Semantically motivated lexical patterns: a study of Dutch and Kambera expressives

Klamer, Marian

SEMANTICALLY MOTIVATED LEXICAL PATTERNS: A STUDY OF DUTCH AND KAMBERA EXPRESSIVES

MARIAN KLAMER

Leiden University

Are linguistic signs really as arbitrary as they are assumed to be, and how exceptional are semantically motivated lexical patterns? These old questions are re-addressed with the help of new quantificational data on two genetically unrelated languages, Kambera (Austronesian) and Dutch (Indo-European). The hypothesis is that semantically complex items such as expressives favor a structurally complex form, and vice versa. An examination of seven distinct types of lexical items found statistically robust semantically motivated lexical patterns. Linguistic signs appear to be less arbitrary than is commonly assumed.*

INTRODUCTION. Since Saussure proposed that 'the linguistic sign is arbitrary' (1916, part 1, ch. 1, 1st principle), this hypothesis has become one of the basic assumptions in modern linguistics, and is standardly repeated in linguistics textbooks. Textbooks also commonly report the flip side of this claim, that sound symbolic forms are 'exceptional' linguistic signs because of their iconic character. But to date, there have been few quantitative studies of particular lexicons in support of these claims.[1]

This article reports on preliminary research on the native lexicon of two genetically unrelated languages: Kambera (Austronesian, spoken on the island of Sumba in Eastern Indonesia) and Dutch (Indo-European, national language of The Netherlands).[2] Within the lexicon of Kambera and Dutch, I focus on one specific domain of the lexicon: the expressive elements. Expressives are usually taken to include morphemes for sense impressions such as onomatopoeias, ideophones and phonestemes, but this article presents evidence that names (nicknames, epithets, terms of endearment, animal and plant names), as well as words with negative connotations pattern together semantically and structurally with words for sense impressions. The examination of the form-meaning relationship of seven types of lexical items in Kambera and Dutch suggests that there is a statistically significant positive correlation between semantic expressiveness and structural complexity.

* The research for this article was carried out in 1999 at the Free University, under a fellowship of the Royal Dutch Academy of Arts and Sciences (KNAW). Related topics within the same research are reported on in Klamer (2000b, 2002). The revision was done as part of the Spinoza research program 'Lexicon and Syntax' of Pieter Muysken, in 2000 and 2001 at Leiden University. I wish to thank Geert Booij, Adrian Clynes, Claartje Levelt, Simon Musgrave, Ger Reesink, Miriam van Staden, Ruben van de Vijver, Henk Westerik, the anonymous *Language* referees as well as Mark Aronoff and Maria Polinsky for their valuable suggestions on earlier versions of this article.

[1] Apart from Clynes 1995, 1998, which inspired the present study, other recent studies of statistical patterns of form-meaning mapping in the native lexicon are Berlin 1994, on sound symbolism in bird and fish names in Jivaro, and Sereno 1994, on the correspondence between phonological structure and lexical category (noun, verb). Older studies include Uhlenbeck 1949, Chastaing 1964, 1965. More recently, Pierrehumbert 1993, Frisch 1996, Frisch et al. 1997 [1995], and Frisch & Zawaydeh 2001 investigated statistical patterns of the Arabic lexicon (in particular, the effects of consonant coocurrence constraints in the verbal roots of Arabic). Pierrehumbert 1994 deals with dissimilarity requirements operating on English consonant clusters across intervening material. Though these works also deal with statistical patterns in the lexicon, they differ from this article in that they are not specifically concerned with the nonarbitrary mapping of form and meaning.

[2] Non-native lexical items are often phonotactically marked. The markedness of such elements is caused by the fact that they originate from a different (non-native) linguistic system. They will not be considered here.

1. ANALYTICAL BACKGROUND. In this section I describe the notions SEMANTICALLY COMPLEX, EXPRESSIVE and STRUCTURALLY COMPLEX as they are used in this study. Semantic complexity: a semantically marked word is a conceptually complex word, which has, in addition to basic semantic features, one or more evaluative, subjective, and/or descriptive semantic feature(s). In other words, semantically complex items, such as expressives, are more specific and less general than common, prototypical, referential lexemes, and they are more determinate and differentiated. For example, expressives such as ideophones usually evoke concrete images and have a quite specific semantics.[3] I focus on expressive items that belong to one of the semantic types shown in Table 1.

TYPE	EXPLANATION	EXAMPLES
Sense words	Words denoting sense impressions (sound, touch, taste, smell, feeling, emotion, and sight, including movements of the body and/or body parts).	English (E):[a] *tweet*, *blob*, *burp*, *bob* Dutch (D): *lallen* 'to jabber, to slur one's words', *lillen* 'to drill, trill' Kambera (K): *tòku* 'knock, bang', *pelung* 'to wriggle, writhe, wrap around'
Names	Personal or place names, nicknames, epithets, terms of endearment, names for plants and animals.	E: *Bob*, *baboon*, *moron* D: *Ruurd* 'male name', *bullebak* 'unfriendly, coarse man' K: *pirih* 'kind of parrot', *helap* 'kind of fish'
Bad words	Taboo words, and lexical items with negative connotations or items that refer to undesirable states.	E: *boob(s)*, *tit(s)* D: *lul* 'prick; jerk', *wrang* 'sour' K: *nyimba* 'be blocking the way', *pengat* 'emaciated, weakened, impaired'

TABLE 1. The semantic types of expressive items.

[a] English examples included for expositional reasons only; they will not be analyzed here; but see Fudge 1970, Clynes 1995, 1998, 2000 and the references cited there.

As shown by the studies in Hinton et. al. (1994), and the references cited there, onomatopoeias, ideophones, sound symbolic forms, and words referring to specific types of movement, touch, taste, smell, feeling, or emotion are generally accepted as expressive semantic types. In other words, items of the Sense category in Table 1 are generally well-established expressives. Expressives of the semantic categories Name and Bad may be more controversial (but see Uhlenbeck 1949, Fudge 1970, Clynes 1995, 2000 for additional evidence). Considering Names and Bad words expressives is motivated by conceptual and structural considerations. Conceptually, Names and Bad words are more complex than, for example, common nouns (a name, epithet, or nickname is quite specific, and more determinate than a common noun like *man*), and, because of their special connotations, negative or taboo items are semantically more specified than neutral lexical items. Structurally, the distinction between sound symbolic forms on the one hand, and names and taboo words on the other, is not sharp, as for instance when names derive from the 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 color, bird's call, animal movement).[4] Such cases suggest that

[3] For example, Kisi (Niger-Congo) has two rice-beating ideophones, whose semantics differ only in the number of people beating the rice: *gbun gbun* 'rice beaten by one person' and *pim pim* 'rice beaten by two or more people' (Childs 1994:188). Kambera has two ideophones to refer to clicking sounds of the mouth: *bèsu* 'click with cheek', and *dòtu* 'click in back of mouth' (see Table 2 below). Clearly, these ideophones are more determinate and differentiated than more general verbs such as 'beat' or 'click'.

[4] For example, in Mundang (Niger-Congo), animal and plant names are part of the same type of expressive vocabulary as ideophones (Elders 2001), 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 (2000:166–69) explicitly discusses 'deideophonic' animal names in a number of Atlantic Creoles.

there is no a priori reason to assume a categorical distinction between items of the type Sense and Name in a language. With respect to the semantic type Bad (taboo words and words with negative connotations), there is crosslinguistic evidence that words from the Bad type may pattern structurally and semantically with the Sense items.[5] 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, than Bad is too. In the remainder of this article I assume that expressive items are semantically complex, and present themselves as any of these three semantic types.

In this study, STRUCTURAL COMPLEXITY is defined in terms of the relative violation of structural constraints on the wellformedness of lexical items.[6] If we view the structural system of language as a set of constraints on the wellformedness of linguistic utterances, we can say that a more constrained item is less structurally complex, because it violates fewer constraints: items that obey the constraints are structurally better 'behaved' than those that violate them.

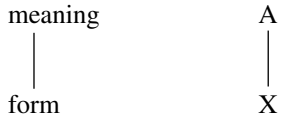
I will show that semantically complex items (Sense, Name, Bad) are formally less constrained than core lexical items. Ideophones are again a case in point: they often use marked segments and/or violate phonotactic constraints of a language; they display very little or exceptional morphology, and show a relative absence of syntax (Mithun 1982, Childs 1994). I propose that the constraints violated by expressive items are of two basic types:

(a) Constraints concerned with the linking of form and meaning, and

(b) Constraints concerned with structural contrasts between linguistic elements.

An example of a constraint concerned with the linking of form and meaning is the constraint on SEMANTIC TRANSPARENCY (Klamer 2000a, 2000b, 2001). This constraint formulates the age-old insight that there is a universal tendency for linguistic items to prefer a direct, one-to-one matching of form and meaning.[7]

(1) SEMANTIC TRANSPARENCY
Match form and meaning one-to-one



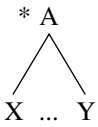
On the morphological side, the constraint is violated by circumfixes, homophones, empty (or meaningless) morphemes, and zero morphemes, as indicated by the diagrams in 2.

[5] This has been reported for Japanese (Kita 1997:98, Hamano 1998), Balinese (Clynes 1995, 1998) and Greek (Joseph 1994, 1997).

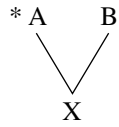
[6] For reasons of space, the discussion here will be limited to morphological and phonological constraint violations, but the implication is that all the constraints in a constraint-based theory of language can in principle be relevant as markedness values in a particular language.

[7] As one referee rightly remarked, the expressive forms in Table 1 are also considered 'marked' in most current phonological theories for purely formal (and often different) reasons. Note that the notion of semantic transparency is not proposed here to replace those considerations, but rather to make explicit WHY certain items are structurally marked.

(2) a. circumfix:



b. homophones:



c. meaningless morpheme:

d. zero morpheme:

Not all schools of linguistics accept the existence of circumfixes,[8] homophones, meaningless morphemes, or zero morphemes. Note, however, that these terms are used pretheoretically here, to refer to morphological phenomena that are typologically less common than ordinary prefixes or suffixes, and as such, are instances of violations of the constraint in 1. Morphemes without a clear semantic content (empty or meaningless morphemes, e.g. the so-called cranberry morphemes, Spencer 1991:40; see also Jackendoff 1997:149) arguably occur in every language, but they are significantly smaller in number than meaningful affixes, hence the constraint violation in 2c. Similarly, it is well known that morphological derivations can be instantiated without an overt morpheme being involved *((a) ring > (to) ring)*, and also that the absence of a morpheme in a certain position may be meaningful. In both cases, we have a distinct morphological meaning but no overt form, and 2d represents that this situation also violates constraint 1. Constraint violations 2a and 2c are examined below: Kambera circumfixes in §2.1, Dutch meaningless morphemes in §2.2, and Kambera meaningless morphemes in §3.3.

The second type of constraint that expressive items may violate is the constraints on structural contrasts between linguistic elements. These constraints on contrasts can be divided into two distinct types with opposite effects: constraints that preserve distinctness and dissimilarity, and thus maximize formal contrasts, and constraints enforcing structural simplicity, thus minimizing contrast. Illustrations of the former are the constraints on (highly) similar homorganic consonant pairs: such pairs are found less frequently than dissimilar homorganic consonant pairs (the OCP effect; Pierrehumbert 1993, Frisch et al. 1997). An instance of such a constraint is the Dutch constraint on identical homorganic liquid consonants (see §3.4). Constraints that enforce structural simplicity and minimize contrast generally take the shape of constraints on complex segments, or complex phonotactics. Instances of such constraints are the Kambera constraints on complex vowels (§2.1), root-final consonants (§3.1), root-initial complex consonants (§3.2), and the Dutch constraint on certain branching onsets (§3.4).

