

Expressives and iconicity in the lexicon

Marian Klamer

Vrije Universiteit Amsterdam

In: F. K. Erhard Voeltz and Christa Kilian-Hatz (eds.), *Ideophones* [*Typological Studies in Language* 44] Amsterdam: John Benjamins, 2001, pp. 165-182.

Expressives and iconicity in the lexicon *

Marian Klamer

Vrije Universiteit Amsterdam

0. Introduction

It has often been observed (e.g. Joseph 1994) that the lexicon of a language consists of a core and peripheral part. The core part is made up by the normal referential elements, whose primary function is to refer to entities, events, concepts, etc. It would at least include body part, kinship and color terms. The peripheral part consists of items with an added non-referential function, whose non-referential (descriptive, evaluative) function is at least as important as their referential function. It includes lexicalisations of vivid sense impressions such as onomatopoeics, ideophones and phonestemes, names, and morphemes with negative connotations or referring to undesirable states (cf. Clynes 1998, Joseph 1997). These semantically and functionally special items are referred to as ‘expressives’ in this paper.

Semantically, expressives are more complex and more specific than normal referential elements. An illustration are the following English pairs, where the first item is a common referential element, and the second an expressive: *the lunchroom* – *the Sizzle*, *breast* - *boob*, *walk* – *stagger*. We may say that the descriptive and/or evaluative function of expressives is at least as important as their referential function.

Traditionally, expressive elements are seen as iconic, sound symbolic forms, on a par with onomatopoeia, phonestemes, Japanese mimetic forms, or ideophones. In this paper we will see that the class of expressives is much larger, and includes forms that are not sound-symbolic at all. We will also see that, though sound symbolism is not involved, the more abstract non-arbitrary relations between form, meaning and function are still motivated by iconicity. I describe this iconicity as a type of diagrammatic iconicity (cf. Peirce 1965, Haiman 1994). This iconicity is the factor which restricts the coinage of new words, and is thus the explanation on the limits of this type of creative language.

1. Iconicity and the concept of markedness

Among the various types of signs, icons are the ones that show a factual resemblance between the sign and the object.¹ In its strictest interpretation, iconicity aligns meaning A and form B because of the inherent similarity of A and B – the ‘icon’, form B, is an image of ‘concept’ A.

When we apply this strict interpretation of iconicity to expressives we immediately run into descriptive problems. Many expressive words refer to sights, motions, or states where no sound is involved, so their phonetic form can never be strictly iconic to begin with. Secondly, because the formal properties of expressives are often not an ‘image’ of the ‘concept’ represented. For example, expressives often have unusual phonotactics, such as a closed syllable rather than an open one, but whether or not a syllable is not a direct image of expressive semantics. Even in the case of onomatopoeics, which are commonly cited as prototypical instances of iconic alignment of sound and meaning, there are language-particular abstractions of the actual sound they describe. For instance, the sound made by a bulky object falling is represented by an item containing a stop consonants and a back vowel in English, Dutch and Kambera (Austronesian), but in Didinga (Nilo-Saharan) the form includes a vowel /I/ and a velar nasal:

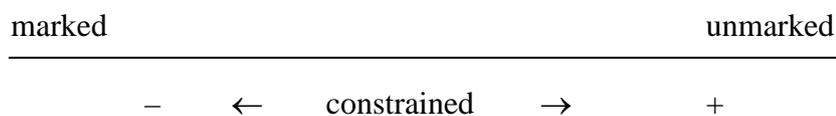
(1) meaning ‘sound made by bulky object falling’

form			
	<i>thud</i>	English	
	<i>plof</i>	Dutch	
	<i>mbùtu</i>	Kambera	(Klamer 1998)
	<i>tdIN</i>	Didinga	(De Jong 1999)

However, we should not take this as evidence to abandon iconicity as a motivation for the form-function relation in expressives, because this would imply that the relation is arbitrary, in contradiction to the findings of e.g. the authors of this volume and those in Hinton *et.al.*(1994). Rather than abandoning iconicity as an explanatory force in word formation, we must refine it. In line with Battistella (1990), the analysis of iconicity presented in this paper draws on the concept of markedness in order to align the form, meaning and function of lexical items. Lexical items occupy a position on continua for their markedness: they can be formally (structurally) more or less marked, but also semantically and functionally. I will now discuss these three markedness continua. In section 3 I present evidence for these continua and the correlations between them, and in section 4 we will see that these correlations are iconic, and how iconicity is a creative factor in the lexicon.

