Explaining some Structural and Semantic Asymmetries in Morphological Typology

Marian Klamer
Leiden University
marian.klamer@let.leidenuniv.nl

1.   Introduction

This paper discusses two types of asymmetries in the typology of words. The first asymmetry concerns the morphological structure of words, the second type concerns their lexical-semantic properties. For both types of asymmetries I first present some empirical evidence, followed by a proposal on how the asymmetries can be explained.

My basic argument will be that the observed structural and semantic asymmetries are two sides of the same coin, and that they can be explained by referring to two quite general well-formedness constraints: Semantic Transparency and Structural Contrast, and one universal semantic principle on form-meaning relationships: Iconicity.

2.   Evidence for the Structural Asymmetries

In this section I present some empirical evidence for the following three typological asymmetries in the morphological make-up of words: prefixing/suffixing is more common than circumfixing[1] (section 2.1); empty morphemes are always a minority in a language's morphology (section 2.2); and compounding is more common than conversion (section 2.3).

2.1   Prefixing/Suffixing Is more Common than Circumfixing

At least since Greenberg 1963, it has often been observed that pre/suffixes are more frequent than circumfixes, both within and across languages. Since they are typologically less marked than circumfixes, the following implicational hierarchy applies: a language with affixes will always have a pre/suffix, but not necessarily a circumfix. When a language has a circumfix, it will at least have one pre/suffix as well.

An example is Dutch, which has many productive and unproductive prefixes and suffixes (cf. Booij 2002), but only one (clear[2]) circumfix *ge__te*, which functions to derive collective nouns: *berg* 'mountain' > *ge-berg-te* 'mountain range'.

Kambera (an Austronesian language spoken on the island of Sumba in Eastern Indonesia; Klamer 1998) has one productive and many unproductive prefixes, as well as several suffixes, but only one circumfix *ka__k*. The circumfix derives verbs from ideophonic roots denoting sounds, motions and sights: *reu* 'sound of people talking' >

---

[1] Cf. Greenberg (1963: 92): If a language has discontinuous affixes, it always has either prefixing or suffixing or both.

[2] Dutch perfect participles may be formed by what looks like a circumfix: prefix *ge__t / g__d* (the voice of the final stop agrees with the voice of the final stem consonant), though various analyses of this affix are possible, see Booij (2002), section 2.4.3.

ka-reu-k 'to talk', *ndiku* 'jerking motion' > *ka-ndiku-k* 'to jerk', *bila* 'light, brightness' > *ka-bila-k* 'to emit light, be bright' (Klamer 1998: 245–247; 2001).

The exceptional status of circumfixes is also evident from the fact that many linguists would argue that circumfixes can (or should be) reduced to a combination of suffixing and prefixing (cf. Spencer 1991: 13), i.e. that they have a 'derived' status in the synchronic morphology of a language. In any case, it is remarkable that the two parts of a circumfix are often formally identical to affixes with other functions. For example, the prefixing part of the Dutch collective noun circumfix *ge__te* is formally identical to the productive nominalising prefix *ge-*, used as in *schrijf* 'write' > *ge-schrijf* 'writing', while its suffix *-te* is formally identical to the unproductive suffix *-te* that derives de-adjectival nouns (as in *leeg* 'empty' > *leeg-te* 'emptyness'.) Observe also that both affixes are nominalising, just like the circumfix is. In other words, either part of *ge__te* is formally and functionally related to another affix, and their combination might be analysed as a derived structure in the synchronic morphology of Dutch.

Similar observations can be made about the Kambera circumfix, though here only the prefixing part is used elsewhere in the morphology as an unproductive prefix: *mboka* 'be fat' > *ka-mboka* 'look healthy, prosperous', *hilu* 'language' > *ka-hilu* 'ear', *beli* 'go back' > *ka-beli* 'turn around; return' (Klamer 1998: 254).

We conclude that circumfixes are less frequent and less common than pre/suffixes, and may often be analysed as derived, complex morphological units.

2.2 Empty Morphemes are Always a Minority in a Language's Morphology

Morphemes such as the *cran* of *cranberry* or *ceive* in *conceive* and *perceive* are forms with no clear meaning of their own and they are not productive. Though we find such forms in probably every language, it is generally agreed upon that morphemes without meaning would never constitute the majority of a language's morphology – they are always a minority class. Often they have special characteristics, for example because they refer to specific semantic domains (e.g. fruits), or because they are part of the non-native lexicon.

In other words, we do not expect to find a language whose morphology only, or mainly, consists of empty (cranberry) morphemes – if it has any of such morphemes, there will also be a class of productive, meaningful morphemes, and this class will be larger.

Similarly, in a language that employs reduplication, we often find empty or meaningless reduplicative elements. For example, the lexicalised relicts of reduplication processes that were productive in the past. Yet, we do not expect to find a language with only empty reduplicative elements. In other words, the existence of empty reduplicative elements implies the existence of productive, and meaningful, reduplicative elements.