2. EXPRESSIVE ITEMS ARE STRUCTURALLY COMPLEX. In this section I discuss two types of lexical items that are commonly considered to be semantically expressive: Kambera ideophones and Dutch epithets.[9] In both cases we find a motivated matching of complex semantics and complex structure.

[8] Circumfixes are e.g. *ge—t* in German *ge-mach—t* 'made', or *ge—te* in Dutch *ge-berg-te* 'mountain range'.

[9] There are, of course, more expressive items in both languages, such as interjections. One referee suggested that, since interjections universally serve 'expressive' purposes and are known to be marked structures, they should have been included in the study. But interjections are structurally an extremely heterogeneous class so it is questionable whether they should be treated as one class in the first place (compare, for example, the structural properties of *ouch*, *holy shit*, *well*, *mamma mia* and *huh*; see de Groot 1963 on Dutch interjec-

2.1. KAMBERA IDEOPHONES. Kambera ideophones are lexical root forms that directly refer to sounds, motions, and sights, and can therefore be classified as belonging to the semantic type Sense words. The class of ideophonic roots is an open class, and they make up approximately 10 percent of the Kambera lexicon. Table 2 gives examples of ideophonic roots (cf. Klamer 1998:15–16, 245–47, 255, 270).

ngùru	'sound of murmur'	ndòri	'silent' (no reaction)
mbùtu	'thudding sound'	pàdi	'quiet, silent' (no sound)
heir	'tearing noise'	reu	'sound of talking'
tòru	'rattling sound'	yidi	'shiver' (in dislike)
nggòru	'crack' (of thunder)	wàdi	'blink'
tòku	'knock', 'bang'	ngàdu	'nod' (motion)
ndùru	'roll' (of thunder)	linji	'jumping motion'
pàka	'smack'	nggidi	'shiver' (of cold)
mbùku	'snap', 'tap'	tila	'convulsion'
mbèri	'rasping, grating sound'	ndiku	'jerk' (to get loose)
bèsu	'click' (w. cheek)	tàta	'vibrate, shake'
dòtu	'click' (in back of mouth)	jila	'gleam, flash'
		rèri	'emit light, sparkle'

TABLE 2. Kambera ideophonic roots.

Syntactically, the ideophonic roots are exceptional because they can surface only in the position of a quote in a special 'quotative' construction, used to report speech acts as well as thoughts and physical perceptions of motions, sounds, and visions (Klamer 2000a). Morphologically, they are special because they are the only root forms that, in order to be used as verbs, must be derived by circumfixation or reduplication, rather than by prefixation or suffixation, as 'normal' roots are. The circumfix to derive ideophonic verbs from roots is *ka-. . .-k*, as illustrated in 3, 4a and 5a. Without this circumfix, the ideophones can appear only in quotative constructions, as in 4b and 5b, that is, they cannot be used verbally.

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

(4) a. Hili odah-ya na hapapa ka-mbùtu-k -danya da marara
again stroke-3SG.ACC the side (fall) thudding -3PL.CONT the gold[10]
'Again (he) stroke the (horse's) side, thudding the gold fell out'
b. Mbutu wà-na
thud report-3SG.GEN
' ''Thud'' it did.'

(5) a. Na- ka-jila-k na uma
3SG.NOM gleam ART house
'The house gleams/shines.'
b. Jila wà-na na uma
gleam report-3SG.GEN ART house
'The house gleamed.'

The circumfix *ka-. . .-k* violates the constraint on semantic transparency because it consists of two separated forms representing one meaning (cf. 2a).

tions). In this preliminary research, I avoided these questions, and concentrated on expressive items of a structurally more homogeneous type.

[10] Abbreviations: 3: 3rd person; ACC: accusative; ART: article; CONT: continuative, GEN: genitive; NOM: nominative; PL: plural; SG: singular.

Kambera ideophones also diverge from the unmarked patterns of the language in their segmental makeup. The vowels in ideophones can be taken from a restricted, marked set. The unmarked Kambera vowels are /i, i:, a, a:, u, u:, e, o/, orthographically represented as *i*, *í*, *à*, *a*, *u*, *ú*, *e*, *o*.[11] These vowels occur in the stressed syllable of all types of roots.[12] But the low vowels /ɔ, ɛ/ and the super-short high vowel [ù] (orthographically represented as *ò*, *è*, *ù*, respectively) occur only in this restricted context of the stressed syllable of an ideophone, as in Table 2. In other words, only ideophones contain marked vowels.[13] I impressionistically describe the general dispreference for the marked vowels /ɔ, ɛ, ù/ in Kambera as a constraint against such vowels: 'Don't be /ɔ, ɛ, ù/'.

In sum, Kambera ideophones show a positive correlation between semantic and structural markedness: the forms match their semantic specificity with a systematic violation of (i) the semantic transparency constraint by using circumfixes, and (ii) the constraint against marked vowels /ɔ, ɛ, ù/ by using the vowels from exactly this set in the stressed syllable. One could also say that they show an 'iconic' relationship between complex form and expressive semantics. But there is no literal iconicity. Kambera ideophones may refer to sights, motions, or states where no sound is involved: cf. Kambera *ndòri* 'silent' (no reaction) and *pàdi* 'quiet, silent' (no sound), *yidi* 'shiver', *wàdi* 'blink', and so on. And, the formal properties of the ideophones are not an 'image' of the 'concept' represented. For example, the fact that Kambera ideophones feature the three vowels /ɔ, ɛ, ù/, which are not otherwise used in the language, is not a direct image of their expressive semantics, yet it is typical for this class of words. So what appears to be involved here is a more abstract notion of iconicity of form and meaning, which has been described as DIAGRAMMATIC ICONICITY (see Peirce 1965, Haiman 1994). Diagrammatic iconicity is the notion that 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 (here, a complex rather than simple set of semantic features) is matched by an element with the same relative position in system B (here, a phonotactically complex rather than simple form). I suggest that the same type of iconicity pertains to the other expressive items discussed in this article.

2.2 DUTCH EPITHET COMPOUNDS. I turn now to a class of items in the Dutch lexicon that are semantically expressive but whose structural properties have not been studied before: the epithet compounds. The members of this class are productively derived nominal compounds and they generally function as epithets; a few are terms of endearment. Representative sets of examples are given in Tables 3 and 4.[14] The meaning of an epithet compound is not usually a sum of its parts: as the examples show, many of the compounds are exocentric. Structurally, the epithet compound consists of a nominal head, a preceding modifier (a noun in Table 3 and a verb/adjective in Table 4) and a linking element. Linking elements in Dutch compounds take the leftmost, modifying element as their (morphological and prosodic) host. There are three linking elements

[11] The /a/ is orthographically represented as *à* and is phonetically realized as [ɑ] or [a].

[12] Vowel length and height contrasts are neutralized in unstressed syllables; vowels used as unstressed nuclei are the cardinal vowels /a, i, u/.

[13] Note that ideophones also contain the unmarked vowel /i/, and that the short low vowel [ɑ/a] (represented as *à*) is also used in nonideophonic roots, for instance *tàka* 'arrive', *nyàmba* 'to worship', *dàngu* 'with, and', *pàdang* 'experience X' (Klamer 1998:14).

[14] Most of the items are listed in Heestermans (n.d. [1990]), and some are from my own experience as a native speaker of Dutch.

rat-ə-kop	'rat face'	snot-ə-pot	'snotty baby/toddler'
rat-ə-head		snot-ə-pot	
aap-ə-kop	'brat, naughty kid'	kip-ə-kop	'chicken face'
monkey-ə-head		chicken-ə-head	
hond-ə-lul	'unhelpful male, jerk'	lel-ə-bel	'dirty girl'
dog-ə-prick		whine-ə-bell	
wrat-ə-kop	'ugly person'	tut-ə-bel	'fussy or bitchy girl'
wart-ə-head		silly.cow-ə-bell	
bril-ə-jood	'ugly person with	lul-ə-pot	'crap; complete
specs-ə-jew	glasses'	prick-ə-pot	nonsense'
kut-ə-kop	'bitch (woman)'	snoes-ə-poes	'sweetheart'
cunt-ə-head		darling-ə-pussycat	
kloot-ə-wijf	'unhelpful, nasty	schat-ə-bout	'darling'
ball-ə-bitch	woman'	treasure-ə-leg	
kloot-ə-vent	'unhelpful or nasty	griet-ə-p-iet	'darling girl'
ball-ə-bloke	male'	girl-ə-p-RED	
bull-ə-bak	'unfriendly, coarse	meis-ə-p-eis	'darling girl'
bull-ə-box	man'	girl-ə-RED	

TABLE 3. Dutch [N + N] epithet compounds.

mank-ə-poot	'cripple'	huil-ə-balk	'crybaby'
limp-ə-leg		cry-ə-pole	
dom-ə-kracht	'mindless hulk'	knor-ə-pot	'grumpy person'
dumb-ə-force		grunt-ə-pot	
jank-ə-pot	'crybaby'	teut-ə-bel	'dilly-dally'
whine-ə-pot		dawdle-ə-bell	
breek-ə-been	'clumsily moving	mors-ə-bel	'person spilling
break-ə-leg	person'	spill-e-bell	food/drink'
drink-ə-broer	'man who drinks (too	hink-ə-poot	'cripple'
drink-ə-brother	much) alcohol'	hop-ə-leg	

TABLE 4. Dutch [A/V + N] epithet compounds.

that can appear in Dutch nominal compounds: [ə], [s], and [(ər) ən] (see, e.g. de Haas & Trommelen 1993:402). Here, we are concerned only with the schwa. The schwa is realized orthographically as *-en-* or *-e-*.[15] In the examples, the linking element is represented and glossed as -ə-.

The linking element -ə- in compounds has three possible historical origins (see, e.g. Haeseryn et. al. 1997:684–85, Booij & van Santen 1998:157): (i) It may be the final lexical vowel in a nominal stem that was lost in isolation, but remained part of lexicalized compounds (Middle Dutch *panne* became *pan* 'pan' in isolation, but remained *panne* in *pan[ə]koek* 'pancake'); (ii) It may be a (genitive) case ending (*heer[ə]huis* 'mansion, lit. house of a gentleman/lord'); (iii) It may be a plural ending (*boek[ə]kast* 'bookcase'). Synchronically, there is a tendency for -ə- to mark plurality in compounds (Haeseryn et al. 1997:685), but plural semantics may also be conveyed by the linking element /-s-/. Note, however, that the origin of the linking element in the [N + N] forms in Table 3 is unclear: all these forms are (relatively) recently coined epithets, so the schwa cannot be an original stem vowel or a case ending. Furthermore, it does not convey plural semantics; it is a meaningless element here, and probably the result of

[15] Reference grammars with detailed information on Dutch compounding are de Haas & Trommelen 1993 and Haeseryn et al. 1997. The literature on Dutch compounds generally agrees that phonological, morphological, and semantic factors play a role in the distribution of the linking elements (see e.g. van den Toorn 1981a, 1981b, 1982a, 1982b, Mattens 1984).

analogical word formation.[16] The point here is that, for historical reasons, the linking schwa is expected to occur in [N + N] compounds *only*, taking the first N as its host.[17] In other words, its occurrence in the [A + N] and [V + N] compounds is entirely unmotivated, because in the regular case, such compounds do not contain a linking element; compare the minimal pairs in Table 5. The schwa in the compounds of Tables 3 and 5 is meaningless.[18]

[A/V + N]		[A/V-ə-N]	
dom-oor	'idiot'	dom-ə-kracht	'mindless hulk'
dumb ear		dum-ə-force	
jank-kind	'child that cries much'	jank-ə-pot	'cry baby'
whine child		whine-ə-pot	
breek-ijzer	'crowbar'	breek-ə-been	'clumsily moving person'
break iron		break-ə-leg	
huil-bui	'crying fit'	huil-ə-balk	'crybaby'
cry fit		cry-ə-pole	
mors-plek	'spot with spilled	mors-ə-bel	'person spilling food/drink'
spill spot	food/drink'	spill-ə-bell	
hink-stap	'hop step'	hink-ə-poot	'cripple'
hop step		hop-ə-leg	

TABLE 5. [A/V + N] compounds in Dutch: minimal pairs with and without schwa.

