages permit a continuum of well-formedness in phonological representations, from optimally well-formed, through to barely tolerated (Trubetzkoy 1939, Fudge 1970, cf also Prince and Smolensky 1993). Importantly, it also enriches it, supplying another source of evidence for the marginality, or otherwise, of phonological structures. Semantic data, properly handled, *can* be used in phonological analyses, despite generally accepted practice.

To the extent that less-than-optimally-well-formed structures surface as the product of competing principles in the phonology and other components of linguistic structure (Prince and Smolensky 1993), this study identifies one of those principles: marginal phonological structures are used productively in the formation of expressive lexis, whereas they are dispreferred in the generation of the non-expressive component of the lexicon.

In a sense, then, expressive semantic content 'licenses' non-optimal phonotactic structures in Balinese morphemes. These results throw still further doubt on the 'double articulation' view of linguistic structure - even in a modified version, such as that proposed by Woodbury (1987).

Like Fudge (1970), I found that the lower the productivity of a given phonological structure, the more likely it is that only

expressive and other peripheral morphemes will occur with it. However, the reverse generalisation does not apply: in Balinese a phonologically marginal structure can occur very frequently, given that peripheral phonological structures are used productively in the creation of expressive vocabulary (§4.2).

The fact that 'pragmatically generated' semantic elements associated, say, with the usage of names (§4.1), in turn influence the phonological shapes we give to newly coined names (or *mutatis mutandis*, any expressive terminology) is further indication of the artificiality of attempts to separate (linguistic) pragmatics and semantics (Wierzbicka 1989).

The finding that expressives pattern almost like a distinct lexical class, with seemingly distinct principles of word-formation applying to them (though see Itô and Mester (1995)), reflects how human cultural values (and emotions) intimately interact with, and shape, linguistic structure (cf Wierzbicka (1985, 1988)). That the names of plants and animals, for example, regularly pattern as expressives reflects their *fundamental role in all human cultures: information from* phonological analysis is relevant not merely to other branches of theoretical linguistics, but to a broad range of social sciences.

Notes

[1]Previous versions of this paper were presented at the ANU Canberra 1992, at the 1992 ALS meeting Sydney, and at UWA, November 1993; thanks in particular to Bapak Ketut Mandi Pinatih, Ketut Nanik Clynes and Juliette Blevins. Thanks also to Avery Andrews, Carl James, Andy Pawley and Malcolm Ross, for comments on a fuller version, in Clynes 1995. Fieldwork for this study was undertaken in Indonesia under LIPI Research permit 2795/SK/1990. I am grateful to LIPI for their support.

[2]I have not had direct access to the unpublished work of Prince & Smolensky (reported, e.g., in Goldsmith 1990, Archangeli & Pulleyblank 1994, Gafos, this volume).

[3]I express the well-formedness continuum in terms of informal 'violable constraints', which I hope translate easily to frameworks such as Prince & Smolensky 1993.

[4]People aged below about 50 no longer regularize loan words like this, due no doubt to the influence of Indonesian.

[5]Database is the principal dictionary of Balinese, Warna et al (1990), which lists around 15,000 root morphemes. Counts were done using Alsop (1990).

[6]Data for this, and all other counts referred to in this paper, is given in full in Clynes 1995, along with more extensive analysis. See also footnote 15 & the Appendix.

[7]Abbreviated forms, such as *'twixt* and *'tween* are not included.

[8]See Clynes 1995 for a proposed explication of this semantic element.

[9]Uhlenbeck has an 'affective-expressive' class, and Trubetzkoy (1969 [1939]:254), all 'words with an 'expressive' coloration'.

[10]Minor semantic classes regularly associated with phonologically peripheral morphemes include (personal and place) names and terms of address, childish terms, names of good things, and true patient verbs; see Clynes 1995.

[11]Like Javanese (Uhlenbeck 1949), Balinese has around 90% disyllabic morphemes. The types CVCVC and CVNCVC together constitute around 50% of all disyllabics.

[12]'Sight' terms include verbs of movement, particularly those which lexicalize movements of the body, and of body parts, such as (in English) skipping, nodding, twitching, and so on.

[13]Balinese **asa** translates variously as 'taste', 'feeling (physical, including touch)' and 'feeling (emotion)'; the emotions are thus lexicalised as a 'sixth sense' in Balinese.

[14]This feature, due no doubt to an OCP violation, is of course common cross-linguistically: it is shared for example with Arabic (McCarthy 1986) and the related Austronesian language Javanese (Yip 1989).

[15]The /k... ...k/ data, and my analysis of it, is given in full in the Appendix below. The /k... ...k/ and /#gVg.../ samples were the first 125 and 20 items respectively of the lists produced by a data base search, using Alsop (1990). The /...tVt#/ and /#tVd.../ samples include all the morphemes of those structures in the data base.

[16]I am grateful to Yvonne Pittelkow of the RSPAS Computer Services Centre, Australian National University, for help with the statistical analysis.