With respect to their formal (physical, structural) properties, lexical items can be arranged along a continuum in which one end is for the optimal word forms, and the other end for the marked word forms. If we view the structural system of language as a set of constraints on the wellformedness of linguistic utterances, then the more constrained items are the unmarked ones, and the formally less constrained items are the marked ones:

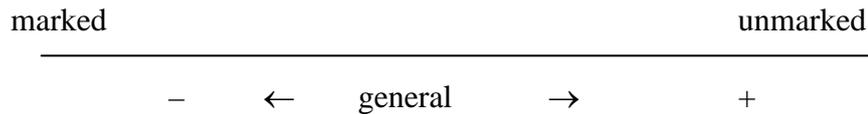
(2) Form:



An illustration of this is the fact that ideophones often use segments not belonging to the regular phonemic inventory and violate phonotactic constraints of the language. They often display very little or exceptional morphology (e.g. reduplication) and rarely possess any syntax unique to their class except the relative absence of any syntax (cf.). They are thus formally less constrained than the core lexical items (cf. Mithun 1982, Childs 1994, among others). Formal markedness is thus a relative notion, defined in terms of degree of formal constrainedness. On this continuum, expressive elements are subject to less constraints, i.e. more marked, than core lexical items.² Section 3 presents some evidence for this.

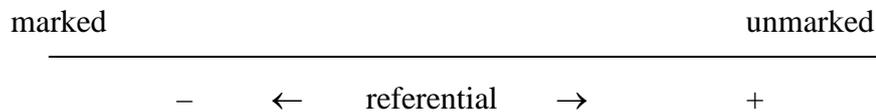
On the continuum of meaning (amount of semantic features), marked elements are characteristically less general than unmarked items, and more determinate and differentiated (Battistella 1990). For example, ‘ideophones can be quite specific, often evoking some concrete imagery. They often appeal to the senses and have a narrow meaning. Kisi (...) has two rice-beating ideophones: *gbun gbun* ‘rice beaten by one person’ and *pim pim* ‘rice beaten by two or more people’ (Childs 1994: 188). As a result of their semantic specificity, expressives have less ability to combine with other features, so the range of contexts in which they appear is smaller than for unmarked items. Unmarked items, on the other hand, are capable of having a general interpretation and can substitute the marked items in some contexts, while the reverse is not true. Illustrations are the English pairs given in the introduction. The semantics of unmarked items is more general and prototypical:³

(3) Meaning:



The third relevant continuum is the continuum of function, ranging from referential to non-referential. As the prime function of language as a sign system is referential, we expect that formally marked, functionally ‘peripheral’ (cf. Uhlenbeck 1950) lexical items, such as expressives, are less referential than unmarked, core lexical items. Expressives are considered non-referential because their function is not primarily referential. They are characteristically used to add vividness, liveliness, and directness to utterances and may be used creatively in verbal art (language games, insults). Their exceptional syntactic behaviour is also evidence for their non-referential function: expressive elements are often used in connection with or juxtaposed to exclamations and interjections, are often syntactically redundant or optional (adjunct) constituents, and do not allow negation (cf. Childs 1994, Kita 1999).

(4) Function:



These three markedness continua can be used to show that the correlation between form, meaning and function is not totally arbitrary, but rather patterned by diagrams of nonarbitrary, iconic relations. As iconic signs are not always images, they are distinguished in ‘images’, ‘diagrams’ and ‘metaphors’ (Peirce 1965). Images partake of simple qualities of their object (e.g. onomatopoeia); diagrams represent the relations, mainly dyadic, of the parts of one thing by analogous relations in their own parts; and metaphors represent the representative part of a sign by representing a parallelism in something else (Battistella 1990:71). The concept of diagrammatic iconicity is defined here as follows:

(5) Diagrammatic iconicity

Not the component parts of the diagram resemble what they stand for, but the relationships among those components: the relative position of an element in system A is matched by an element with the same relative position in system B.

In this paper I argue that the markedness patterns described above are in a nonarbitrary relation to each other and that the diagrammatic iconic mapping between them explains why certain choices are made from among competing possibilities in the coining and perception of new words. We will also see that the alignment of formal marking and semantic markedness is a tendency and not compulsory.

2. Formal and semantic markedness

Before considering the correlation between marked form and meaning in actual language data, we translate our notions of formal (structural) and semantic markedness in diagnostic terms.

Formal markedness I define in terms of constraint violation: a lexical item is formally marked if it violates one or more formal constraints of the language. The violated constraints may be phonological, morphological or syntactic. Within the limits of this paper I only discuss some phonological constraints in a number of Austronesian languages, plus one general constraint on semantic transparency (cf. immediately below), but the implication is that all the constraints in a constraint-based theory of language can in principle be relevant as markedness values in a particular language.

A constraint that is systematically violated by expressives across languages is the constraint on Semantic Transparency (Klamer 1999a, 1999b), a constraint that refers to the universal tendency that linguistic items prefer a direct, one-to-one matching of form and meaning.:

(6) Semantic Transparency: Match form and meaning one-to-one:

meaning	A
form	X

The following configurations violate this constraint, and the illustrations given indicate that unusual morphological structures often represent such configurations.

- | | |
|--|--|
| <p>(7) (a) * A</p> <p style="padding-left: 40px;">/ \</p> <p style="padding-left: 40px;">X...Y</p> <p>(e.g. circumfix)</p> | <p>(b) * AB</p> <p style="padding-left: 40px;">\ /</p> <p style="padding-left: 40px;">X</p> <p>(e.g. homophones)</p> |
| <p>(c) * —</p> <p style="padding-left: 40px;"> </p> <p style="padding-left: 40px;">X</p> <p>(e.g. empty prefix)</p> | <p>(d) * A</p> <p style="padding-left: 40px;"> </p> <p style="padding-left: 40px;">—</p> <p>(e.g. zero morpheme)</p> |

In the next section we will see how this constraint is violated by expressives in Kambera and Balinese. Examples of relevant phonological constraints are the ones that involve infrequent, complex or unusual segments or prosodic entities, i.e. constraints similar to the markedness and alignment constraints of Optimality Theory. We will see illustrations of these in the next section.