With this meaningless morpheme, the epithet compounds with an adjectival or verbal modifier thus violate the constraint on semantic transparency discussed above (see 1 and 2c).

My conclusion is that (Kambera) ideophones and (Dutch) epithet compounds are semantically expressive forms that are also structurally complex because they systematically violate the constraint on semantic transparency, while the Kambera ideophones also violate a constraint against the marked vowels /ɔ, ɛ, ù/.

3. STRUCTURAL COMPLEXITY CORRELATES WITH SEMANTIC COMPLEXITY. In the preceding section I considered semantically expressive items and looked at their structural

[16] Van de Toorn (1982b) assumes that in general, analogy is one of the forces behind the derivation of nominal compounds with linking elements. The question in what way analogy influences the choice of linking morphemes in Dutch [N + N] compounds is addressed in detail in Krott et al. 2000.

[17] Perhaps significantly, Dutch reference grammars like de Haas & Trommelen 1993 and Haeseryn et al. 1997 discuss the linking schwa only in connection with [N + N] compounds. Though [A + N] and [V + N] compounds are discussed, and a few of them do feature a linking schwa, its distribution in those compounds is not further discussed.

[18] An alternative explanation for the presence of the schwa in the forms in Table 4 would be to analyze them as (derived from) nominal phrases, consisting of a nominal head and an adjective with the inflectional suffix -ə, e.g. [*mank*$_A$ *-ə poot*$_N$]$_{NP}$ 'crippled leg'. For the [V + N] compounds this analysis would imply a stage where verbs become adjectives by conversion. There are several reasons why this analysis cannot be correct. First, the stress pattern of the items in Tables 3 and 4 is that of compounds, not of phrases: in phrases, the phrasal head receives main accent: *een zwart bóek* 'a black book', whereas in compounds main stress is on the leftmost element: *een zwártboek* 'document containing negative/critical evaluations'. In the epithets above, main stress is always on the leftmost element. Thus, *mánkəpoot* is a nickname for a crippled person, but *(een) mankə póot* is a phrase, '(a) crippled leg'. A second piece of evidence is morphological. Dutch diminutives are neuter, and adjectives preceding a neuter noun do not inflect. Therefore, if the -ə in the above forms were an adjectival inflection in a nominal phrase, we would expect it to disappear when the compound noun is diminutive, contrary to fact.

(i) [*wrat-ə-kop-je*]$_N$ **wratkop-je*
wart-ə-head-DIM

[*mank-ə-poot-je*]$_N$ **mankpoot-je* cf. [*een* [*mank poot-je*]$_{AP}$]$_{NP}$
cripple-ə-leg-DIM a cripple leg-DIM

properties. I now reverse this process, compiling sets of structurally complex items and considering their semantics to see if there is a correlation. The null hypothesis is that there is no correlation; the alternative hypothesis is that there is a positive correlation. The Kambera data are based on a corpus analysis by Onvlee (1984), a comprehensive Kambera-Dutch dictionary, while the Dutch data are based on a corpus analysis of the standard dictionary of Dutch (Van Dale Groot Woordenboek der Nederlandse Taal), and on an experiment with Dutch native speakers.

The formal complexity of the items discussed here has been motivated independently in the literature on Kambera and Dutch (for Kambera, see van der Hulst & Klamer 1997a, 1997b, Klamer 1998; for Dutch, see Booij 1995:35–43, and the references cited there). The semantic classifications of the items came from independent sources as well. For the Kambera data, I relied on the translations given in Onvlee 1984; for the Dutch data I consulted the intuitions of seven native speakers of Dutch who were unaware of the topic under investigation.

In section 3 I examine formally complex words in Kambera and Dutch: (i) Kambera roots with a final lexical consonant ((C)VCVC) that are prosodically complex, (ii) simple Kambera roots (CVCV) that start with a complex initial segment, and (iii) Kambera words with an empty (meaningless) prefix *la-*. I also consider two types of structurally complex Dutch words: words that violate the constraint on two homorganic identical liquid consonants (Booij 1995:42–43), and words that violate a constraint on the branching onset *wr-* (Booij 1995:35–43).

3.1. KAMBERA MARKED ROOT FORMS. The majority of Kambera roots conform to the canonical CVCV pattern, but a large percentage of the lexical roots—approximately 30%—are consonant final ((C)VCVC). These roots end with one of the consonants: *k*, *ng* (= /ŋ/), *r*, *l*, *p*, *t*, or *h*. Illustrations are given in Table 6.[19]

tehik	'sea'
uhuk	'sit'
nggidik	'tremble'
padang	'field
múhung	'be rotten, gone bad'
unung	'drink something'
watar	'corn'
tangar	'watch someone/something'
winggir	'turn around (intr.)'
ka-lipar	'be lame'
engal	'kind of tuberous plant'
ka-bunggul	'be short'
banjal	'leave something, put away, store something'
mungal	'fall out, slip out, slip off'
helap	'kind of sea fish'
holap	'weak, limp, flabby'
jakap	'walking of a tall, imposing man'
langgap	'crowl'
ma-ngadat	'be afraid'
akat	'have bad character'
pirih	'kind of parrot: Trichoglossus heamatodus'
duruh	'continue something'
punduh	'skip, jump'
pàlih	'lick mouth/lips'

TABLE 6. Kambera (C)VCVC roots.

[19] The table includes VCVC forms—vowel-initial stems in Kambera pattern phonotactically like consonant-initial stems (Klamer 1998:16–21). A glottal stop is inserted automatically as their initial consonant

Despite the fact that such (C)VCVC roots are unmarked in terms of lexical frequency (they make up one-third of the Kambera roots!), data from prosody, reduplication, and language games indicate that the final consonant is not fully integrated into the prosodic template that is generally applicable for Kambera root forms—the trochaic foot. (Evidence and independent motivation of the prosodic complexity of (C)VCVC roots is given in van der Hulst & Klamer 1997a, 1997b and Klamer 1998:16–34). In other words, though Kambera consonant-final roots are morphologically simple (they are morphological roots, not derived words), they are prosodically complex, because they consist of a foot combined with an additional consonant.

Taking this analysis as correct, I set up a corpus analysis addressing the question of whether there is evidence that the complex (C)VCVC root forms are also semantically complex.

A corpus analysis of Kambera (C)VCVC root forms

HYPOTHESIS: structurally complex (C)VCVC root forms are also semantically complex.

MATERIALS

From a corpus of approximately 7300 Kambera words (Onvlee 1984) I compiled two sets of roots with the simple initial consonant *p-*. They are given in Appendix B. (Evidence for the unmarked status of this consonant is given in §3.2 and Appendix A). One set contained all the Kambera words starting with /p/ that have the complex root form (C)VCVC. To ensure that the structural complexity of these forms was not caused by segments other than the final consonant, roots containing a marked vowel /ɛ, ɔ, ù/ (see §2.1) and /ɑ/ were excluded from the set. This test set consisted of 106 items; the control set contained all structurally simple root forms (CVCV) with initial /p/. Only items with a C2 from the same set of consonants that can appear as final C (C3) in the test set (i.e. /k, ŋ, l, r, h, t, p/), were included in the control set. Again, roots containing a vowel /ɛ, ɔ, ù, ɑ/ were excluded from the set. The control set consisted of 62 items.

PROCEDURE

Both sets of roots were examined for their semantics, as indicated by the translation given in Onvlee 1984. Their semantics were classified according to the presence or absence of any of the expressive semantic types Sense, Name, and Bad. I therefore translated Onvlee's general descriptions in terms of those semantic types. This procedure contains a certain element of subjectiveness, especially in the case of the Sense and Bad type words. However, I classified words as Sense type only when the translation or the accompanying example sentences clearly conveyed an expressive sense (e.g. 'type of sound', 'kind of motion', 'movement of the body', 'facial expression', etc.) Examples are *pilik* 'to wag (tail), to flap, to flutter', *porak* 'to snap, to crack', *púpur* 'have a sour, surly, grumpy face'. Only those words whose translations had clearly negative, unpleasant, or taboo connotations were classified as Bad type, for example, *pengat* 'emaciated, wasted, weakened, impaired', *púbuk* 'half, incomplete, not whole, not real, untrue', *pukul* '1. to miss, to not touch or hit, 2. mutilated, numb'. See Appendix B for details.

(Klamer 1998:11–12). The present discussion is only concerned with consonant-FINAL stems—the shape of the first syllable of the stem is irrelevant.

RESULTS AND DISCUSSION

The results, distinguished per consonant (C3 and C2) are presented in Tables 7 and 8. The distribution of the semantically expressive elements in the test set with CVCVC roots is: out of an N of 96, 41 roots had expressive semantics and 55, nonexpressive semantics (Table 7). The distribution of the semantically marked elements in the control set with CVCV roots is: out of an N of 62, 14 roots had expressive semantics and 48 nonexpressive semantics (Table 8).

SEMANTICS/C3	k	ŋ	l	r	h	t	p	TOTALS
Expressive								
Sense	8	3	4	10	4	4	0	33
Name	0	1	0	0	2	0	0	3
Bad	1	1	1	0	0	2	0	5
Other	11	17	7	8	4	8	0	55
Totals	20	22	12	18	10	14	0	96

TABLE 7. Semantics of Kambera (C)VCVC roots with initial consonant =/p/, per C3.

SEMANTICS/C2	k	ŋ	l	r	h	t	p	TOTALS
Expressive								
Sense	4	1	0	3	1	1	1	11
Name	0	0	1	0	0	1	1	3
Bad	0	0	0	0	0	0	0	0
Other	5	7	11	8	6	6	5	48
Totals	9	8	12	11	7	8	7	62

TABLE 8. Semantics of Kambera CVCV roots with initial consonant =/p/, per C2.

A one-tailed chi-square test[20] of the figures in Table 9 gives a value $X^2 = 6.72$, which indicates that the chance that the distribution of semantically expressive and nonexpressive forms in the test set is produced by random effects is less than 1 percent (0.01). In other words, there is a positive correlation between the presence of a coda in a Kambera root and the semantic markedness of that root, which allows us to reject the null hypothesis.

SEMANTICS/FORM	COMPLEX	SIMPLE	TOTALS
Expressive	41	14	55
Other	55	48	103
Totals	96	62	158

$\chi^2 = 6.72$; df = 1; $p < 0.01$ (one-tailed)

TABLE 9. Semantics of Kambera (C)VCVC and CVCV roots (Tables 7–8).

This finding is supported by facts from genetically related (Central Malayo-Polynesian) languages of Roti and Timor. Jonker (1906) reports that sound imitations in Roti and motion verbs in Roti and Timor end in /k/ (Jonker 1906:333), while in Dengka and Oenale (languages spoken on the island of Roti), final liquid consonants mark words (nouns/verbs) describing motions (Jonker 1906:341–342).[21] And Middelkoop

[20] To evaluate nominal data like these, a chi-square test is an adequate instrument. The test is one-tailed because the hypothesis is directional (i.e. we expect a positive correlation between semantic and structural complexity).

[21] Jonker analyzes the final consonant as a fossilized 'emphatic marker or adverb' *la*, which at that time was still in use in the related language Termanu (Jonker 1906:342). The marked status of certain roots may thus be the result of grammaticalization processes.

(1950:393–94) gives as one of the two functions of the final consonant /m/ in languages of Timor that it is a marker of plant/tree names.