2.3 Compounding Is Typologically more Common than Conversion

Compounding is a word-formation process that is distinct from other derivational processes, because it combines two lexemes into one new one while there is no bound morpheme involved in the process. Conversion (also referred to as zero-derivation) resembles compounding in that it is also a morphological process that does not involve any bound morphology (cf. Aronoff 1994: 15–16).

As a first step in the typological comparison of these processes, I would like to address the question which of the two is more commonly used in a language that has both of them, such as Dutch. In Dutch, the process of compounding goes in various directions: it is possible to productively derive nominal, adjectival, and numeral compounds on various types of bases (verbal compounds exist but are not productive). The base of a compound can be either a morphologically simple or a complex form (e.g. compounds can be derived from derived compounds, [*fiets-band*] [*ventiel-* [*dop-je*]] 'small lid of a bicycle tyre valve'). In contrast, the direction of productive conversion is quite limited: N>V is the productive pattern (*zon* 'sun' > 'to sunbathe'), while V > N (*kook* 'cook' > 'boiling'), A > N (*gek* 'mad' > madman), A > V (*wit* 'white' > 'to whiten') are marginally productive, or have a restricted domain of application. There is no conversion of nouns or verbs into adjectives. The base for a conversion is preferably morphologically simple – it is not easy to find derivationally complex nouns that feed conversion. In other words, Dutch conversion is subject to a lot more structural restrictions than compounding is. In addition, the semantics of compounds in relation to their morphological structure is also more transparent for compounds than for converted forms.

In Kambera, compounding is a productive process, deriving both nouns and verbs (Klamer 1998: 40, 58, 115, 117) but conversion does not exist. Similarly, in standard Indonesian, compounding derives both nominal and verbal forms (Sneddon 1996: 23–25), but conversion is not mentioned as a derivational process in Indonesian reference grammars or textbooks. Note however, that in Kambera as well as in (substandard) Indonesian we often find words with no nominal or verbal affixes which are used as so-called 'multifunctional' items: lexemes without a clear lexical category/word class that function in both verbal and nominal contexts. For example, in Kambera *tanda* can be used as a noun 'sign, symbol', as well as a verb 'to know, recognise'. Multifunctional items are distinct from words undergoing conversion, because the lexical category of their base form is unclear. In conversion, the lexical category of the base can usually be established, e.g. on semantic grounds.

In sum, while neither compounding nor conversion involves the addition of bound morphological material, we formulate the hypothesis that, if a language has both, compounding is more common than conversion. Why would this asymmetry exist?

3. Explanation of the Structural Asymmetries

In this section I present a proposal on how the three typological asymmetries discussed above might be explained. The basic idea behind my explanation is that structurally simple forms are cross-linguistically more common than complex ones. In this view, prefixes would then be structurally simpler than circumfixes, meaningful morphemes simpler than empty ones, and compounding simpler than conversion.

Why would this be the case? In what sense are circumfixes, empty morphemes and conversion structurally complex? If we envisage a linguistic system as a set of constraints on the wellformedness of utterances, we may say that those linguistic items that obey the constraints are structurally less complex than those which violate the constraints. Put in a different way, structurally complex items violate more wellformedness constraints than simple items do.

In the context of the present discussion, this implies that prefixes and suffixes are 'better behaved' than circumfixes or empty morphemes, and that compounding is structurally less complex than conversion, because the latter violate some wellformedness constraints that are obeyed by the former.

The question is then: What do these constraints look like? They must be of a quite general and abstract type, because they apply to formally quite distinct phenomena. I propose that the first of these constraints goes back to the age-old insight that linguistic signs should be semantically transparent:

(1)   *Semantic Transparency*
        'Match form and meaning one-to-one':   meaning   A
                                           |
                                     form   X

This (classic) constraint assumes that the 'ideal' linguistic system is one where every form corresponds to one meaning only, and every meaning has a single formal expression. Of course, deviations from this ideal exist, but these are considered marked, minority constructs, that are historically less stable, and less favoured in e.g. language acquisition. In principle, the constraint applies to all linguistic modules (e.g., syntax, morphology, phonology) but the discussion here is limited to its application on the word level.[3]

On the word level, we observe that the constraint is not violated by a meaningful pre/suffix (3a) or a compound (3b), while it is violated by circumfixes (3c): one meaning is expressed through two forms, meaningless morphemes (3d): a form with no meaning attached to it,[4] and by conversion (3e).

(2) a.   affix: b.   compound: c.   circumfix:
           A   AB   * A
           | | ╱ ╲
           X   XY   X   Y

      d.   meaningless morpheme: e.   conversion:
           * __   * A
             | |
            X   __

In conversion (3e) we add a meaning or function A (e.g. a category change) but this has no overt formal expression, which is a violation of the Semantic Transparency constraint. On the other hand, in compounding (3b), we combine two form-meaning pairs into one (new) form-meaning pair, so compounding does not violate the

---

[3] In syntax, this constraint would for example imply that a difference in word order is never truly optional but always relates to a difference in meaning: since there are two distinct forms, ideally each of them must have its own meaning (see e.g. Williams 1997).