The notion of semantic markedness can be translated into three diagnostic types, following up on a proposal by Clynes (1995, 1998) Semantically, expressive elements are marked because they are very specific. I propose that they are restricted to only three semantic types: Sense, Name and Bad, specified as follows:

- (8) **Sense**: lexicalisations of vivid sense impressions: sound, touch, taste, smell, feeling, emotion and sight (including lexicalisations of movements of the body and of body parts). E.g. Kambera *tòku* ‘knock, bang’ (sound), *holap* ‘be weak, limp, flabby’

Name: personal or place names, hypocoristic names, terms of endearment; names for plants and animals E.g. Kambera *pirih* ‘k.o. parrot’, *helap* ‘kind of fish’

Bad: lexical items with negative connotations or referring to undesirable (mental, bodily, atmospheric) states and referents. E.g. Kambera *nyimba* ‘be blocking the way’

In the study of sound symbolic forms and their semantics, the first type is always mentioned. The second and third semantic type may seem somewhat surprising to some. However, it is a well-established fact that names, nicknames and animal names often reflect audible or visible characteristics of the named person or animal (e.g. body shape, hair colour, bird’s call, animal movement), whereas plant names may reflect the typical shape of (parts of) the plant (e.g. leaves, branches). For example, in Mundang (Niger-Congo), animal and plant names are part of the expressive vocabulary (Elders 1999). In Estonian, bird names are expressive, and in Finnish too, to some extent (Antilla 1976). In Greek, nicknames pattern with the expressives (Joseph 1997). As names can be seen as lexicalisations of sense impressions, the distinction between the types Sense and Name is a fluid one.

In many languages, words with bad or negative connotations formally pattern with the expressive items. This has been described, for example, for Japanese (Kita 1997:98, Hamano 1998), Balinese (Clynes 1995, 1998) and Greek (Joseph 1997). Further examples from Malay, West Tarangan and Dutch will be presented in section 3. Note that speakers of a language need not be consciously aware of this fact. For example, that /çr-/ is a marked onset in Dutch (Booij 1995) is known at least among Dutch linguists, but that all the words with this marked onset only belong to the semantic types Sense and Bad (i.e. refer to sense impressions or have negative connotations) has so far gone unnoticed.

Now we have translated our notions of formal and semantic markedness into diagnostic values and types, we consider in the next section the correlation between marked form and meaning in some actual language data.

3. The correlation between marked form and meaning

The correlation between form and meaning in sound symbolic words such as onomatopoeia and ideophones is very well-known. Therefore I will just give one illustration of this by discussing the ideophones of the Austronesian language Kambera. Then I will go on discussing data from other Austronesian languages that only on second sight appear to match form and function in an iconic fashion.

3.1. Kambera

Kambera (Klamer 1998, 1999b) is one of the few Austronesian languages with ideophones. Ideophones are prime examples of how formal, functional and semantic markedness can be aligned. The ideophonic roots of this language describe sounds, sights and motions, and are commonly used to add vividness and liveliness to a narrative text. Some ideophones are more conventionalised than others, and new forms are constantly being created. Kambera has only one circumfix, *ka-k*, which is the affix used to derive verbs from ideophonic roots. Some illustrations of Kambera ideophonic roots and their derivations:

(9)

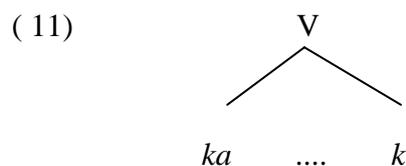
Root form	Derived verbal form	Gloss of verbal form
Sounds		
<i>ngùru</i>	<i>ka-ngùru-k</i>	‘murmur’
<i>hèri</i>	<i>ka-hèri-k</i>	‘tear something’
<i>tòru</i>	<i>ka-tòru-k</i>	‘rattle’
<i>pàka</i>	<i>ka-pàka-k</i>	‘smack’
<i>tiku</i>	<i>ka-tiku-k</i>	‘creak/click’
Motions		
<i>yidi</i>	<i>ka-yidi-k</i>	‘shiver’ (in dislike)
<i>wàdi</i>	<i>ka-wàdi-k</i>	‘blink’
Sights		
<i>rèri</i>	<i>ka-rèri-k</i>	‘ablaze’ (fire)/ ‘shine’ (ring)
<i>bila</i>	<i>ka-bila-k</i>	‘light/brightness’

Semantically, all these ideophones belong to the *Sense* type. The forms are phonologically marked because they are the only Kambera lexical items with marked vowels. The unmarked vowels of Kambera are the cardinal vowels /i,u,a/ with a long/short opposition. The marked vowels are the short lowered vowels [ɛ, ɔ] (represented as è, ò) and the super-short high vowel [ɨ]. The marked central vowel is the short low vowel [ʌ] (represented as à). We describe the general dispreference for these marked vowels in Kambera as a markedness constraint against such vowels:

(10) * /ɨ, è, ò, à/ “Don’t be /ɨ, è, ò, à/”

Because the marked vowels occur only in ideophones, ideophones violate this constraint, which is otherwise not violated in the language. But the marked vowels are not a categorical property of ideophones: though /è, ò, ù/ occur in ideophonic roots only, /à/ is used in a number of non-ideophonic roots as well, e.g. *tàka* ‘arrive’.