(6) Oenale: *kaur* 'nod', *kakaler* 'shake one's head'
Dengka: *nggonggal*, Oenale *nggonggonggar* 'shake something'
Dengka: *kapel*, Oenale *kaper* 'beckon'
Dengka: *lenggal* 'open itself', *lofal* 'come loose, snap'
Timor: *ekam* 'pineapple' *nisum* 'fruit-bearing *gebanga* tree', *nanum* 'kind of *ficus*'

Thus, the correlation between the formal complexity of (C)VCVC roots and their semantic expressiveness has been attested in genetically related languages too. Note, however, that in none of the languages are the final consonants EXCLUSIVELY expressive elements. So the observed correlation is a tendency rather than a categorical feature of the (C)VCVC roots.

3.2. KAMBERA COMPLEX CONSONANTS. Another type of complex forms in Kambera is CVCV roots with an initial complex consonant. The full set of Kambera consonants is given in Table 10 (orthographic notation is included in brackets). Observe that Kambera has six complex consonants: four prenasalized stops and two implosive stops.

	LAB	ALV	VEL	GLOT
voiceless stops	p	t	k	
voiced implosive stops	ɓ (b)	ɗ (d)		
voiced affricate		dʒ (j)		
nasals	m	n	ŋ (ng)	
prenasalized stops	mb	nd	ŋg (ngg)	
prenasalized affricate		ndʒ (nj)		
fricative				h
liquids		l, r		
approximants	w	j (y)		
prenasalized approximants		nj (ny)		

TABLE 10. Kambera consonant segments.

Prenasalized and implosive stops are generally considered articulatorily more complex than simple stops; this is represented as formal complexity in models of segmental phonology (e.g. Walli-Sagey 1986, and subsequent literature). In addition, the following test suggests that complex segments in Kambera have a low distributional frequency. I compiled a random set of 3,617 words from Onvlee 1984. This set included both root forms and morphologically derived words. I then counted the initial consonants of the roots (of those words).[22] The most frequent root-initial consonants of Kambera appeared to be the consonants /t, p, r/: 917 or 25.35% of the 3,617 forms considered had /t, p, r/ as initial root consonant. The least frequent root-initial consonants in Kambera turned out to be the prenasalized affricate /ndʒ/ (*nj*) and the prenasalized glide /nj/ (*ny*): they were the initial root consonant in 330 forms or 9.1% of 3,617. (More detailed figures are given in Appendix A.). In other words, while /t, p, r/ have a high token frequency, the plain stops /t, p/ are simple from an articulatory and structural point of view, while /ndʒ/ and /nj/ are articulatorily complex and have a low frequency. With this information on structurally complex and simple initial consonants in mind, the following corpus analysis was performed.

[22] In other words, prefixed words like *pa-hàla* or *ta-binu* were counted, respectively, as root-initial /h/ and /b/; they were not counted as forms with initial /p/ or /t/ because these consonants are part of the prefix, not of the root.

A corpus analysis of Kambera CVCV roots with complex initial consonants

HYPOTHESIS: root forms with complex initial consonants are semantically expressive.

MATERIALS

I compiled two sets of roots from the Onvlee (1984) corpus. One test set contained all the Kambera CVCV roots with the initial consonants /nj/ (*ny*) (28 items) and /ndʒ/ (*nj*) (41 items). The control set contained roots with an unmarked initial consonant /t/ or /p/ and consisted of 40 items. As before, roots containing the vowels /ɔ, ɛ, ù, ɑ/ were not included in either set.

PROCEDURE

The sets of roots were classified, according to the presence or absence of any of the expressive semantic types Sense, Name, and Bad, based on the translations given in Onvlee 1984.

RESULTS AND DISCUSSION

The results for the roots with the complex initial consonants /ndʒ/ (written as *nj*) and /nj/ (written as *ny*) are given in Tables 11 and 12. The distribution of the semantically expressive elements in the control set is given in Table 13.

The frequency of the semantic expressives in the test set was compared to their frequency in the control set, and the results are presented in Table 14.

Applying a chi-square test on these data gives the result $X^2 = 10.58$, which is significant at the 0.01 level. This indicates a very low possibility (less than 1 percent

SEMANTICS	CVCV root with /ny/
Expressive	
Sense	9
Name	3
Bad	6
Sense/Bad	1
Total	19
Other	9
Total	28

TABLE 11. Semantics of CVCV roots with C1 = ny.

SEMANTICS	CVCV root with /nj/
Expressive	
Sense	13
Name	7
Bad	8
Total	24
Other	17
Total	41

TABLE 12. Semantics of CVCV roots with C1 = nj.

SEMANTICS	CVCV root with /t/ or /p/
Expressive	
Sense	7
Name	3
Bad	2
Total	12
Other	28
Total	40

TABLE 13. Semantics of CVCV roots with C1 = /t/ or /p/.

SEMANTICS/FORM	C1 = ny, nj	C1 = t, p	TOTALS
Expressive	43	12	55
Other	26	28	54
Totals	69	40	109

$\chi^2 = 10.58$; df = 1; $p < 0.01$ (one-tailed)

TABLE 14. Semantics of CVCV roots with complex and plain initial consonant.

chance) that the semantically expressive/nonexpressive distribution in the test sets is produced by random effects. I therefore conclude that the use of the complex initial consonants *nj* and *ny* shows a significant correlation with the expressive semantics of the roots of which they are a part.

3.3. KAMBERA EMPTY PREFIX. A third group of formally complex lexical items in Kambera are words with the 'prefix' *la-*. The element *la-* is not a productive prefix: the root forms of derivations with *la-* do not occur independently, and it is impossible to derive new *la-* forms on the basis of existing roots. Native speakers assert that *la-* does not have a meaning. The argument that words with *la-* should be analyzed as complex morphological forms is thus strictly formal: morphologically derived words in Kambera consist of a root plus an unstressed prefixed *Ca*; the prefixes are *pa-*, *ka-*, *ma-*, *ta-* and *ha-*. Prefixes never bear stress; word stress remains on the first syllable of the root. Forms with the prefix *la-* are phonotactically identical to derived words, that is, the prefix *la-* is unstressed, though *la-* does have primary stress when it is the root-initial syllable (see also Klamer 1998:16–31, 260–61).

(7) a. ˈ*laku* 'walk, go', ˈ*ladi* 'bench', ˈ*lala* 'melt' (ˈ indicates primary stress)
b. *la-*ˈ*lei* 'be married to a woman', *la-*ˈ*ngora* 'wipe off', *la-*ˈ*mihi* 'clean away X'

The number of words with a prefix *la-* is extremely restricted: Table 15 gives an exhaustive list. Only the root *lei* of *la-lei* 'be married to a woman' can be used independently (this root form means 'husband (coarse)',[23] not 'woman'). The presence of the meaningless prefix *la-* in these words constitutes a violation of the semantic transparency constraint, and therefore I consider *la-* items as structurally complex forms.

la-lei	'be married to a woman'	la-mbungur	'flower spec.' (*Datura factuosa*)
la-ngora	'wipe off'	la-mboya	'name of medicinal plant'
la-wihir	'turn one's back to, 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/feel weak'	la-nggapa	1. 'tree with thin bark'
la-mbiri	'have a sleepy expression'		2. 'very thin'
la-ngudu	'be in a heap (bodies, clothes)'	la-ngira	'tree sp. used for canoes'
la-wújur	'with bended back'	la-ngaha	'tree sp.' (*Barringtonia asiatica*)
la-muji	'suck'	la-yia	1. 'ginger plant'
la-nggori	'burp'		2. 'brother in law'
la-ngidip	'hickup, gasp'	la-hona	'red onion'
la-nggeha	'be thin'	la-bawa	'white onion'
la-nggudu	'tied with feet together'	la-mbaru	'centipede'
la-mbonga	'deep, dark, large hole'	la-mbàku	'civet cat'
la-pàpu	'ulcer in armpit/groin'	la-wora	'iguana'
la-ngiha	'gums'	la-nggudu	'tuberous plant sp.' (*Toca palmata*)
la-ngàdi	'type of coral'		

TABLE 15. Kambera words with the 'empty' prefix *la-*.

[23] *Mbapa* is the common term for 'husband'.

From the translations given, it appears that the majority of the *la*-derivations belong to a restricted set of semantic types: the nouns are mostly plant or animal names, the Name type, and the verbal forms mostly denote very specific movements or positions/states of the body, the Sense type. In addition, words like *lambonga* and *lapàpu* may be categorized as Bad words. My conclusion is that the Kambera *la*-forms show a significant positive correlation between their semantic and structural complexity.

In sum, we have considered three classes of Kambera lexical items in which structural complexity is correlated with semantic expressiveness: prosodically complex (C)VCVC roots, CVCV roots with a complex initial consonant, and words with an empty prefix *la*-.[24]

3.4. DUTCH MARKED STEMS. I now consider the matching of semantic and structural markedness in Dutch phonotactically marked stems. Booij (1995:35–43) describes the following two phonotactic constraints in Dutch as tendencies with exceptions: (i) the tendency to avoid two homorganic identical liquid consonants, and (ii) the tendency to disallow branching onsets. I discuss the constraints in turn, and observe that the stems that violate these constraints show a preference for a marked semantic interpretation. This observation is the basis for an experiment on the semantics assigned to Dutch nonwords discussed at the end of this section.

Two homorganic identical liquid consonants are generally avoided. The sequence *rVr* is out when the vowel is short, and marginally allowed when the vowel is long; the sequence *lVl* is strange with long vowels and marginally allowed with short vowels (Booij 1995:42–43).[25] To capture the 'strange' and 'marginally allowed' forms we generalize the constraint as prohibiting two identical liquid consonants within one morpheme, irrespective of the length of the vowel (V).

[24] One referee raised the question how polysemy is treated in the correlation. For instance, if a word has both a 'bad' interpretation and a 'non-bad' one, how is it counted? Forms that were interpreted by me (as a non-native speaker of Kambera) as polysemous are indicated with (P) in Appendixes B and C, and their semantic classification is given. Many polysemous words have only nonexpressive meanings, and are thus irrelevant. Of the forms with at least one expressive meaning, I took the most literal and specific meaning as the basic one, and the one to count. Thus, *njingi* 'motion of looking sidewards; look after someone' is counted as a Sense word (describing a motion of looking sidewards, which is metaphorically extended to the activity of looking after someone), but *pelu* 'comb of rooster; vine with fruit as a rooster's comb' is not counted as expressive because the plant name is derived from the word that refers to a rooster's comb, and this word is not expressive. A related, and perhaps more difficult, question is how homophonous forms are to be counted. The total number of Kambera items in Appendixes B and C is 267. Of these, there appear to be 31 homophonous items (indicated with (H)). Of those, 17 do not have an expressive meaning for either of the forms, and are thus not counted as such. Example: *tiki* 1. 'utter' 2. 'almost'. Of the remaining 14 homophones with an expressive meaning of one of their forms, 6 have expressive meanings for both their forms and are thus counted as expressive; and 2 forms have one expressive Kambera form, and one nonexpressive form borrowed from Indonesian. In the one case the form was counted as expressive (*njanga*), in the other case it was counted as nonexpressive (*pikir*). This leaves 6 cases where we have homophones where one of the forms is expressive, and the other nonexpressive. For instance, the form *poku* 1. 'sound of thudding, clapping', 2. 'capital'. Such forms were all counted as expressive. Admittedly, this is a subjective choice. Note, however, that the alternative classification of these forms as NONexpressives would change the overall figures only slightly: of a total number of Kambera items N = 267, the figures would become: 104 expressive, 163 nonexpressive, whereas according to the present classification 110 items are expressive and 157 nonexpressive. In any case, the alternative classification would not change the test results.