[4] Cf. Croft (2003: 104), who notes that it is typologically rare to find one meaning expressed through two or more forms (as in (3c)) or forms with no meaning attached to it (as in (3d)). He adds that such rare configurations are historically unstable, referring to the loss of the double marking of negation in the history of French (one meaning-two forms becomes one meaning-one form).

constraint.[5] In general, then, compounding is structurally simpler than conversion, since it conforms more to the ideal of one form-one meaning matching.

We conclude that the cross-linguistically less common morphological patterns can be considered to be more complex forms because they violate the constraint on Semantic Transparency: they are structurally less 'optimal' that the forms that do comply to the constraint.

Now, if the typological asymmetries observed in section 2 are indeed correct, they indicate that cross-linguistically, morphologically simple structures are preferred over complex ones. Why would this be so? I suggest that the explanation might be sought outside language itself, in the language user. Most language users are both speaker and hearer. It is generally believed that economy considerations play a role in structuring linguistic communication. As speakers, we strive to be economical in speech production, so that we say as much as possible with as little effort as possible, i.e. we reduce formal contrasts (cf. (4a). As hearers, on the other hand, we want to be economical in the processing of what we hear, so that utterances must be as distinct as possible. As hearers, then, we prefer reduced formal identity (cf. 4b). In other words, 'economy' concerns of hearer and speaker are the motivation of a second family of structural constraints, the constraints on structural contrasts between linguistic elements:

(3)   *Constraints on Structural Contrast between linguistic elements*
        a.   "No formal contrast" (i.e., "Favour increased similarity")
                (Economy in production; speaker's perspective)

        b.   "No formal identity" (i.e., "Favour increased dissimilarity")
                (Economy in processing; hearer's perspective)

Constraints on structural contrasts between linguistic elements are well-known in phonology. Examples of constraints on formal contrast (4a) are constraints on certain complex segments or complex phonotactics, and examples of constraints on formal identity (4b) are the constraints on similar homorganic consonant pairs such as the OCP (Pierrehumbert 1993).

In morphology, an example of a constraint on formal contrast (> "Favour increased similarity") would be one that penalises morphologically complex structures: an isolating language where every single linguistic unit represents a single meaning unit would then be the ideal.[6] An example of a morphological constraint on formal identity would be a constraint on homophonous morphemes. If several distinct functions are expressed by one single form, processing becomes increasingly difficult; so our preference is to link different meanings to different morphemes. These constraints may be used to explain observed asymmetries, but they cannot be categorical: it is obvious that not all languages are isolating, and of course, homophonous morphemes do exist. I come back to this in section 7.

---

[5] Here we refer to endocentric compounds and not to exocentric ones: endocentric compounds do not violate the constraint since their interpretation is a sum of their parts, while the interpretation of exocentric compounds is much less regular.
[6] Note, however, that in such a language the structural contrast between individual *lexemes/words* is maximal; so minimal morphological complexity does not lead to minimal structural contrast in the overall make-up of a language.

4. Evidence for the Semantic Asymmetry

After discussing structural asymmetries in section 2 and 3, I now turn to an asymmetry that relates to the semantics of certain types of words. In the present section I present evidence that certain types of words show a remarkable semantic pattern. The types of words under consideration are not random: I only look at words that have a "complex" morphological make-up in the sense discussed in the previous section. That is, we look at the semantics of words which violate the structural constraints discussed above. The words under consideration are all from Austronesian languages: Kambera, Ilocano (Philippines, Rubino 2001), and Kéo (Flores, Baird 2002). We look at the semantics of words with a meaningless prefix in Kambera (4.1), words with a circumfix in Kambera (4.2), words with a meaningless reduplication in Ilocano (4.3) and lexicalised compounds in Kéo (4.4).

Since morphemes, like lexemes, are generally arbitrary signs (i.e. onomatopoeic morphemes hardly exist), we do not expect to find a direct correlation between the phonetic make-up of a morpheme and its meaning, or between its position (pre/suffix) and its meaning. For example, there is no a priori reason why a verbalizing affix should be a prefix rather than e.g. a circumfix, or why it would have the particular phonetic make-up it has (e.g. why a nominalising prefix in Dutch has the shape *ge-* rather than *li-, pa-* or any other string of sounds). In general, then, we say that the relation between the shape of a morpheme and its meaning is arbitrary.

I mention this very obvious generalisation here because in the cases discussed below we *do* find a direct correlation between the shape of the words and their semantics. We will see that these words have a "complex" morphological structure (in the sense of section 3) and tend to have the semantics of a particular, circumscribed, semantic domain: the domain of "expressives". In other words: complex forms link to expressive semantics.

'Expressive' items belong to one of the semantic types of "Sense words", "Names" or "Bad words", as explained in Table 1 (for additional motivation, see the Appendix and Klamer 2002).