Verbal derivations from ideophonic roots are also morphologically marked: they are the only Kambera items with a circumfix, violating the Semantic Transparency constraint by mapping two forms on one function:



The verbs derived from ideophonic roots are also phonotactically marked because they have a final C, namely the *-k* part of the circumfix *ka-k*. Now, the unmarked root form in Kambera is a bimoraic trochee CVCV and this form covers approx. 50 % of the roots. At the same time, roots with closed syllables are dispreferred. In a constraint-based model of language we can describe morpheme structure conditions in terms of surface constraints on markedness and alignment (cf. Booij 1998). The condition here can be formulated as an alignment constraint where the right edge of a root morpheme (M) must be aligned with a vowel (V):

(12) “Align (V, R, M, R)” “Avoid roots with closed final syllables”

But 30% of the Kambera roots are CVCVC forms, examples:

(13)	<i>tehik</i>	‘sea’	<i>watar</i>	‘corn’
	<i>padang</i>	‘field’	<i>engal</i>	‘k.o. tuberous plant’
	<i>múhung</i>	‘be rotten, gone bad’	<i>mungal</i>	‘fall out, slip out, slip off’
	<i>helap</i>	‘k.o. sea fish’	<i>bàbat</i>	‘bamboo slat (part of loom)’
	<i>holap</i>	‘weak, limp, flabby’	<i>hoput</i>	‘be very angry, put out’
	<i>pirih</i>	‘parrot <i>Trichoglossus heamatodus</i> ’		
	<i>palih</i>	‘lick mouth/lips’		

The constraint should therefore not be taken to imply that Kambera does not have roots with closed final syllables. Rather, the constraint penalises closed syllables in output forms, and as a result, an epenthetic (‘paragogic’) vowel [u] is always inserted to ‘repair’ lexical roots ending in a closed syllable. In other words, lexical codas become the onsets of an additional syllable in output forms: /pu.duk/ → [pu.du.ku] ‘kiss’.

Despite their frequency, the CVCVC roots are thus formally marked forms, because they violate the constraint on roots with a closed final syllable. The question is now whether the formal markedness correlates with semantic markedness. This is indeed what is found: of a test sample of 145 CVCVC forms, 71% belong to the types Sense, Name or Bad, while 29% were unclassified:

Table 1. The final consonant of CVCVC roots, C1=/p/ (Onvlee 1984)

C3	k	N	l	r	h	t	p	Totals
Sense/Name/Bad	46	7	14	14	11	11	0	103
Unclassified	8	18	2	7	2	5	0	42

We conclude that the presence of a lexical final consonant of a Kambera root signals semantic markedness.

It is therefore also no surprise to find that verbal derivations of ideophonic roots end in a consonant (-*k*). Note however that the final C is *not* a categorical property of expressives, because not all CVCVC roots are expressive (only 71%); and not all final C's signal expressiveness (e.g. final /N/ also functions in word derivation as a valency changing (applicative) suffix).

In other words, the Kambera ideophones are semantically marked special forms referring to sounds, sights and motions. They violate three formal markedness constraints of the language: on segmental selection, on phonotactic structure and on morphological marking. This correlation of formal and semantic markedness is an example of diagrammatic iconicity.

3.2. Balinese

In Balinese too, semantic and formal markedness are aligned (Clynes 1995, 1998). Balinese expressives violate at least one, but usually more of the six constraints listed below. Balinese nicknames are an especially clear instance of this: the nicknames are inelegant ‘bad names’ and are otherwise meaningless. All of them violate at least one constraint. Illustrations:

(14) Constraint violation by Balinese bad names (Clynes 1995: 51-52, 1998: 21-22)

Onset: Every syllable must have an onset

Cluit, Joet

* Complex^{ONS} : Avoid complex onsets

Kl↔mug, Namprut, Gomblos, Cluit

* σ [/h/ : Avoid /h/ as onset

Cibuhut

Root= $\sigma\sigma$ Roots must be bisyllabic

Cidaku, Cibuhut, Maseni

Vowel harmony: Cooccurring [+ATR] vowels agree in height

Kedi, Keni, Maseni, Toti

Consonant disharmony

Two consonants with the same place of articulation do not cooccur in a root

Cidaku, Namprut, Toti, Latep, Petet

Balinese also has a special type of ‘inherent reduplication’ forms (Clynes 1995: section 4.2, p. 71). Because they contain meaningless reduplicated elements, they violate the Semantic transparency constraint. Semantically, they belong to any of the three types we defined above. Examples:

(15)	<i>mémé</i>	‘mother’
	<i>crukuk</i>	‘k.o. bird’
	<i>kedongdong</i>	‘k.o. fruit tree’
	<i>lumbalumba</i>	‘dolphin’
	<i>kisikisi</i>	‘whisper’
	<i>tektek</i>	‘chop up’
	<i>sepsep</i>	‘suck’
	<i>cadcad</i>	‘criticise’

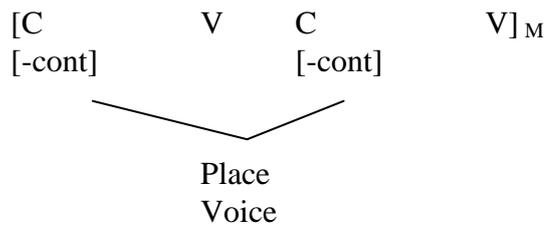
In conclusion, even though we only dealt with a subpart of Balinese expressive forms (cf. Clynes 1998 for a full discussion, including animal cries and onomatopoeia), we found that the language has semantically marked lexical elements that systematically violate markedness constraints.

Note that the choice and number of the constraints involved is a language particular issue. In Balinese, there are different, and more constraints involved than in e.g. Kambera. The Semantic Transparency constraint however seems to have relevance crosslinguistically.

3.3. Malay

In an earlier stage, the canonical Malay root C1 V C2 V (C) was subject to a constraint that where C1 and C2 are both stops *and* are homorganic, they must also agree in voicing:

(16) ‘If tautomorphemic consonants harmonise for place, they also harmonise for voice:’
 (“Place=Voice”)



This constraint excludes morphemes where homorganic voiced stops, voiceless stops, nasals or semivowels occur together as C1 and C2, e.g. **komo* and **panda*.⁴ However, quite a number of Malay words violate the Place=Voice constraint – an example is the commonly used form *bapa* ‘father’. Adelaar (1983) presents a list of these exceptional forms, compiled on the basis of Wilkinson’s 1932 *Malay-English dictionary*. Of the 108 forms presented in this paper, 51 forms are loans, and thus irrelevant in the present context. Of the 57 native forms, however, we found that 22 can be semantically classified as names of plants or animals, and 10 refer to special states, movements or sounds, names, or bad things such as evil spirits. The remaining 25 forms are a semantically and formally mixed bag, including morphologically derived forms such as reduplications and contracted forms. Thus, in Malay, a significant proportion of the words violating the constraint on tautomorphic consonant harmony belong to the expressive semantics class.

3.4. West Tarangan

West-Tarangan (Southeast Maluku, Nivens 1992) prohibits morpheme-internal consonant clusters:

(17) $*\dots CC\dots]_M$ “Avoid morpheme-internal consonant clusters”

The vast majority of the roots in this language are disyllabic (C)V(C) V(C) roots (Nivens 1992: 148, 159). We express this as the following constraint:

(18) Root = $\sigma\sigma$

Nivens (1992) reports that the forms in (19) illustrate the exceptional case when the constraint on consonant clusters is violated. The initial consonant is a liquid /r,l/ or a front glide /y/:

(19) West Tarangan exceptional consonant clusters (Nivens 1992: 148-149)⁵

C1=/r/		C1=/l/		C1=/y/	
<i>kart</i> □w	‘rat’	<i>siNalNalta</i>	‘fungus (sp.)’	<i>kaytEa</i>	‘corn’
<i>gark</i> □w	‘orphaned’	<i>ElkEy</i>	‘black cockatoo’	<i>gaykapa</i>	‘crab (sp.)’
<i>korkorba</i>	‘bat (sp.)’	<i>kaybElbElbal</i>	‘bird (sp.)’	<i>gaybu#n</i>	‘large’

<i>korbarda</i>	‘shrimp (sp.)’	<i>kildu#n</i>	‘turtle (sp.)’	<i>Naypep</i>	‘womb’
<i>g□rpan</i>	‘white hair’	<i>□p□t</i>	‘heron’	<i>gaymoy</i>	‘slow’
<i>garsī</i>	‘crab (sp.)’	<i>galsErasEran</i>	‘tuber (sp.)’	<i>gaynaka</i>	‘sago leaf’
<i>□rmol</i>	‘seasnake’	<i>almu#n</i>	‘shadow’	<i>ayN□tay</i>	‘difficult’
<i>marna#n</i>	‘turtle (sp.)’	<i>sEln□r</i>	‘mollusk (sp.)’	<i>payr□ra</i>	‘bird (sp.)’
<i>k□rNam</i>	‘parrot’	<i>t□NNum</i>	‘plant (sp.)’	<i>maylEwa</i>	‘tree (sp.)’
<i>irlow</i>	‘cassava’	<i>maljamajaman</i>	‘bed’		
<i>□rjEla</i>	‘tree’				

Though this is not observed in Nivens 1992, in the context of the present paper it is striking that out of the 30 forms 22 are plant/animal names and 2 belong to the Bad type (‘slow’ and ‘difficult’), leaving only 8 forms that are not expressive. As it is,⁶ this data supports the hypothesis that the violation of a phonological constraint, in this language, on consonant clusters, signals semantic expressiveness.