[25] Booij (1995:42) describes the constraint as a syllable constraint rather than a morpheme structure constraint on the basis of the existence of words such as *Lola* 'id.' (name) and *rara* 'guess what?'. But note that *Lola* is a semantically marked item (Name), and *rara* does not need to be analyzed as a lexical root morpheme—it could also be a partial reduplication of the verbal stem *raad* 'guess'. Because crosslinguistically, the domain of the constraint on (near-)identical consonants is generally the morpheme rather than the syllable (cf. the OCP effect), I have formulated the domain of the Dutch constraint in morphological terms too. But nothing in the analysis depends on this assumption.

(8) a. * r V: r, * r V r e.g. **roor*, **ror*
b. * l V: l, * l V l e.g. **leel*, **lil*

A list of the forms violating the constraints in 8 is given in Table 16. Note that this is an exhaustive list of these forms, and that this list includes both stem and derived forms. In other words, there are indeed extremely few of these strange and marginally allowed forms in Dutch. The data are from Van Dale Groot Woordenboek and Booij 1995.

a. Words violating */r V: r/ and */r V r/		
Rur	/rʏr/	'name of tv program' (intentional coining, Booij 1995:43, n. 33)
Ruurd	/ry:rt/	'male name'
raar	/ra:r/	'strange, weird'
roer	/ru:r/	'rudder, helm; k.o. call bird; strong bowel movements; pipe'
roer-en	/ru:r/-INF	'to stir, to mix with circular movement'
roer-ing	/ru:r/-NOM	'circular movement; bustle, commotion, stir'
reur-ing	/rœ:r/-NOM	'dialect form of *roer-ing*'
reer-en	/rɪ:r/-INF	'to moo; to rant and rave, to storm; to cry loudly' (obsolete)
b. Words violating */l V: l/ and */l V l/		
lall-en	/lɑl/-INF	'to jabber, babble, slur one's words' (cognate of *lollen*)
lel	/lɛ1/	'ear lobe; clout; whopper'
lell-en	/lɛl/-INF	'to moan, nag, whine; give a hefty kick; box around the ears'
lill-en	/lɪl/-INF	'to drill, trill, esp. of soft material' (meat, dead body, pudding)
lol	/lɔl/	'fun, lark, trick'
loll-en	/lɔl/-INF	'to lull, murmur, mumble, mutter; cry of cat in heat'
lul	/lʏl/	'prick; jerk'
lull-en	/lʏl/-INF	'to murmur, mumble, mutter' (cognate of *lollen*); 'to dawdle over one's work; be weak, shiftless; to talk nonsense'

TABLE 16. Dutch words violating constraint on homorganic identical liquids.

All the words in Table 16 have at least one marked interpretation: the items in (a) belong to one of the marked semantic types Name (3), Sense (4), or Bad (1), while the items in (b) are of the Sense (6) and Bad (2) type. Again, the data show a positive correlation between a complex form—violating a general constraint on homorganic identical liquids—and expressive semantics. But since the number of items is small (N = 16), the observed patterns could be a coincidence. I therefore took the observation as the basis for an experiment about the semantics of Dutch nonwords, described at the end of this section.

The second type of constraint discussed here relates to the tendency of Dutch to disallow branching onsets. This type of constraint follows the universal SONORITY SEQUENCING GENERALIZATION (Selkirk 1982), which states that segments decrease in sonority towards the edges of a syllable. For Dutch, it implies that an onset cannot consist of a cluster of two sonorants (nasals, liquids, or glides). In particular, the branching onsets in 9 are disallowed (Booij 1995:35–39).

(9) * nas + liquid i.e. **mr-*, **nr-*, **ngr-*, **ml-*, **nl-*, **ngl-*
* nas + glide i.e. **mj-*, **nj-*, **ngj-*, **mw-*, **nw-*, **ngw-*
* liquid + glide i.e. **rw-*, **rj-*, **lw-*, **lj-*
* liquid + nas i.e. **rn-*, **rm-*, **rng-*, **ln-*, **lm-*, **lng-*
* glide + nas i.e. **jn-*, **jm-*, **jng-*, **wn-*, **wm-*, **wng-*
* glide + liquid i.e. **jl-*, **wl-*, **jr-*, **wr-*

Of this family of branching onset constraints, there is ONE that can be violated by certain lexical items: the glide plus liquid combination **wr-* (= */ʋr/). Since this onset consists of a labiodental fricative and a liquid—two segments with the same degree of sonority—it would in principle be prohibited by the SSG (Booij 1995:35). But Van Dale Groot Woordenboek lists 17–20 distinct stem forms with initial /ʋr/.[26]

wraak	'revenge'	Bad
wraddel	(> *wrat*) 'neck flab'	Bad
wrak	1. 'wreck' 2. rickety, ramshackle'	Bad
wrang	1. 'sour, acid' 2. 'unpleasant, nasty' (> *wringen* 'squirm (mouth)')	Bad/Sense
wrat	'wart'	Bad
wreed	'cruel, harsh'	Bad
wrijven/wreef	'rub (INF/PAST SG.)'	Sense
wregelen	'to twist, to intertwine'	
wrensen	'to contract (the lips); to whinny (stallion to mare)'	
wrielen	'young bird's soft calling for food (INF) (onomatopoeic)'	
wriemelen	'wriggle, squirm, fiddle with (frequentative)'	Sense
wrikken	'to lever/prize/scull with effort'	?Sense
wringen	'to wring/squirm/wrench with effort'	Sense
wroeten	'root (up), rout'; 'rout (frequentative)'	
wroeging	'remorse'	Bad
wrocht	'to wind' (frequentative of Middle Dutch *wrigen*)	Sense

TABLE 17. Dutch lexical items with onset /ʋr-/.

Observe that many of these forms belong to the semantic types Sense or Bad.[27] In other words, a violation of the onset constraint **wr-* appears to be matched with an expressive interpretation. But is the violation of the onset constraint **wr-* also a productive strategy to interpret unknown word forms? This question is addressed in the following experiment.

Experiment: The semantics of Dutch nonwords

HYPOTHESIS: the violation of the constraint on homorganic liquids, and the violation of the constraint **wr-* are productive strategies to make sense of unknown word forms.

PREDICTIONS: (i) forms violating the constraints will be given an expressive interpretation, and (ii) forms not violating the constraint will be assigned an unmarked, prototypical interpretation.

MATERIALS

Two sets of Dutch nonwords were compiled: a test set of structurally complex items violating the constraint on two identical homorganic liquids or the constraint */ʋr-/, and a control set of the structurally simple counterparts of these. To avoid a bias for either a nominal or verbal interpretation of the nonwords, I included both stems and

[26] Actually, *Van Dale* (1999) contains 95 lemmas with initial /ʋr/. However, the number of distinct stem forms in this set can be reduced to 17–20 (the exact number depends on the analysis of certain forms). The remaining 75 to 78 forms are not stems but (i) compounds (*wraak* 'revenge'-wraakactie 'action of revenge'), (ii) derivations (*wraak* 'revenge'- *wraakbaar* 'deserving revenge') or inflections (*wringen* 'wring, wrench (INF)'-*wrong* 'wring, wrench (PAST, SG)'). In addition, in the list of 95 lemmas, some alternative forms go back to the same base form (*wraak~wrake~wraken~wreken~wrok*; *wreef~wrigen~wrijven*; *wregelen~wriggelen*; *wrevel~wreed*; *wringen~wrong~wrongel*; *werken~wrocht*; *wroegen~wroeging*; *wroeten~wroetelen*.

[27] *Wraddel*, *wregelen*, *wrensen*, *wrielen*, and *wrocht* are not currently used anymore. *Van Dale* represents them as historically related to Sense/Bad words in Middle Dutch.

derived forms with the suffix *-en*. The derived forms could be interpreted as either noun or verb, because there are two homophonous suffixes *-en*, one to derive plural nouns and one to derive infinitive verbs. The entire set of nonwords consisted of 111 items (See Appendix D for details).

(10) Nonwords used in the experiment on the semantics of Dutch nonwords
   a. 37 words with 2 identical homorganic liquids l-l, r-r (19 stems, 18 derived forms)
   b. 35 counterparts of the words in (a) with an initial l/r, and a final consonant /p, t, k, s/ (18 stems, 17 derived forms)
   c. 14 words with a /ʊr-/ onset (7 stems, 7 derived forms)
   d. 13 counterparts of the words in 10c with simple onset /n,h/ (7 stems, 6 derived forms)
   e. 12 counterparts of the words in 10c with simple onset /w-/ (6 stems, 6 derived forms)

The subjects were seven undergraduate students of linguistics. They were presented with a printed list of the 111 nonwords, in a randomly mixed order, headed by the written instruction: 'Indicate which meaning you find most suitable for the following nonsense words'. The list had the following four response categories (in this order):

(11) Response categories
   1. Action—to stand for the canonical or prototypical semantics of a verb
   2. Sound—to stand for a Sense type verb, i.e. a verb with expressive semantics
   3. Object name—to stand for the canonical, prototypical semantics of a nominal
   4. Epithet—to stand for an item of the Name and/or Bad type, i.e. a word with expressive semantics

RESULTS AND DISCUSSION

All the subjects filled out the entire list independently of each other and gave one response per nonword. Nonwords that were given the interpretations Action or Object were assigned unexpressive semantics. Nonwords with the interpretations Sound or Epithet were assigned expressive semantics.

It turned out that the derived forms with *-en* were almost uniformly classified as Action. This result is particularly striking when compared with the results of the classification of the stem forms, which show a great deal of variation (see Appendix D for details). It seems that in almost all cases, the subjects interpreted the derived forms as infinitive verbs and classified them with the canonical verbal meaning 'action'. This indicates that the semantic classification of the derived words was mainly based on the presence of the suffix *-en*, and not on the structural properties of the stem. Because the present experiment is not concerned with the semantics of suffixes like *-en*, I excluded the forms derived with *-en* from further statistical analysis: only the semantic classifications of the stem forms were statistically analyzed. The results are summarized in Tables 18–20.

SEMANTICS/FORM	COMPLEX (/l—l/ or /r—r/)	SIMPLE (/l—stop/ or /r—stop/)	
Expressive	(S 32 + E 17) = 49	(S 10 + E 10) = 20	69
Other	(A 3 + O 81) = 84	(A 3 + O 103) = 106	190
Totals	133	126	259

$\chi^2 = 14.56$; df = 1; $p < 0.001$ (one-tailed)
A = Action, O = Object name, S = Sound, E = Epithet

TABLE 18. Semantic classification of stems with /l—l/ or /r—r/ versus /l—stop/ or /r—stop/.

Applying a chi-square test gave the result $\chi^2 = 14.56$, which is significant at 0.001 level. This indicates that nonwords that violate the constraint on homorganic identical liquids are much more liable to be assigned an expressive interpretation than forms that do not violate the constraint. This suggests that the similar correlation observed for the real words in Table 16 is probably significant as well. The results of the experiment also indicate that the positive correlation between complex forms and expressive semantics is a productive, synchronically used strategy to make initial sense of unknown word forms,[28] and not just an inert, statistical pattern of the lexicon due to historical factors, for example.

The second set of nonwords was stem forms violating the constraint */ʋr-/ and their simple counterparts. Table 19 shows the result of the semantic classification of these nonwords.

SEMANTICS/FORM	/ʋr—/	/h—/ or /n—/	TOTALS
Expressive	(S 1 + E 19) = 20	(S 2 + E 10) = 12	32
Other	(A 4 + O 25) = 29	(A 0 + O 37) = 37	66
Totals	49	49	98

$\chi^2 = 2.96$; df = 1; $p < 0.10$ (one-tailed)
A = Action, O = Object name, S = Sound, E = Epithet

TABLE 19. Semantic classification of stem forms /ʋr—/ versus /h—/ or /n—/.