Table 1. The semantic types of expressive items

TYPE	EXPLANATION	EXAMPLES
Sense words	Words denoting sense impressions: sound, touch, taste, smell, feeling, emotion and sight, incl. movements of the body and/or body parts.	English: <u>tweet, blob, burp, bob</u>
Names	Personal or place names, nicknames, epithets, terms of endearment, names for plants and animals.	English: <u>Bob, baboon, moron</u>
Bad words	Taboo words, and lexical items with negative connotations or items that refer to undesirable states.	English: <u>boob(s), tit(s)</u>

Expressive items are conceptually more complex, and more specific (less general) than common, prototypical referential lexemes. For example, *jabber* is

semantically more specialised, and conceptually more complex than *talk*: since *jabber* is a special kind of talking, *jabber* has at least one feature more than *talk*: an evaluative, subjective, and/or descriptive semantic feature. Expressive items are used less frequently than lexical items with more general meanings because they refer to very specific events or referents, (hence) they are not usually phonologically reduced, less easily accessible on-line, and never subject to grammaticalisation (cf. Hopper and Traugott 1993: 87, Slobin 2001: 432/3).

Having established what it means to say that an item has an 'expressive' semantics, let us now return to the semantic asymmetry that can be observed in the lexicon of a number of Austronesian languages.

4.1   The Semantics of Words with a Meaningless Prefix in Kambera

Kambera has a limited number of formally derived words with the prefix *la-*. They are listed in (5). With one exception, none of them has a root form that is still used independently. The prefix *la* has no independent meaning and does not occur elsewhere in Kambera morphology. The argument to analyse the words in (5) as morphologically complex forms is purely formal (cf. Klamer 1998 for further motivation).

(5)   Kambera words with the 'empty' prefix *la-*

la-lei	be a husband'	*la-mbungur*	'flower spec.' *Datura factuosa*
la-ngora	'wipe off'	*la-mboya*	'name of medicinal plant'
la-wihir	'turn one's back, give way to X'	*la-wungu*	'tree sp. with hard wood'
la-mihi	'clean away X'	*la-wina*	'bean sp.' *Cajanus Cajan*
la-manga	'be weak'	*la-nggapa*	1. 'tree with thin bark' 2. 'very thin'
la-mbiri	'look sleepy'	*la-ngira*	'tree sp. used for canoes'
la-muji	'suck'	*la-ngaha*	'tree sp.' *Barringtonia asiatica*
la-nggori	'burp'	*la-yia*	1. 'ginger plant' 2. 'brother in law'
la-ngidip	'hickup, 'gasp'	*la-hona*	'red onion'
la-ngudu	'be in a heap'	*la-bawa*	'white onion'
la-nggeha	'be thin'	*la-mbàku*	'civet cat'
la-wújur	'with bended back'	*la-wora*	'iguana'
la-nggudu	'tie w. feet together'	*la-nggudu*	'tuberous plant sp.' *Toca palmata*
la-mbonga	'deep large hole'	*la-ngàdi*	'type of coral'
la-mbaru	'centipede'	*la-ngiha*	'gums'
la-pàpu	'ulcer in armpit/groin'		

When we consider the semantics of these *la*-derivations, we observe that they are both verbs and nouns. The nouns are mostly plant or animal names (cf. the right column), whereas a sizable number of the verbal forms denote a position or state of the body, or movements/sounds that are related to the mouth. In other words, the nouns are Names, and quite a number of verbs are Sense words. The large majority of *la*-derivations can

thus be characterised as semantically "expressive" in the sense defined above. There is thus a remarkable semantic asymmetry to be observed in the class of words with the meaningless prefix *la-*.

4.2   The Semantics of Words with a Circumfix in Kambera

As mentioned in section 2.1, Kambera has one circumfix, *ka__k*, which derives verbs denoting sounds, motions and sights from ideophonic roots:

(6)   *mbùtu*   'thud' (sound)   >   *ka-mbùtu-k*   '(fall) with a thud'
        jila   'flash' (sight)   >   *ka-jila-k*   'gleam; flash (as lightning)'

Morphologically, words derived with *ka__k* are special because they are the only Kambera forms that are derived by circumfixation. (In addition, they also have an exceptional phonotactic make-up, as well as special syntactic properties, not discussed here.) Both the root forms and the derived verbs denote sounds, motions and sights; and can thus all be classified as "Sense" words.

4.3   The Semantics of Words with Meaningless Reduplication in Ilocano

Ilocano, spoken in the Philippines and described by Rubino (1999, 2001) has a very elaborate morphology, including several morphemes and morphological processes that are especially related to sounds, i.e. derive onomatopoeic words.

Rubino (2001) presents an overview of the onomatopoeic morphology. His overview also contains a set of 45 lexical items of the shape $C_1V_1.C_2V.C_1V_1C_2$, for example *bu.ki.buk* 'scatter, overturn'. Structurally, the roots in this set are made up of two identical CVC sequences separated by a vowel, resulting in a tri-syllabic lexical item, e.g. *bug-a-bug* 'to be mixed (varieties of rice)', *bas-i-bas* 'hurl a long object'. Rubino analyses the derived items as "roots" (2001: 317), which I take to imply that there is no meaningful root unit *bug/bugi* or *bas/basi* etc. in Ilocano morphology. In other words, formally these items are reduplications, but the base of the reduplication is non-existent. Rubino further remarks that most of the words in this set are "no longer" onomatopoeic.

