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

$(\ldots)V_{[+ATR]}CV_{[+ATR]}(\ldots)$ 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) Hypo coristic [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