Applying the chi-square test gave the result $\chi^2 = 2.96$, which is significant at 0.10 level, a 10% likely result of chance. Though this is much less significant than the results of Table 18, one can still observe the tendency that a word that violates the constraint */ʋr-/ has a greater chance to be assigned an expressive interpretation than a word that obeys the general phonotactic constraints of Dutch. This suggests that the similar correlation observed for the real words in Table 17 is probably significant as well—another effect of the productive strategy Dutch speakers employ to interpret structurally marked forms as having some marked meaning.

The third set of nonwords with the simple onset *w*- (/ʋ/) was contrasted with the structurally marked words violating the constraint */ʋr-/. The results are presented in Table 20.

SEMANTICS/FORM	/ʋr—/	/ʋ—/	TOTALS
Expressive	(S 1 + E 19) = 20	(S 2 + E 11) = 13	33
Other	(A 4 + O 25) = 29	(A 1 + O 28) = 29	58
Totals	49	42	91

$\chi^2 = 0.95$; df = 1; n.s.
A = Action, O = Object name, S = Sound, E = Epithet

TABLE 20. Semantic classification of stem forms /ʋr—/ versus /ʋ—/.

Applying a chi-square test gave the result $\chi^2 = 0.95$, and this is not significant. In other words, words with a complex onset /ʋr—/ are not semantically more expressive than words with a simple onset /ʋ—/. Recall that Table 19 showed that the contrast between /ʋr-/ and a simple nasal or fricative onset is significant. The lack of contrast between /ʋr—/ and /ʋ—/ may therefore lie in the nature of the Dutch /ʋ/. As a labial-dental fricative, it is phonetically and typologically more marked than segments like /n/ or /h/. In other words, in Table 20 we could be comparing one set of marked forms with another set of marked forms. For the moment, I leave this issue open, and simply

[28] Other OCP effects are reflected in language games, morphological derivation, and in the interpretation of complex words. See Frisch et al. 1997:6, and the references given there.

conclude that the use of a *wr-* onset in contrast to simple *h/n-* onsets appears to be semantically driven, while the semantics of words with a simple *w-* onset do not differ from those with complex *wr-*.

In conclusion, the lexicon shows positive correlations between expressive semantic categories and violations of certain structural constraints of the language, and this pattern was observed in existing lexical items (§§2.1, 2.2, 3.1, 3.2). The experimental results reported in this section indicate that it is also an active strategy that speakers employ to make sense of nonwords.

4. CONCLUSIONS AND DISCUSSION. I considered seven classes of items in the lexicon of Kambera and Dutch that show a positive correlation between complex form and expressive semantics. I observed several crosslinguistic regularities, which can be formulated as hypotheses for further research.

1. Sense words (*tweet, blop*, etc.) pattern together semantically and structurally with names and with 'bad' words.

2. Expressive items violate relatively more structural constraint(s) of a language than do nonexpressive items.

3. The formal complexity of expressives involves the violation of language-particular constraints, which are of two basic types: (a) constraints concerned with the linking of form and meaning (maintaining semantic transparency) and (b) constraints on structural contrasts between linguistic elements (either preserving structural distinctness or enforcing structural simplicity).

Observations in line with these hypotheses have been made about expressive elements elsewhere: see Clynes 1995, 1998 on Balinese (Austronesian); Klamer 2000b, 2001 on Kambera, Ilocano, West Tarangan, Tetun, Javanese, and Malay (Austronesian); Anttila 1976 on Estonian and Finnish (Finno-Ugric); Voeltz & Kilian-Hatz 2001 on the African languages Hausa, Zulu, Ewe, Wolaitta, Didinga, and Ciluba; on the Australian languages Jaminjung, Warrura, Gooniyandi, and Gunin/Kwini; and on Quechua; and Bartens 2000 on some Atlantic Creoles. However, still lacking are detailed quantificational data on the lexicons of individual languages (Balinese is an exception).

My findings have the following consequences for the organization of the lexicon:

(i) The lexicon contains not only arbitrary signs but also words whose structure is semantically driven. Languages may have small classes of such words (Dutch is an example of such a language), but there are also languages like Kambera with large classes of expressives. In Kambera at least one out of four lexical items is expressive: approximately 10% of the Kambera lexical roots are ideophones, which are all semantically expressive (§2.1), while 30% of the Kambera lexicon is made up of (C)VCVC roots, about half of which are expressive (§3.1).

(ii) The lexicon contains different strata, but the boundaries between those strata are vague rather than categorical. Expressive items are only gradually, not categorically, distinct from core lexical items. Such gradual differences can be expressed in terms of variable degrees of semantic and formal complexity of lexical items.

(iii) The fact that the patterns of form-meaning matching discussed here are noncategorical implies that the lexicon also contains core lexical items that share the complex formal characteristics of expressives, as well as expressives with a simple structure. My prediction is that such items will always be a minority in their class.

(iv) Some subclasses of the lexicon are symbolic, others iconic. We have seen that the expressive class is iconic in the sense that complex form matches with complex semantics. Additional corpus analyses on lexicons of individual languages can perhaps

establish more iconic lexical classes. For instance, the hypothesis could be that words with a prototypical verbal meaning (Action) and/or a prototypical nominal meaning (Object name) have a structurally simpler form than words with meaning X or Y.

(v) Experimental evidence on Dutch nonwords indicates that the lexicon contains patterns of nonarbitrary form-function mapping that reflect productive strategies to make sense of nonwords. The logical counterpart of this experiment would be an experiment that investigates the opposite direction of the form-meaning mapping, by testing the forms that speakers of Dutch associate with particular meanings. The hypothesis would be that the meanings Action and Object name are associated with nonwords that are structurally simpler than the nonwords that will be associated to the meanings Sound and Epithet.

(vi) If indeed the matching of form and meaning is preferably iconic, a newly coined referential element will never be formally more complex than the most complex expressive in a language, and a new expressive element will never be simpler than the most simple referential element of that language (complexity being defined in terms of the violation of language-specific constraints). In other words, a language disprefers to create expressives with simple shapes and referential items with complex shapes.

Appendix A: frequencies of initial and root-internal consonants in CVCV(C) Kambera roots (Data based on Onvlee 1984)

Only underived CVCV(C) root forms with unmarked vowel segments have been counted.

Frequency of initial consonants:

t = 326, p = 297, r = 294, ngg = 264, l = 251, h = 233, mb = 218, m = 213, k = 197, nd = 198, d = 177, ng = 174, y = 172, b = 141, n = 132, nj = 125, j = 116, ny = 89, **total** = 3,617.

Conclusion: in terms of frequency and feature make-up, the unmarked initial consonants are /t, p/.

Initial consonant /b/, second consonant varies:

b-b = 33, b-l = 30, b-ngg = 25, b-t = 21, b-k = 19, b-h = 18, b-d = 10, b-nd = 10, b-nj = 8, b-n = 7, b-ng = 7, b-w = 5, b-mb = 4, b-p = absent, b-j = absent, b-y = absent, b-ny = absent, b-w = absent, **total** = 197.

Initial consonant /p/, second consonant varies:

p-l = 32, p-r = 31, p-k = 30, p-p = 27, p-t = 25, p-h = 24, p-d = 22, p-nd = 17, p-b = 14, p-ng = 14, p-nj = 11, p-n = 10, p-ngg = 10, p-j = 10, p-w = 4, p-ny = 3, p-m = 3, p-mb = 3, p-y = absent, **total** = 288.

Initial consonant /n/, second consonant varies:

n-n = 15, n-d = 13, n-h = 11, n-t = 10, n-k = 9, n-ngg = 7, n-p = 7, n-r = 7, n-l = 6, n-m = 6, n-nd = 6, n-mb = 5, n-nj = 5, n-b = 4, n-j = 3, n-ng = 3, n-w = 1, n = ny = absent, n-y = absent, **total** = 119.

Initial consonant /r/, second consonant varies:

r-k = 29, r-t = 26, r-r = 25, r-b = 24, r-mb = 22, r-ngg = 20, r-nd = 17, r-p = 17, r-h = 16, r-d = 15, r-m = 12, r-ng = 12, r-nj = 11, r-j = 8, r-w = 8, r-n = 6, r-y = 1, r-ny = absent, r-l = absent, **total** = 267.

No initial consonant, second consonant varies:

ø-ng = 4, ø-r = 3, ø-nd = 3, ø-j = 2, ø-k = 2, ø-t = 2, ø-ny = 1, ø-p = 1, ø-h = 1, ø-b = absent, ø-d = absent, ø-m = absent, ø-n = absent, ø-ngg = absent, ø-mb = absent, ø-nj = absent, ø-w = absent, ø-y = absent, **total** = 19.

Conclusion: the unmarked second consonants are /p, t, h/. Note: Frequent roots with two identical consonants were suspected to show harmony effects. To establish the unmarked second consonant, such roots were not considered. If harmony effects are disregarded, /p, t, h/ occur in all roots in approximately the same relatively high frequency. I take this to indicate that these segments are the unmarked root-internal consonants. The liquids /l, r/ are of high frequency in some roots but low in others, so for lack of unambiguous data I decided not to include them among the unmarked segments. Finally, note that the prenasalized segments cannot be treated as one class (marked or unmarked)—their markedness should be established separately for every prenasalized segment.

Conclusion: the unmarked (1st, 2nd) consonants in Kambera roots are /p, t/.

Appendix B: Kambera marked root forms

1. Test set of Kambera CVCVC roots

The translations as given in Onvlee 1984 were the source for the semantic categories of the words in the test and control sets. But as these translations are in Dutch and may therefore not be accessible to everyone, I have provided brief English translations below. (P = interpreted by the present author as polysemous form; H = interpreted as homophonous form, see note 24).

KAMBERA	TRANSLATION	SEMANTIC TYPE
padang	field, open space	
pahang	pair, form a pair	
pahar	field, terrain	
paindah	to bounce, glance off, ricochet (off)	
paindal	idem	
pakang	to tie tightly	
pandak	short, stocky	
panjang	belong together, gather together	
pawang	to tend (a herd), to keep watch over	
pedah	to smoothen, to pave	
pedang	not be upright; with bended head (P)	Sense/Bad
pekat	thin; too small (P)	Bad
pelar	to smoothen, make a surface even	
pelit	to rub, grate, chafe	
pelung	to wriggle, writhe, wrap around	Sense
pelur	bullet	
penang	to crush, pound, mash (corn)	
pendang	to tie, to fasten	
pengat	emaciated, wasted, weakened, impaired	Bad
pepang	tool to wind thread used for weaving	
pepar	to tilt over, not remain straight, to stagger, wobble	Sense/Bad
perang	to fell, to cut (tree, shrubs)	
pidih	to clasp	
pijar	to knead, pinch, squeeze, crush	Sense
pijik	to bend hand backwards at wrist (while dancing), to bend knee inwards, be knock-kneed	Sense
pikat	be cripple, to totter, stagger, toddle	Sense
pikir	1. turn one's head away out of unwillingness; 2. think (<pikir 'think', Indonesian) (H)	
pikul	weight unit (62.5 kg)	
pílik	to wag (tail), to flap, to flutter	Sense
pílit	to move along the side of a terrain, to not traverse it	
pímbung	to wane (rice)	
pinang	'what's-his-name'	
pínding	thin wall of bamboo or wood	
pingat	strong, powerful, robust	
pinjang	1. to borrow (<pinjam 'borrow' Ind.) 2. continuous, incessantly (H)	
pipit	follow the border, walk along the edge	
pirang	when	
pirih	small kind of parrot	Name
pirung	touch briefly	
pitak	press with the hand, massage	
pitar	intelligent, smart	
pitik	1. a drill, to drill. 2. kind of illness (H)	
pitil	pick up between thumb and forefinger	
piting	kind of bamboo	Name
pobul	young and round, chubby, plump	Sense
pobur	cracking sound of breaking wood	Sense
podah	wipe off, clean	
pohak	1. trod on, trample 2. to clear, purify, sift (H)	