(7)   *The semantics of words with a meaningless reduplication in Ilocano*

reduplication	non-existent base	meaning
yaba-yap	**yab(a)*	'flap (flags), flutter'
ngasa-ngas	**ngas(a)*	'wear out (shoes); suffer injury'
pali-pal	**pal(i)*	'black magic'
wisa-wis	**wis(a)*	'fishing tackle'
guyu-guy	**guy(u)*	'suggest; convince'
bala-bal	**bal(a)*	'scarf, muffler; wrap snugly'
rangi-rang	**rang(i)*	'dry, parched land'
wida-wid	**wid(a)*	'swing the arms when walking'
nuru-nur	**nur(u)*	'erode from water contact'

The translations suggest that most of them are not related to sounds, i.e. they are not onomatopoeic. In the list of (7), I would classify *yabayap* and *widawid* as "Sense" words, while *ngasangas* and *palipal* are "Bad" words, words with negative connotations. In a similar way, of the 45 examples given in the paper we can classify 16 as a "Sense" or a "Bad" word, i.e. about one third of the items are semantically expressive. This is a remarkable semantic asymmetry, considering the wide semantic range of the words (from 'fishing tackle' to 'black magic' to 'suggest'...!). If the potential semantic range of the given forms is so wide, why would one third of them cluster in the particular, rather circumscribed, semantic domain of expressives?

4.4   The Semantics of Compounds in Kéo

A similar semantic asymmetry is found in a particular set of morphologically complex words in Kéo, a language spoken in Central Flores in Eastern Indonesia, and described by Baird (2002).

Kéo is an isolating language – it has no inflectional morphology and no productive morphological derivation. The only sub-lexical element in the language is the numeral clitic *ha-* 'one'. Kéo has some lexicalised reduplicated forms and a limited number of compounds. Many of the Kéo compounds are semantically opaque, and Baird (2002: 182) suggests that they are lexicalised inheritances from ritual, parallel speech. The compounds attested by Baird are all listed in the grammar. There are 47 compounds listed, of which 20 items semantically belong to the class of Sense, Name or Bad words. For example (cf. Baird 2002: 171–182) (a question mark as gloss indicates lack of independent meaning of a word):

(8)   *Kéo lexicalised compounds*

da'é-dondo	'space'	*méké-suné*	'flu'
place-place		cough-sniffle	
mutu-tiwo	'gathering'	*pémba-jawa*	'sit cross-legged'
gathering-meeting		hold on lap-corn	
dera-kiri	'afternoon'	*mundé-mi*	'place name Mundemi'
sun-slant		citrus.fruit-sweet	
topo-dhupa	'machete sheath'	*méso-mélo*	'sit restlessly'
machete-?		move-?	

Of the items given in (8), those in the left column are not expressive, while I would interpret those in the column on the right as expressive (from top to bottom: Bad, Sense, Name, Sense). Classifying the 47 items in Baird 2002 in a similar way, 20 turned out to be expressive. Considering the semantic range covered by the compounds ('machete sheath' to 'afternoon' to 'flu'), it is again a remarkable asymmetry that 42,5 % of the items cluster in the semantic domain of expressives.

In sum, in this section I have presented four case studies from three Austronesian languages which illustrate that certain morphologically complex forms show a tendency to associate with expressive semantics. In the next section I suggest an explanation for this semantic asymmetry.

5.   Explanation of the Semantic Asymmetry

The semantic asymmetry observed in the classes of words above can be explained when we consider the link between the meaning and the structural complexity of the items in question. Recall that expressive items (Sense, Name and Bad words) are assumed to be semantically or conceptually more complex than common, referential items: an expressive word has one or more evaluative, subjective, and/or descriptive semantic feature(s), and is more specific than a referential lexeme.

Recall also that the circumfix *ka__k* in Kambera is structurally complex because it violates the Semantic Transparancy constraint, as does the empty prefix *la-* in this language. When derivations with *ka__k* and *la-* are semantically expressive, we observe a match between the structural complexity of these items and their semantic complexity, a matching that might be called an 'iconic' matching of form and function.

Turning now to the Ilocano words with an empty reduplicative element, we observed that these items are structurally complex for two reasons. Firstly, since they contain a reduplicative syllable they are longer than common roots: Ilocano roots usually have only two syllables – not three. Secondly, they contain a meaningless (reduplicative) morpheme, and therefore violate Semantic Transparancy. We observed the asymmetry that one third of the items are semantically expressive. This asymmetry can be explained when we assume that these cases too show a preference for an iconic matching of complex form and complex semantics.

Finally, how to account for the skewed semantics of the Kéo compounds?