3.5. Evidence from other language families

The data from the few Austronesian languages discussed above can be supplied by data on many other languages that display a correlation between marked form and meaning. On the Symposium on Ideophones (Koeln, January 1999) similar observations were made for expressives in e.g. Estonian and Finnish, in the African languages Hausa, Zulu, Ewe, Wolaitta, Didinga and Ciluba; the Australian languages Jaminjang, Warrura, Gooniyandi, Gunin/Kwini, and in Quechua.

For example, Newman (1999) observes that Hausa ideophones use segments from normal phoneme inventory but in an unusual position. Phonotactically, Hausa ideophones are marked because they can violate the constraint that word-final consonants cannot be voiced:

- (20) *C]
- |
- voice
- Examples: *tsulum* sound of a small object falling in water
- tsigil* emphasising smallness

They can also violate the constraint that final vowels should not be long:

- (21) * V:]
- Example: *batso#* *bàtsò#* poorly made, ugly looking

And Schultze (1999) shows that the so-called ‘coverbs’ in Jaminjung (a non-Pama-Nyungan Australian language) have coupled marked formal to marked semantic properties. Among other things, the coverbs can violate phonological wellformedness constraints of Jaminjung: they may be monosyllabic rather than di/trisyllabic, they allow consonant clusters

(liquid-stop) in word-final codas, and they may contain the mid vowel /e/, otherwise not part of the phonological inventory of Jaminjung. Examples:

(22)	<i>goob</i>	‘come off’	<i>wirrb</i>	‘wipe’
	<i>yoorrg</i>	‘show’	<i>ngoojoolb</i>	‘cough’
	<i>deb</i>	‘knock down’	<i>thebberr</i>	‘pierced w. something sticking out’

Finally, in Dutch, there is has a small set of words with the onset /çr/. This onset is marked because it violates the Sonority Sequencing Generalisation, generally applicable in Dutch, which states that segments decrease in sonority towards the edges of a syllable (Booij 1995). The onset cluster /çr/ consists of a labiodental fricative and a liquid – two segments with the same degree of sonority. A standard dictionary of Dutch (*Van Dale Groot Woordenboek* 1993) reveals that all base forms with initial /çr/ semantically synchronically belong to the types Sense or Bad (or are historically derived from Middle Dutch Sense verbs). They are given in (23):

(23)		
	<i>wraak</i>	‘revenge’
	<i>wrak</i>	(i) ‘wreck’ (ii) ‘rickety, ramshackle’
	<i>wrang</i>	(i) ‘sour, acid’ (ii) ‘unpleasant, nasty’ (> <i>wringen</i> ‘squirm (mouth)’)
	<i>wrat</i>	‘wart’
	<i>wraddel</i>	(> <i>wrat</i>) ‘neck flab’
	<i>wreed</i>	‘cruel, harsh’
	<i>wrevel</i>	‘resentment, rancour, peevishness’
	<i>wrok</i>	‘resentment, grudge’
	<i>wriemelen</i>	‘wriggle, squirm, fiddle with’ (frequentative)
	<i>wrikken</i>	‘lever, prize, scull’
	<i>wringen</i>	‘wring, squirm, wrench, pinch’
	<i>wrielen</i>	‘soft cry for food of small birds’
	<i>wriggelen</i>	‘wind’ (frequentative of M. Dutch <i>wrigen</i>)
	<i>wrijven</i>	‘rub’
	<i>wroeten; wroetelen</i>	‘root (up), rout’ ; ‘rout (frequentative)’
	<i>wroeging</i>	‘remorse’
	<i>wrong</i>	‘roll, knot, wreath’
	<i>wrongel</i>	‘coagulated milk’ (> <i>wringen</i> ‘wring, wrench’)
	<i>wreef</i>	‘instep’ (> Middle Dutch <i>wrigen</i> ‘wind, tilt, grow bent’)

In sum, it appears that crosslinguistically, certain phonologically/morphologically marked forms are also semantically marked. The positive correlation between the marked form and marked meaning of these lexical items is so strong that we cannot but conclude that the relation between form and meaning in these items is not arbitrary but motivated. In the next section we will see that the motivating factor is diagrammatic iconicity.