KAMBERA	TRANSLATION	SEMANTIC TYPE
pojur	to squirt (out), spurt	Sense
pokar	to cut down, fell, to process (wood)	
pokat	a trap, to catch in a trap	
popang	be larger than, to rise above, to stick out above	
popar	strong winds and storm with high seas	
porak	to snap, crack	Sense
poruk	movement of approaching or attacking something	Sense
pubak	soft, subdued sound, e.g. bubble, gurgle	Sense
púbuk	half, incomplete, not whole, not real, untrue	Bad
pubur	sound of breaking pottery	Sense
pudah	finish, clear out, clear away	
pudang	close eyes tightly	Sense
puduk	to kiss	
puhak	to shift, to cleanse	
púhuk	bald, without hair or feathers	
puhur	to squirt, spurt	Sense
pujil	remove skin, e.g. foreskin	
pujur	to squirt, spurt	Sense
pukal	untie, let loose, take away/off	
pukang	to be full; to stuff, cram oneself	
	(rude word for 'to eat') (P)	Bad
pukat	(put out) large net to catch sea animals	
pukul	to miss, to not touch or hit; mutilated, numb (P)	Bad
púkul	to hit, tool to hit something	
pulak	to uncover, to strip, to bare	
púlih	kind of tree	Name
pulur	make bald, shave (off)	
punang	1. splinter 2. be filled with (H)	
pundil	swinging movement	Sense
punduh	hop, skip, jump	Sense
púnduk	door jamb; stab with a weapon; wooden pen (P)	
pungal	untie, make loose, fell	
punggur	short, round, stocky, with squat body	Sense
púnggur	large quantity, mass, heap (stones)	
punggut	short, round, stocky, with squat body	Sense
punjil	turn around	
púpat	remove, take away	
pupuh	swollen, bloated	Sense/Bad
pupuk	sopping sound	Sense
púpur	have a sour, surly, grumpy face	Sense
puput	to blow	
purak	kind of pleating	
púruh	to tug, to jerk; rub, massage, scrape (P)	Sense
puruk	to fling, to sling	Sense
purut	flap one's wings, to flutter	Sense
púrut	1. strip, to stem (berries, leaves)	
	2. furrowed, shrivelled (face) (H)	Sense
putal	1. moisty ball or lump (soil, dough)	Sense
	2. become lumpy, to clot, to curdle	
putar	turn, turn round, turn back	
pútir	turn round, turn back	
putuk	thudding, flopping sound of falling down	Sense

2. Control set of unmarked CVCV roots

KAMBERA	TRANSLATION	SEMANTIC TYPE
paha	take as a wife; partner, part of a pair	
pahi	1. confirm, strengthen 2. space in between (H)	
paki	1. good, attractive	
	2. wear clothes (<Ind. pakai 'wear, use') (H)	

KAMBERA	TRANSLATION	SEMANTIC TYPE
paku	(hit a) pen, nail; sound or movement of touching by hitting (P)	Sense
pala	cross, traverse, pass	
palu	hit, instrument to hit with	
panga	space in between, slit	
pangi	embrace, enclose	
papa	one of a pair (partner, adversary, opponent)	
papu	cheek	
para	female genitals	
pari	1. kind of rice 2. ray (fish), tail of ray, weapon (H)	Name
paru	half; side of a dice (P)	
pata	break something	
Pati	male name	Name
patu	four	
peka	announce, express verbally	
peku	possible, good, well done	
pela	later, in the future	
pelu	comb of rooster; vine with fruit as red as a rooster's comb (P)	
pepa	push	
pera	at the side of, besides	
peri	rub, rub off	
pihu	seven	
piku	100,000	
pila	(have a) scar; male name (P)	Name
pili	take away, remove	
píli	tilted	
pingi	stem, source, reason	
pingu	in pingu-pànga 'everywhere, all over the place'	Sense
píngu	to know, to be able to	
pipi	cheek	
pípi	1. smoothen 2. kind of tree 3. kind of bird (H)	Name
pira	how many	
piri	slanting, sloping	
piti	take	
pohu	squeeze; mixed with, amidst of (P)	
poki	blind	
poku	clapping, thudding sound	Sense
pola	stem, source; reason (P)	
polu	1. long (reed, prawn). 2. part of loom (H)	
pongu	wound, wounded	
poru	mixed	
pota	break, split	
puha	throw down	
puhi	put in/on	
puhu	1. navel, middle. 2. movement of entering (H)	Sense
puki	sound/motion of swiftly moving feet	Sense
puku	cracking, snapping sound (rope, back)	Sense
pula	button; ear knob (P)	
puli	let go, set free	
púli	be identical; come together (P)	
púnga	very pale (face)	
pungu	pole, stem, arm, hand	
pupu	kind of poisonous snake	Name
púpu	sign marking the border, surround	
puri	making disorderly, irregular movements/turns	Sense
púri	1. bitter, sour, acid; facial movements when eating something bitter or sour 2. knot of thread (H)	

KAMBERA	TRANSLATION	SEMANTIC TYPE
puru	descend	
puta	hold tightly, hold back	
púti	1. wind. 2. rich (H)	
putu	1. take up, sort out 2. thudding, smacking sound of falling down (H)	Sense

APPENDIX C: KAMBERA MARKED ROOT-INITIAL CONSONANTS

1. Test set of CVCV roots with marked initial consonant ny

KAMBERA	TRANSLATION	SEMANTIC TYPE
nyabu	smacking sound of pig eating	Sense
nyaki	sound of crunchy chewing	Sense
nyama	to chew	
nyanga	secure; ready	
nyanyi	in a small amount, a little bit	Bad
nyapa	ripped, in shreds, tattered	
nyara	chase, chase away	
nyawa	power, strength, energy	
nyeli	worm	Name
nyimba	be blocking the way	Bad
nyiwa	flesh with seeds (in melon, pumpkin etc.); to peel such fruit (P)	Name
nyobi	bolt down food (dog)	
nyola	walk with large steps, stride	Sense
nyolu	eat with smacking lips, to feast on	Sense
nyomba	1. sturdy, robust 2. wallow in mud (H)	Sense
nyonga	1. stupid(ity), foolish(ness), fake 2. put/sit down alternating one after the other (H)	Bad/Sense
nyonya	1. worn out (e.g. baskets, clothes) 2. 'Madam' (<Indonesian nyonya 'Madam') (H)	
nyonyi	hit, thrash, flog; exhausted, worn out (people) (P)	Sense
nyora	1. protrude 2. teacher's wife (<Indonesian nyora 'teacher's wife') (H)	
nyuka	support	
nyúlu	1. k.o. jellyfish 2. tilt, slant, cant, distort (H)	Name
nyunju	go straight on, follow on	
nyura	1. utter, speak out 2. blemish, ailment (H)	Bad/Sense
nyúru	upcoming movement of sea water at high tide	Sense
njadi	be able to, succeed (Ind. loan)	
njaka	deficient, insufficient	Bad
njaki	kind of oyster	Name
njala	wrong, bad	Bad
njanga	1. branch of a tree 2. attend, look after (<Ind. jagar 'look after') (H) (cf. njangga)	Name
njangga	attend, look after (cf. njanga)	
njanji	promise (Ind. loan)	
njara	horse	Name
njata	very large number	
njata	ancestral name	Name
njati	kind of tree (<Ind.: jati)	Name
njawa	Java	Name
njepa	change, exchange, reimburse, compensate	
njeri	beard, (hair, fringes) hanging down	
njibi	smash to pieces, splinter	Sense
njídi	hobble, limp, walk with a limp	Sense
njika	(crop) consumed, devoured entirely by animals (pigs)	
njíku	(walk with a) limp	Sense
njili	be exhausted, tired	Sense
njilu	replacement, reimbursement	

KAMBERA	TRANSLATION	SEMANTIC TYPE
njima	large sea shell fish	Name
njini	motion of penetrating soil; level(ed) (P)	Sense
njingi	motion of looking sidewards; look after someone (P)	Sense
njingu	back to front, inside out; crooked, false, fake (P) (only in compound njingu njànga)	Bad
njínji	deviating, diverging from the norm; be insecure (P)	Bad
njípa	crossed	
njípu	pass, exceed, surpass; trespass, breach the law (P)	Sense/Bad
njiru	thundering noise (only in compound njiru njàra)	Sense
njiwa	earring	
njobu	bay, cove	
njodi	motionless spinning (of, e.g., a top)	Sense
njonga	space in between houses, side of house	
njongu	depth, cove, hollow	
njubu	sharp, pointed	
njuda	sleepy	
njuka	support, prop	
njulu	change skin (snake, shrimp); become young again (P)	
njúlu	wander, roam	Sense
njunga	sit motionless, aimless, workless	Sense
njúngu	swarm	Sense
njunja	movement of soft, tender material (breast, fat)	Sense
njunju	watery, wet (pap, pulp)	Sense
njura	prop, hold up (water, planned marriage)	Bad
njuru	expert (<Ind. juru 'expert')	
njúru	wet, damp (sand on beach, excrements)	Sense/Bad

2. Control set of CVCV roots with initial plain stops t and p

KAMBERA	TRANSLATION	SEMANTIC TYPE
tutu	touch (on), be on target, be true/correct	
túta	be near, guard	
títa	race, gallop	Sense
tata	kind of wild chicken	Name
tetu	kind of sea fish	Name
tota	be erect	
toti	calm, quiet	
taka	forbid, hold back	
taku	spoon, spoon out	
teki	take as wife	
tika	1. kneel, crouch 2. too much, in abundance 3. almost (H)	Sense
tiki	1. utter 2. almost (variant of tika) (H)	
tiku	sound of knocking, creaking, cracking	Sense
tíku	1. sheath (of sword or knife) 2. head (H)	
toka	drive up by hitting (esp. buffalo)	
toku	1. ridge of roof 2. shop (<Indonesian toko 'shop') (H)	
tuka	pull tight, stretch, harness	
tuki	1. spotted, black and white 2. require, demand (H)	
túki	round, spherical, whole	
tuku	1. mortar (for pounding pinang) 2. to forge 3. throw at (H)	
túku	higher than surroundings, not even or flat	
tahi	1. bright (of color) 2. frying pan (H)	
tehu	stunted, dwarf-like	Bad
tohu	with disgusting odor	Bad
tuhu	soft farting sound	Sense
pata	break something	
Pati	male name	Name

KAMBERA	TRANSLATION	SEMANTIC TYPE
patu	four	
piti	pick up/take something	
puta	hold tightly	
púti	roll rope	
paki	1. nice, attractive 2. clothes, wear clothes (<Ind. pakai 'wear, use') (H)	
paku	(hit a) bolt, pin, nail	
peka	proclaim, announce	
peku	good, well-arranged	
piku	100,000	
poki	blind	
poku	1. sound of thudding, clapping	
	2. capital (to merchandise) (H)	Sense
puki	movement of racing feet; 'running for it' (P)	Sense
puku	sound of snapping rope or back bone; trot (of a horse) (P)	Sense

Appendix D. Dutch nonwords

Set of nonwords used in the experiment on the semantics of Dutch nonwords, with the semantic classification given in response. (Note that there is an unambiguous spelling-to-pronunciation mapping in the Dutch orthography of the items in this set.)