In the isolating language Kéo, there is obviously a very strong preference for morphologically simple forms, since – apart from the set of compounds discussed here and some lexicalised reduplications – morphologically complex forms do not occur in this language. In other words, the constraint on formal contrast in (3a), which disallows morphologically complex structures, is generally obeyed in Kéo. The compounds are exceptional because they are morphologically complex forms, and they show a tendency to match their structural complexity with the complex semantics of expressiveness.

In sum, the fact that a significant part of the complex forms discussed here are semantically expressive is not a coincidence, but something that can be explained: in many cases, the general principle of Iconicity seems to be applied, and a complex form is matched with a complex semantics.

Put differently, expressive words constitute a subclass in the lexicon which shows a non-arbitrary connection between form and meaning, and Iconicity is the principle steering the lexical semantic asymmetries observed. Note that this is a tendency observed for certain types of words; it does not apply categorically in all languages for all words: there are many morphologically complex items that are not expressive, and there are also many simple words with a complex semantics.

6.   Conclusions and Discussion

In this paper I presented a number of structural and semantic asymmetries at the word level; some of them obvious, others perhaps more controversial.

I argued that pre/suffixes are crosslinguistically more common than infixes, that meaningless affixes are always a minority in a language, and that compounding is typologically less marked than conversion. I then explained the skewed distribution of

these morphological patterns as a cross-linguistic preference for simple morphological structures over complex ones. This preference can be expressed by formulating structural constraints on the wellformedness of linguistic forms.

I suggested that the relevant constraints are (a) constraints concerned with the one-to-one linking of form and meaning (maintaining Semantic Transparancy), and (b) constraints on structural identity and structural contrast.

I then demonstrated a striking asymmetry in the semantics of certain classes of morphologically complex words in the Austronesian languages Kambera, Ilocano and Kéo. The four types of morphologically derived words all showed a strong preference for an expressive semantics. If it is correct to assume that expressive items are semantically more complex that common referential items, we can explain this semantic skewing as the outcome of the application of the universal principle of Iconicity: link a complex form to a complex meaning.

As a consequence, we understand why certain types of morphological processes are typologically less common than others, and why cross-linguistically, expressives appear to have a preference for complex structures.

The explanations proposed here are not new: economic and iconic motivations for certain linguistic forms or patterns have been proposed in various linguistic research traditions (both typological and generative), as well as for various sub-disciplines of linguistics, including morphology. For morphologists, the ideas presented here may sound similar to those presented as the theory of Natural Morphology (Mayerthaler 1981, Dressler 1985, 1987, and references cited there). Natural Morphology is a theory of what constitutes a natural, or unmarked morphological system, and how we can predict and explain deviations from that system. In this theory, the most 'natural' type of morphology is fully transparent: every morpheme has one form and one meaning, and every meaning corresponds to only one form (the 'bi-uniqueness' principle, e.g. Dressler 1987: 111 v.v.). 'Bi-uniqueness' is an explication of the intuition that has always been implicit in the classical notion of the morpheme as the minimal form-meaning unit. Natural morphology regards deviations from the most natural, transparent state as unnatural or marked, and the assumption is that cross-linguistic patterns, historical change, language acquisition, speech errors and language disorders show a statistical tendency to prefer the natural, unmarked state to the unnatural, marked one.

The present paper agrees with natural morphologists such as Dressler (1987) in that typological asymmetries in morphology can and should be explained with very simple, general constraints on the linking of form and function.

The constraints should, however, not be used to characterise possible (and impossible) morphological systems, but rather to calculate which systems are more probable than others (cf. Croft 2003: 283).

It is clear that the constraints discussed here cannot be categorical. For example, the Semantic Transparency constraint would exclude the existence of polysemous and synonymous affixes, as well as allomorphy, yet we find such morphemes in many languages. The Dutch diminutive is but one illustration of allomorphy, where one meaning is expressed through five forms: *koning-kje* 'king-DIM', *riem-pje* 'belt-DIM', *huis-je* 'house-DIM', *oven-tje* 'oven-DIM', *tong-etje* 'tongue-DIM'. An example of synonymy in English affixes (where various forms have one common meaning) is the fact that *-ship, -dom, -hood* all indicate a 'state or quality of being', compare *friendship, serfdom, motherhood*. And an illustration of polysemy in an English affix (where one

form has various meanings) is the suffix *-ist.* Canonically this suffix means 'one who does X', as in *rapist* 'one who rapes', but it also appears in words like *racist,* and *sexist* where it means something like 'one who is prejudiced against a group'. The Semantic Transparency constraint does not exclude such phenomena, but is a way to express that such items are structurally more complex than morphemes with a one-to-one mapping of meaning and form.