[17]Where $Z = p_1\text{-}p_2/(p_1q_1/N_1 + p_2q_2/N_2)^{1/2}$, and p_x = the proportion of 'peripherals' in sample number 'x', q_x = the proportion of 'unclassified' elements in sample number 'x', and N_x = total number of items in sample 'x'.

[18]This list contains fictional nicknames, created by making minimal alterations (eg, by substitution of one phoneme of almost the same place/manner for another) to the real nicknames of people in the village where I did my fieldwork. See Clynes (in preparation).

[19]/s/ is the only laminal to occur word-finally; glides never do.

References

Alsop, John R. 1990. Fiesta. Version 4.5. Summer Institute of Linguistics. Computer software.

Archangeli, Diana & Douglas Pulleyblank. 1994. Grounded *phonology. Cambridge MA: The MIT Press.*

Bolinger, Dwight L. 1977. Meaning and form. London: Longman.

Carr, Denzel 1966. Homorganicity in Malay/Indonesian in expressives and quasi-expressives. Language 42:370-377.

Chomsky, N. and M. Halle. 1968. The sound pattern of English. New York: Harper and Row.

Clynes, Adrian. 1989. Speech styles in Javanese and Balinese: a comparative study. MA thesis. Australian National University.

_____. 1995. Topics in the phonology and morphosyntax of Singaraja Balinese. Dissertation. Australian National University.

Eisman, Fred B. Jr. 1990. Bali: Sakala & Niskala. Volume II, essays on tradition and craft. Periplus Editions, Berkeley-Singapore. chapter 11, Onomatopoeia: a penchant in Balinese speech.

Fudge, Erik. Phonological structure and 'expressiveness'. Journal of Linguistics 6:161-188

Gimson. A. C. 1983. An introduction to the pronunciation of English. Third edition. London: Edward Arnold.

Gonda. J. 1940. Some remarks on onomatopoeia, sound symbolism and word formation a propos of the theories of C.N. Maxwell. Tijdschrift voor Indische Taal-, Land- en Volkenkunde 80:133-210. Reprinted in: J. Gonda, 1975. Selected studies. Volume V Indonesian linguistics. Leiden: E.J. Brill.

Goris, Roelof. 1954. Prasasti Bali. Bandung: Lembaga Bahasa dan Budaya/Masa Baru.

Hatch, E. & A. Lazaraton. 1991. The research manual: design and statistics for applied linguistics. New York: Newbury House.

Hinton, Leanne, Johanna Nichols & John J. Ohala. 1994. Sound Symbolism. Cambridge: Cambridge University Press.

Hymes, Dell. 1974. Foundations in sociolinguistics. Philadelphia: University of Pennsylvania Press.

Itô, Junko & R. Armin Mester. 1995. Japanese Phonology. In: J. Goldsmith (ed.) The handbook of phonological theory. Oxford & Cambridge, Mass.: Blackwell.

Jakobson Roman, and Linda R. Waugh. 1979. The sound shape of language. Bloomington: Indiana University Press.

Prince, A. & P. Smolensky. 1993. Optimality theory. Ms. Rutgers University & University of Colorado, Boulder.

Scherzer, Joel. 1993. On puns, comebacks, verbal dueling, and play languages: Speech play in Balinese verbal life. Language in Society 22:217-233.

Spiegel. Murray R. 1972. Schaum's Outline of Theory and Problems of Statistics in SI Units. New York: McGraw-Hill.

Trubetzkoy, N.S. 1969. Principles of Phonology. University of California Press. [GrundzÜge der phonologie, Transactions du Cercle Linguistique de Prague 7, 1939, transl. C.A.M. Baltaxe]

Uhlenbeck, E. M. 1949. De structuur van het Javaanse morpheem. Verhandelingen van het Bataviaasch Genootschap 78.

_____. 1950. The structure of the Javanese morpheme. Lingua 2:239-270.

Warna, I Wayan & Panitia Penyusunan Kamus Bali [Dictionary Team]. 1990. Kamus Bali-Indonesia [Balinese-Indonesian Dictionary]. Denpasar: Dinas Pendidikan Dasar Bali. Departemen Pendidikan dan Kebudayaan.

Wierzbicka, Anna. 1985. Lexicography and conceptual analysis. Ann Arbor: Karoma.

_____. 1988. The semantics of grammar. Amsterdam: John Benjamins.

Woodbury, Anthony C. 1987 Meaningful phonological processes: a consideration of Central Alaskan Yupik Eskimo prosody. Language 63:685-740.

Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. Phonology 6:349-374.