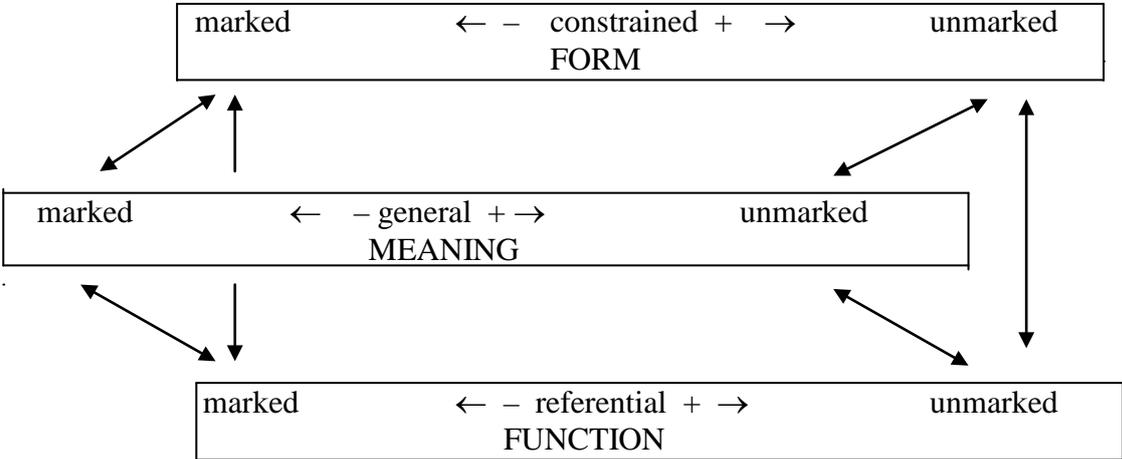
Note that the formal properties of expressives are not characteristically found in the ‘core’ lexical items of a language, though some core items may share features with expressives. Expressives are, in a sense, systematically marked forms, which implies that not all the native morphemes in the lexicon are optimal and unmarked: some are optimal while being grammatically *marked*. In other words, the optimal shape of core lexical items is

defined by a minimal violation of constraints, while the optimal shape of peripheral items is characterised by the explicit violation of at least one (and usually more) wellformedness constraints of the language.

4. Diagrammatic iconicity and the lexicon

Obviously, the alignment of marked forms and marked semantics in the data discussed above is not iconic in the strictest sense. For example, the onset cluster /çr/ in Dutch *wraak* 'revenge' or the monosyllabicity and final consonant cluster of *yoorrg* 'show' does not represent any factual resemblance between the signifier and the signified. Because this relation is not strictly iconic, we appeal to the notion of diagrammatic iconicity: the relationships among the components of the diagram are iconic – the relative position of an element in system A is matched by an element with the same relative position in system B.

In section 1 the continua for the formal, semantic and functional markedness of lexical items were introduced. They are repeated in below in a slightly different format:



← *peripheral lexical items*..... *core lexical items* →

We described the formal markedness of an item in terms of the number of constraints violated by it, its semantic markedness in terms of the more or less specific semantics it has, and its functional markedness in terms of its referentiality. In the cases discussed, the item's markedness values on the three continua are matched. In other words, the item's markedness values are iconically aligned, the iconicity being diagrammatic. The iconic relation is indicated by the two-sided arrows between the three continua.

Iconicity is thus seen as the factor that is responsible for a different patterning of core and peripheral items in the lexicon, because it is responsible for the assimilation of the markedness values of an item on the different continua. In this view on the lexicon, the distinction between core and peripheral lexical items is gradual, and the iconic matching of form, meaning and function applies to both types of items. Peripheral lexical items such as expressives occupy the space towards the left-hand side, while core lexical items pattern towards the right-hand side. But as the alignment of formal and semantic markedness is a tendency rather than a categorical property, there is some overlap of both types of lexical items. In other words, we do find core lexical items that share the marked characteristics of

expressives, and expressives with unmarked properties. In this model, however, we predict that such items will form a small minority.

We know that not everything is possible in the coining and interpretation of new lexical items, and this model helps to see why. Because the matching of function and formal/semantic markedness is preferably iconic, a newly coined referential element will never be more marked than the least constrained expressive in a language, and a new expressive element will never be more constrained than the most constrained referential element of that language. In other words, the model explains why languages prefer not to create expressives with canonical shapes, nor referential items with marked shapes. It predicts that items which are not motivated by iconicity are dispreferred in creative language.

*** Acknowledgements:** I wish to thank Geert Booij, Tucker Childs, Adrian Clynes, Claartje Levelt and Ruben van de Vijver for feedback on the data and ideas presented in this paper. The research for this paper was supported by a Fellowship of the Netherlands Academy of Arts and Sciences (KNAW).

Notes

¹ Other signs include indexes, which show a factual contingency between the sign and the object (e.g. a wind vane as the index of the wind direction) and symbols, which reflect conventional contiguity (Peirce 1965, Battistella 1990).

² The unconstrainedness of marked forms implies the unmarkedness values of Battistella's (1990): crosslinguistic 'optimality' and formal 'simplicity'.

³ I.e., properties are more conceptually complex, and hence more marked, the less closely and clearly they reflect attributes of prototypical or experientially more basic categories (Battistella 1990: 27, 41-44).

⁴ The constraint does not apply to (i) combinations of initial /t/ and medial /d,n/, and (ii) initial /d,n/ with medial /t/. This suggests that a different place feature may be involved, and Adelaar (1983:65) therefore assumes that /t/ is dental, while /d,n/ are alveolar.