Since there are various motivations for linguistic structure, both functional and formal, and since these motivations relate to distinct linguistic modules (phonetics, phonology, morphology, syntax, semantics), and/or to language external factors (sociolinguistic, psycholinguistic), etc., the motivations for certain linguistic structures compete with each other in many ways. It is impossible to predict the outcome of this competition for a language; indeed, it is usually assumed to be arbitrary (cf. Croft 1995: 504–509). In other words, the synchronic grammar of a particular language always involves a lot of arbitrariness, and not everything in language is explainable in terms of a completely deterministic set of formal or functional principles.[7]

Since morphological systems are the outcome of many different, competing synchronic and diachronic forces, historical developments may lead to a complex, 'unnatural' or 'marked' situation. For example, the occurrence of clitics within other morphemes is crosslinguistically very unusual, but such 'endo-clitics' are attested in Udi (Harris 2002) as the outcome of a unique combination of particular historical changes and certain morpho-syntactic features in this language. In other words, it cannot be maintained that diachronic change, language acquisition, speech errors etc. always strive towards more natural, less marked, simpler structures.

Yet, strikingly asymmetrical crosslinguistic patterns do exist, and call for explanation. In this paper I presented evidence for some of such asymmetrical patterns in morphological structure as well as in semantics. I argued that their skewed patterning might be explained with some very general constraints and principles. These constraints may also be used to calculate which morphological types are more probable than others.

References

Aronoff, M. (1994), *Morphology by itself,* Cambridge (Mass.), MIT Press.

Baird, L. (2002), *A grammar of Keo,* Ph.D. dissertation, Canberra, Australian National University.

Booij, G. (2002), *The Morphology of Dutch*, Oxford University Press.

Clynes, A. (1995), *Topics in the phonology and morphosyntax of Balinese,* Ph.D. dissertation, Canberra, Australian National University.

Clynes, A. (1998), *The emergence of the marked: interactions of phonology and semantics in Balinese*, University of Brunei Darussalam, ms.

Croft, W. (1995), *Autonomy and functionalist linguistics*, "Language" 71, 490–532.

Croft, W. (2003), *Typology and universals*, [2nd edition], Cambridge, Cambridge University Press.

Dressler, W.U. (1985), *Morphonology*, Ann Arbor, Karoma Press.

Dressler, W.U. (1987), *Word formation as part of natural morphology*, in W.U. Dressler, W. Mayerthaler, O. Panagl & W.U. Wurzel (eds.), *Leitmotifs in Natural Morphology,* Amsterdam, John Benjamins, 99–126.

---

[7] If it were, all languages would be alike, internally invariant and no languages would change (Croft 2003: 282).

Greenberg, J.H. (1963), *Some universals of grammar, with particular reference to the order of meaningful elements*, in J.H. Greenberg (ed.), *Universals of language*, Cambridge (Mass.), MIT Press, 73–113.

Humboldt, W. von (1836), *Uber die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwickelung des Menschengeschlechts,* Berlin, Dummler.

Harris, A.C. (2002), *Endoclitics and the origins of Udi morphosyntax,* Oxford, Oxford University Press.

Hopper, P. & E. Traugott (1993), *Grammaticalization*, Cambridge, Cambridge University Press.

Klamer, M. (1998), *A grammar of Kambera,* Berlin/New York, Mouton de Gruyter.

Klamer, M. (2001), *Expressives and iconicity in the lexicon*, in F.K. Erhard Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, Amsterdam, John Benjamins, 165–181.

Klamer, M. (2002), *Semantically motivated lexical patterns: a study of Dutch and Kambera expressives*, "Language" 78.2, 258–286.

Klinken, C. van (1999), *A grammar of the Fehan dialect of Tetun,* Canberra, Pacific Ling. 155.

Mayerthaler, W. (1981), *Morphologische Natürlichkeit,* Wiesbaden, Athenäum.

Pierrehumbert, J. (1993), *Dissimilarity in Arabic verbal roots*, in *Proceedings of NELS 23*, 367–381.

Rubino, C. (1999), *Ilocano Grammar and Dictionary*, Honolulu, University of Hawaii Press.

Rubino, C. (2001), *Iconic morphology and word formation in Ilocano*, in F.K. Erhard Voeltz & C. Kilian-Hatz (eds.), *Ideophones*, Amsterdam, Benjamins.

Seuren, P., & H. Wekker (1986), *Semantic transparency as a factor in creole genesis*, in P. Muysken & N. Smith (eds.), *Substrata versus universals in creole genesis*, Amsterdam, John Benjamins, 57–70.

Slobin, D.I. (2001), *Form-function relations: how do children find out what they are?*, in M. Bowerman & S.C. Levinson (eds.), *Language acquisition and conceptual development,* Cambridge, Cambridge University Press, 406–449.

Sneddon, J.N. (1996), *Indonesian reference grammar,* St Leonards NSW, Allan & Unwin.

Spencer, A. (1991), *Morphological theory,* Oxford, Blackwell.

Williams, E. (1997), *Blocking and anaphora*, "Linguistic Inquiry" 28, 677–628.

Appendix 1. Motivation for Sense, Name and Bad words as Expressive or semantically complex

(See Klamer 2002 and the references cited there.)