1. Structurally complex words with two identical homorganic liquids (*l-l*, or *r-r*)

a. 19 stem forms = singular noun, 1st person plural verb inflection
laal, *laul*, *leel*, *leil*, *leul*, *liel*, *lil*, *loel*, *lool*, *luil*, *rar*, *raur*, *reir*, *reur*, *rier*, *rir*, *roor*, *ror*, *ruir*
Semantic classification: 3 Action, 32 Sound, 81 Object name, 17 Epithet

b. 18 derived forms with *-en* = plural noun, infinitive verb form, 1/2/3 person plural verb inflection
lalen, *laulen*, *lelen*, *leilen*, *leulen*, *lielen*, *loelen*, *lolen*, *luilen*, *rarren*, *rauren*, *reiren*, *reuren*, *rieren*, *rirren*, *roren*, *rorren*, *ruiren*
Semantic classification: 119 Action, 6 Sound

2. Structurally simple counterparts of the words in 1, words with initial *l*/*r*, and final *p*, *t*, *k*, or *s*

a. 18 stem forms: *laap*, *laut*, *leet*, *leik*, *leup*, *liek*, *lit*, *look*, *luip*, *rak*, *raut*, *reip*, *reup*, *ries*, *rip*, *roop*, *rop*, *ruip*
Semantic classification: 3 Action, 10 Sound, 103 Object name, 10 Epithet

b. 17 derived forms: *lapen*, *lauten*, *leten*, *leiken*, *leupen*, *lieken*, *litten*, *loken*, *luipen*, *rakken*, *rauten*, *reipen*, *reupen*, *riezen*, *rippen*, *ropen*, *ruipen*
Semantic classification: 115 Action, 4 Object name

3. Structurally complex words with *wr-* onset

a. 7 stem forms: *wraut*, *wreit*, *wreut*, *writ*, *wroot*, *wrot*, *wruit*
Semantic classification: 4 Action, 1 Sound, 25 Object name, 19 Epithet

b. 7 derived forms: *wrauten*, *wreiten*, *wreuten*, *written*, *wroten*, *wrotten*, *wruiten*
Semantic classification: 49 Action

4. The structurally simple counterparts of the words in (3), words with simple onset *n* or *h*

a. 7 stem forms: *heut*, *hoot*, *naut*, *neit*, *nit*, *not*, *nuit*
Semantic classification: 0 Action, 2 Sound, 37 Object name, 10 Epithet

b. 6 derived forms: *hoten*, *nauten*, *neiten*, *nitten*, *notten*, *nuiten*
Semantic classification: 42 Action

5. The simple onset counterparts of the words in 3, words with a simple onset *w-*

a. 6 stem forms: *waut*, *weit*, *weut*, *woot*, *wot*, *wuit*
Semantic classification: 1 Action, 2 Sound, 28 Object name, 11 Epithet

b. 6 derived forms: *wauten*, *weiten*, *weuten*, *woten*, *wotten*, *wuiten*
Semantic classification: 42 Action

REFERENCES

Anttila, Raimo. 1976. Meaning and structure of Finnish descriptive vocabulary. (Texas Linguistic Forum 5.) Texas.

Bartens, Angela. 2000. Ideophones and sound symbolism in Atlantic Creoles. Helsinki: Annales Academiae Scientiarum Fennicae.

BERLIN, BRENT. 1994. Evidence for pervasive synesthetic sound symbolism in ethnozoological nomenclature. In Hinton et al., 76–93.
BOOIJ, GEERT. 1995. The phonology of Dutch. Oxford: Clarendon Press.
BOOIJ, GEERT, and ARIANE VAN SANTEN. 1998. Morfologie: de woordstructuur van het Nederlands. Amsterdam: Amsterdam University Press.
CHASTAING, MAXIME. 1964. Nouvelles recherches sur le symbolisme des voyelles. Journal psychologique normale et pathologique 61.75–88.
CHASTAING, MAXIME. 1965. Dernières recherches sur le symbolisme vocalique de la petitesse. Revue philosophique de la France et de l'étranger 155.41–56.
CHILDS, TUCKER G. 1994. African ideophones. In Hinton, et al., 178–99.
CLYNES, ADRIAN. 1995. Topics in the phonology and morphology of Balinese [based on the dialect of Singaraja, North Bali]. Canberra: Australian National University dissertation.
CLYNES, ADRIAN. 1998. The emergence of the marked: Interactions of phonology and semantics in Balinese. University of Brunei Darussalam, MS.
CLYNES, ADRIAN. 2000. Phonological structures and expressiveness: The role of iconicity in 'the emergence of the marked'. Proceedings of AFLA 7, ed. by Marian Klamer, 15–32. Amsterdam: Free University, Department of Linguistics.
DE GROOT, A. W. 1963. De interjectie. Studies op het gebied van het hedendaagse Nederlands [Acta Linguistica Rheno-Trajectina], ed. by A. W. de Groot and H. Schultink, 13–22. The Hague: Mouton de Gruyter.
DE HAAS, WIM, and MIEKE TROMMELEN. 1993. Morfologisch Handboek van het Nederlands. The Hague: SDU Uitgeverij.
ELDERS, STEFAN. 2001. Defining ideophones in Mundang. In Voeltz & Kilian-Hatz, 97–110.
FRISCH, STEFAN. 1996. Similarity and frequency in phonology. Evanston, IL: Northwestern University dissertation.
FRISCH, STEFAN; MICHAEL BROE; and JANET PIERREHUMBERT. 1997 [1995]. Similarity and phonotactics in Arabic. Bloomington, IN and Chicago: Indiana University and Northwestern University, MS.
FRISCH, STEFAN, and BUSHRA ZAWAYDEH. 2001. The psychological reality of OCP-place in Arabic. Language 77.91–106.
FUDGE, ERIK. 1970. Phonological structure and 'expressiveness'. Journal of Linguistics 6. 161–88.
HAESERYN, W.; K. ROMIJN; G. GEERTS; J. DE ROOIJ; and M. VAN DEN TOORN. 1997. Algemene Nederlandse Spraakkunst. 2nd revised edn. Groningen: Martinus Nijhoff.
HAIMAN, JOHN. 1994. Iconicity. The encylopedia of language and linguistics, ed. by R. E. Asher and J. M. Y. Simpson. Oxford: Pergamon Press.
HAMANO, SHOKO. 1998. The sound-symbolic system of Japanese. Stanford: CSLI Publications.
HEESTERMANS, HANS. n.d. [1990]. Luilebol! Het Nederlands scheldwoordenboek. Amsterdam: Thomas Rap.
HINTON, LEANNE; JOHANNA NICHOLS; and JOHN J. OHALA (eds.) 1994. Sound symbolism. Cambridge: Cambridge University Press.
JACKENDOFF, RAY S. 1997. The architecture of the language faculty. Cambridge: MIT Press.
JONKER, J. C. G. 1906. Over de eind-medeklinkers in het Rottineesch en Timoreesch. Bijdragen tot de Taal-, Land- en Volkenkunde 59.263–343. Leiden: Royal Institute of Linguistics and Anthropology (KITLV).
JOSEPH, BRIAN D. 1994. Modern Greek *ts*: Beyond sound symbolism. In Hinton et al., 222–36.
JOSEPH, BRIAN D. 1997. On the linguistics of marginality: The centrality of the periphery. Chicago Linguistic Society 33.197–213.
KITA, SOTARO. 1997. Two-dimensional semantic analysis of Japanese mimetics. Linguistics 35.379–415.
KLAMER, MARIAN. 1998. A grammar of Kambera. Berlin/New York: Mouton de Gruyter.
KLAMER, MARIAN. 2000a. How report verbs become quote markers and complementisers. Lingua 110.69–98.
KLAMER, MARIAN. 2000b. Austronesian expressives and the lexicon. Proceedings of AFLA 6, ed. by Catherine Kitto and Carolyn Smallwood, 201–19. Toronto: University of Toronto.

KLAMER, MARIAN. 2001. Expressives and iconicity in the lexicon. In Voeltz & Kilian-Hatz, 165–81.
KROTT, ANDREA; R. HARALD BAAYEN; and ROBERT SCHREUDER. 2000. Analogy in morphology: modelling the choice of linking morphemes in Dutch. Nijmegen: University of Nijmegen, MS.
MATTENS, W. H. M. 1984. De voorspelbaarheid van tussenklanken in nominale samenstellingen. De nieuwe taalgids 77.333–43.
MIDDELKOOP, P. 1950. Proeve van een Timorese grammatica. Bijdragen tot de Taal-, Land- en Volkenkunde 106.375–517. Leiden: Royal Institute of Linguistics and Anthropology (KITLV).
MITHUN, MARIANNE. 1982. The synchronic and diachronic behaviour of plops, squeaks, croaks, sighs and moans. International Journal of American Linguistics 48.49–58.
ONVLEE, LOUIS. 1984. Kamberaas (Oost-Soembaas)-Nederlands Woordenboek. Dordrecht: Foris.
PEIRCE, CHARLES S. 1965. Collected papers of Charles Sanders Peirce. Cambridge, MA: Harvard University Press.
PIERREHUMBERT, JANET. 1993. Dissimilarity in Arabic verbal roots. North Eastern Linguistic Society 23, 367–81.
PIERREHUMBERT, JANET. 1994. Syllable structure and word structure. Papers in laboratory phonology 3: Phonological structure and phonetic form, ed. by Pat Keating, 168–90. Cambridge: Cambridge University Press.
SELKIRK, ELISABETH O. 1982. The syllable. The structure of phonological representations, ed. by Harry van der Hulst and Norval Smith, 337–84. Dordrecht: Foris.
SERENO, JOAN A. 1994. Phonosyntactics. In Hinton et al., 263–75.
SPENCER, ANDREW. 1991. Morphological theory. Oxford: Blackwell.
UHLENBECK, E. M. 1949. De structuur van het Javaanse morpheem. Bandoeng: Nix & Co.
Van Dale Groot Woordenboek der Nederlandse Taal. 1999. 13th revised edn., ed. by G. Geerts and H. Heestermans. Utrecht/Antwerpen: Van Dale Lexicografie.
VAN DER HULST, HARRY, and MARIAN KLAMER. 1997a. The uneven trochee and the structure of Kambera roots. Dam Phonology [HIL phonology papers 2], ed. by Marina Nespor and Norval Smith, 39–57. The Hague: Holland Academic Graphics.
VAN DER HULST, HARRY, and MARIAN KLAMER. 1997b. The prosodic structure of Kambera roots and words. Proceedings of the seventh international conference on Austronesian linguistics, ed. by Cecilia Odé and Wim Stokhof, 105–23. Amsterdam: Rodopi.
VAN DE TOORN, M. C. 1981a. De tussenklank in samenstellingen waarvan het eerste lid een afleiding is. De nieuwe taalgids 74.197–205.
VAN DE TOORN, M. C. 1981b. De tussenklank in samenstellingen waarvan het eerste lid systematisch uitheems is. De nieuwe taalgids 74.547–52.
VAN DE TOORN, M. C. 1982a. Tendenzen bij de beregeling van de verbindingsklank in nominale samenstellingen I. De nieuwe taalgids 75.24–33.
VAN DE TOORN, M. C. 1982b. Tendenzen bij de beregeling van de verbindingsklank in nominale samenstellingen II. De nieuwe taalgids 75.153–60.
VOELTZ, F. K. ERHARD, and CHRISTA KILIAN-HATZ (eds.) 2001. Ideophones. Amsterdam: John Benjamins.
WALLI-SAGEY, ELISABETH. 1986. On the representation of complex segments and their formation in Kinyarwanda. Studies in compensatory lengthening, ed. by Leo Wetzels and Engin Sezer, 251–95. Dordrecht: Foris.

Faculty of Arts
Leiden University
Dept. of SE Asia and Oceania
PO Box 9515
2300 RA Leiden
The Netherlands
[m.a.f.klamer@let.leidenuniv.nl]

[Received 7 March 2000;
first revision received 20 November 2000;
second revision received 26 September 2001;
accepted 30 September 2001]