⁵ Note that the phonotactic status of the initial consonant in the third column, /y/, differs from that of the liquids in the first two columns. Firstly, because its consonantal status is unclear, and secondly, because it appears to be part of a (fossilised, lexicalised) prefix *Cay*. As Nivens does not contain further information on this, I follow his analysis and assume that /y/ here is a consonant in a cluster.

⁶ From Nivens (1992) it is unclear whether the list of 30 items is exhaustive, or indeed representative for words with a consonant cluster, nor does it contain observations on the semantics of these forms.

References

- Antilla, Raimo. 1976. *Meaning and structure of Finnish descriptive vocabulary*. [Texas Linguistic Forum 5.] Texas.
- Adelaar, Sander. 1983. "Malay consonant harmony: an internal reconstruction". In James T. Collins (ed.), *Studies in Malay dialects Part I*. [NUSA Volume 16]. Jakarta: Universitas Atma Jaya.

-
- Battistella, Edwin L. 1990. *Markedness: The Evaluative Superstructure of Language*. Albany: State University of New York Press.
- Booij, Geert. 1995. *The phonology of Dutch*. Oxford: Clarendon Press.
- Booij, Geert. 1998. "Morpheme structure constraints and the phonotactics of Dutch". In Harry van der Hulst and Nancy Ritter (eds.), *The syllable: views and facts*. Berlin: Mouton de Gruyter.
- Childs, Tucker G. 1994. African ideophones. In Hinton *et.al.* (eds.), 178-199.
- Clynes, Adrian. 1995. *Topics in the Phonology and Morphology of Balinese [based on the dialect of Singaraja, North Bali]*. PhD dissertation, Australian National University, Canberra, ACT.
- Clynes, Adrian. 1998. 'The emergence of the marked: interactions of phonology and semantics in Balinese'. Ms., University of Brunei Darussalam.
- Elders, Stefan. 1999. "Defining ideophones in Mundang". Paper presented at the International Symposium on Ideophones, Köln, January 25-27, 1999.
- Haiman, John. 1994. 'Iconicity'. In: R.E. Asher and J.M.Y. Simpson (eds.) *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon Press.
- Hamano, Shoko. 1998. *The sound-symbolic system of Japanese*. Stanford, California: CSLI Publications.
- Hinton, Leanne, Johanna Nichols, and John J. Ohala (eds.). 1994. *Sound symbolism*. Cambridge: Cambridge University Press.
- Jong, Nicky de. 1999. "The ideophone in Didinga". Paper presented at the International Symposium on Ideophones, Köln, January 25-27 1999.
- Joseph, Brian D. 1994. "Modern Greek ts: beyond sound symbolism". In Hinton *et.al.* (eds.), 222-236.
- Joseph, Brian D. 1997. "On the linguistics of marginality: the centrality of the periphery". In Kora Singer, Randall Eggert and Gregory Anderson (eds.), *CLS 33: Papers from the main session*. Chicago: Chicago Linguistic Society, 197-213.
- Kita, Sotaro. 1997. "Two-dimensional semantic analysis of Japanese mimetics". *Linguistics* 35. 379-415.
- Kita, Sotaro. 1999. Kita, Sotaro. 1997. "Two-dimensional semantic analysis of Japanese *giongo/gitaigo* (sound mimetics/manner mimetics)". Paper presented at the International Symposium on Ideophones, Köln, January 25-27 1999.
- Klamer, Marian. 1998. *A grammar of Kambera*. Berlin/New York: Mouton de Gruyter.
- Klamer, Marian. 1999a. "How report verbs become quote markers and complementisers". To appear in *Lingua*.
- Klamer, Marian. 1999b. "Ungrammatical but optimal: evidence from Austronesian expressives". Paper presented at the Austronesian Formal Linguistics Association (AFLA) VI, University of Toronto, April 16-18, 1999
- Klinken, Catharina van. 1997. *A grammar of the Fehan dialect of Tetun*. PhD diss., Australian National University, Canberra, ACT.
- Mithun, Marianne. 1982. The synchronic and diachronic behaviour of plops, squeaks, croaks, sighs and moans. *International Journal of American Linguistics* 48, 1: 49-58.
- Newman, Paul. 1999. "Are ideophones really as weird and extra-systematic as linguists make them out to be?". Paper presented at the International Symposium on Ideophones, Köln, January 25-27, 1999.
- Nivens, Richard. 1992. 'A lexical phonology of West Tarangan.' In: Donald A. Burquest and Wyn D. Laidig, *Phonological studies in four languages of Maluku*. Dallas: Summer Institute of Linguistics and The University of Texas Publications in Linguistics.

-
- Onvlee, Louis. 1984. *Kambaraas (Oost-Soembaas)-Nederlands Woordenboek*. Dordrecht: Foris.
- Peirce, Charles S. 1965. *Collected papers of Charles Sanders Peirce*. Cambridge: Harvard University Press.
- Schultze-Berndt, Eva. 1999. Ideophone-like characteristics of ‘coverbs’ in Jaminjung (Australia). Paper presented at the International Symposium on Ideophones, Köln, January 25-27, 1999.
- Uhlenbeck, E. M. 1950. “The structure of the Javanese morpheme.” *Lingua* 2: 239-270.