Appendix

Analysis of morphemes of shape /#k... ...k#/ (cf Table 2), by semantic class. Sample is the first 125 items produced by a database search, using Alsop (1990)

1. SENSE 71 Items (45 Sound; 26 movement, last 11 of which are True Patient verbs, involving movement):

1(a) SOUND: **kaak**, **kaek**, **keek**, 'sound of someone bringing up phlegm'; **kauk** 'call out loudly'; **kebruk** 'sound of fart'; **kebyak** 'sound of a whip cracked'; **kebék** 'noisy (of a crowd)'; **kecik** 'make noise (of a crickets wings)'; **kakak** 'laugh loudly'; **kaplak**, **keplak**, **kaplek**, **kaplok**, **kapluk** 'slap; slapping noise'; **kampak**, **kapak** 'hatchet '; **kandik** 'axe' (the last three

are assumed to be iconic of the sound they produce, based on the common meaning and phonological shape); **kedokdak** 'boil, making a bubbling sound (of water)'; **kedék** 'laugh'; **kempluk** 'small gong, used to beat time'; **kepak** ~~ 'sound of birds wings in flight'; **keplék** 'sound of a flat object falling into water'; **kepwak** 'bamboo inst. which makes a *pwak* noise'; **kerak** 'scream'; **kerik** 'scream'; **kerok** 'sound of frog, monkey'; **ketekok** 'sound of a hen clucking'; **ketepluk, ketipluk** 'small drum'; **ketok** 'sound of hammer'; **kiak, kiek** ~~ 'cheep (vi)'; **kiik** 'sound of chick or duckling'; **kikik** 'snigger'; **kiok, kéok** 'sound of a hen in pain'; **kinyuk** 'make a smacking noise with the lips'; **klecek** 'sound of a nearly-ripe coconut'; **klecok** ~~ 'sound of person walking in mud'; **klesek** 'sound as of people touching each other' **kletak** 'beating sound'; **kletek** 'sound of wood being cut with an axe'; **kletok** 'sound of a plank, board being struck'; **kletuk** 'sound as of walking stick on floor'; **kleték** 'sound of hard object being beaten';

1(b) MOVEMENT: **kaplik, kaplék** 'floppy (of breasts)'; **kecek** 'press with fingernail'; **kemik** 'move the mouth'; **kemuk** 'chubby (of cheeks)' ~~ 'move (of chubby cheeks, when eating); **kilak-kiluk** 'have many bends and turns'; **klisik** 'move; give off a sound'; **kipek** 'move face to one side' ~~ look repeatedly to the sides; **kisik** ~~ 'make tiny movements'; **kituk** 'shake (head)'; **kletik** 'bounce off sthg (of gravel, pebble)'; **kocok** 'shake back and forth'; **kerék** 'scrape a match to light it, sweep a broom'; **kocak** 'wobble, come loose (of teeth)'; **klincak** 'always going off walking'; **kepik** 'pick (leaf, flower)'; **keplok, kepluk** 'hit until burst'; **kepuk** 'break off, snap off'; **keplek** 'suddenly faint'; **kerik** 'scrape with a knife'; **keruk, keduk** 'hollow out, scrape out; scraping noise'; **ketek** 'chop into small bits'; **kiruk, kicuk** 'make a hole in (with a knife)';

2. BAD 16 items: **kacék, kecék** 'not enough (wrongly counted, of money)'; **kacak** 'scattered about (vi)'; **kadéngklok** 'trip up'; **kagek** 'unexpectedly forced to stop'; **kagok** 'speechless, unable to reply'; **kaluk** 'disorderly, noisy (crowd)'; **kapok** 'repent an action'; **katék** 'go off alone, stand apart'; **katuk** 'have sex (crude)'; **kayak** 'stretched out exhausted on the ground'; **kejok** 'too short'; **kemaruk** 'ravenous'; **kerék** 'skin disease'; **kisek** 'packed, overcrowded'; **kolok** 'deafmute';

3. PLANT/ANIMAL 18 items: **kacwak** 'cockroach'; **kaikik** 'kind of banana'; **kaliampwak** 'k.o. tree'; **keciplak, kecipluk** 'k.o. shrub'; **kapuk** 'kapok'; **karuk** 'k.o. plant'; **katak** 'frog'; **katik** 'stalk (of plant)'; **kirik** 'k.o. duck'; **kelik** 'k.o. bird'; **kepokpok** 'k.o. manggo'; **ketoktok** 'k.o. plant'; **klopak, klopék** 'sheath surrounding flowers/fruit (kelopak)'; **klutuk** 'k.o. spirit/animal'; **kedukduk** 'k.o. plant'; **klampwak** 'k.o. tree'

4. LOANS/HI/HON 7 items: **kaduk** 'even though.HI'; **kedik** 'small amount.HI'; **kekerik** 'border.HON'; **kekosok** 'rice flour, used in a ceremony' Jav.; **kibik** 'cubic' Dutch; **kidik** 'a small amount.HI'; **kodak** 'camera' Engl.

5. OTHER 6 items: **kadék** 'title of 3rd-born child' NAME; **kampék** 'small bag' SMALL; **kobok** 'small bowl' SMALL; **kesyek** 'VERY [lonely (of place), deserted]'; **kabak** 'boy-/girl-friend' FAMILIAR; **keñcak** 'clever, diligent' GOOD;

6. UNCLASSIFIED 6 items (4%): **katok** 'underpants'; **keték** 'count' (vt); **kléték** 'yoke (n&v)'; **kitak** 'brick'; **klopok** 'wooden box'; **kolek** 'dish, boiled bananas'