Items from the Sense category in Table 1 are generally well-established expressives (see Hinton et.al. 1994). Items from the categories Name and Bad may be more controversial as 'expressives', but it should be noted that the distinction between sound symbolic forms on the one hand, and names and taboo words on the other, is not sharp. For example, names often derive from vocabulary used to refer to sounds, motions, and shapes, reflecting visible or audible characteristics of the named person, plant or animal (e.g. body shape, hair colour, bird's call, animal movement). For example, in Mundang (Niger-Congo), animal and plant names are part of the same type of expressive vocabulary as ideophones (Elders 1999), in Estonian and Finnish, bird names are expressive forms to some extent (Antilla 1976), and in Greek, nicknames pattern with the other expressive forms (Joseph 1997). Bartens (2000: 166–169) explicitly discusses 'de-ideophonic' animal names in a number of Atlantic Creoles. This suggests that there is no categorical distinction between Sense items and Names in a language. With respect to the semantic type Bad (taboo words and words with negative connotations); there is cross linguistic evidence that words from the Bad type may pattern structurally and semantically with the Sense items (for Japanese: Kita 1997:98, Hamano 1998; for Balinese: Clynes 1995, 1998, and for Greek: Joseph 1994, 1997). In addition, there are cases where the distinction between the types Bad and Name is fluid (cf. English baboon as animal name and epithet in English), so if Name is a semantically complex type, then Bad is too.

Appendix 2. Additional evidence for the iconic matching of form and function in expressives

The evidence presented above concerned morphologically complex items that were semantically complex. Klamer (2002) contains quantitative evidence from other linguistic domains: there are certain classes of Kambera and Dutch words with a complex phonotactics or with complex segments that show a statistical tendency to match that formal complexity to expressive meanings. Below follows some additional phonological evidence from Austronesian languages that suggests a similar iconic patterning in the lexicon of these languages. This particular evidence shows that phonotactically/prosodically complex base forms in Tetun, Ilocano, and Balinese tend to be semantically expressive.

Fehan Tetun

Root forms in Fehan Tetun (Central Timor, Van Klinken 1999) are generally di- (55%) or tri-syllabic (43 %). Only 2% of the roots have 4 syllables.
Trisyllabic words are prosodically complex: they consist of one disyllabic foot and an extrametrical syllable. In general, we can say that a Tetun root is a disyllabic foot:

(a)   Root = PrWd = F = $_{\sigma\sigma}$

Four-syllable roots violate this constraint. 9 illustrations given by Van Klinken (1999:16):

(b)   *akitou*   'dove'   *bibiliku*   'drum' (noun)
       banokae   'kind of sea shell' *labadain*   'spider'
       kaibók   'leaf vegetable'   *tualekik*   'wake songs'
       sibalebok   'parsley'   *liurai*   'executive noble'
       maufinu   'danger'

Note that 7 of these forms are semantically expressive (Name, Bad). If this list is representative for the class of four-syllable lexemes in Tetun, it suggests that semantic expressiveness is matched with a complex form: a form that violates the constraint against prosodically complex roots.

Ilocano

Ilocano (Phillipines, Rubino 1999, 2001) roots are usually disyllabic CV(C).CV(C):
(c)   Root = F = $_{\sigma\sigma}$

There are less than 5 monosyllabic roots, e.g. *wak* 'crow' and *waw* 'thirst' and three- or four-syllable roots (all monomorphemic) are generally expressive: most of them represent sounds (repetitive or rustling), as in:

(d)   *sa.ʔid.dek* 'hiccup'   *sa.ʔib.bek, sa.ʔin.nek* 'sob'
       ta.rat.tat 'sound of typing'   *ka.ra.sa.kas* 'rustling sound of leaves'
       dis.su.or 'waves breaking'   *ka.ra.si.kis* 'rustling sound of bamboo'
       sa.ra.i.si 'waterfall'   *ʔu.bu.ʔub* 'fumigate'
       dil.la.wit 'instant, brief period of time'
       sa.rung.kar 'visit'

Balinese

In Balinese, semantic and formal markedness are also aligned, as argued by 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 expressives in this language: they are meaningless but inelegant words that have 'bad' connotations. All of them violate at least one structural constraint that applies elsewhere in the language. Illustrations:

(e) Constraints violated by Balinese 'bad names' (Clynes 1995: 51–52, 1998: 21–22)

Onset:
"Every syllable must have an onset": violated by the bad names: *Cluit, Joet.*

* Complex ^{ONS} :
"No complex onsets": violated by the bad names *Klemug, Namprut, Gomblos, Cluit.*

* _σ [/h/ : "No /h/ as onset":
violated by the bad name *Cibuhut.*

Root= _{σσ} :
"Roots must be bisyllabic": violated by the bad names *Cidaku, Cibuhut, Maseni*

Vowel harmony:
"Cooccurring [+ATR] vowels agree in height": Violated by the bad names *Kedi, Keni, Maseni, Toti.*

Consonant disharmony:
 "Two homorganic consonants do not cooccur in a root": violated by the bad names *Cidaku, Namprut, Toti, Latep, Petet.*