

NOUN CLASS, GENDER, AND THE LEXICON–SYNTAX–MORPHOLOGY INTERFACES : A COMPARATIVE STUDY OF NIGER-CONGO AND ROMANCE LANGUAGES

Alain Kihm (CNRS, Laboratoire de Linguistique formelle, Paris)

1. Introduction

Not all grammatical categories are created equal as far as universality is concerned. For instance, Number, i.e. the formal indication of whether one or more than one token of a given entity concept is being referred to, is an excellent candidate universal, inasmuch as no natural language seems to be devoid of the category, even though there is variation as to its applicability (on this, see Chierchia 1998). Gender, on the other hand, appears a poor candidate. Consider a rudimentary gender system like that of English : noun phrases denoting interestingly sexed organisms (i.e. human beings and pets) are masculine or feminine ; all others are neuter.¹ The allotment is made perceivable only by the choice of a coreferential singular pronoun, *he*, *she*, or *it*. Even such a cognitively simplistic (and realistic) system is entirely absent in a language like Turkish where all noun phrases are pronominalized as *o* ‘she/he/it’, irrespective of their denotation. And Turkish is by no means an isolate : many languages all over the world are like it in this respect.

Traditional lore about gender has long reflected this state of affairs : when marked in any way, gender was either considered to be as in English, i.e. the linguistic expression of the obvious and obviously vital sexual division of the animate and, through mythology, part of the inanimate world as female, male, or nil ; or it was as in French, namely arbitrary, not to say absurd – why *la table*, but *le cable* ? – but for that portion of the vocabulary denoting sex-relevant entities.² In both cases it was uninteresting from the serious linguist’s point of view, except for the agreement phenomena it may trigger (on this and much more, see Corbett 1996). The cognitive trivialness and/or arbitrariness of gender was taken as the reason why so many languages do without it altogether.

In recent years, a shift away from this tradition has occurred. Numerous studies have appeared which aim to deal seriously with gender and related categories as a classification device necessarily reflecting some central property of human cognition as it expresses itself in the language faculty (see, e.g., Allan 1977 ; Serzisko 1982 ; Craig 1986 ; Aronoff 1994 ; Croft 1994). Most recently (to my present knowledge) Aikhenvald (2000) gave us a thoroughgoing typological review of “noun categorization devices” in a vast array of languages, the least merit of which is not that it includes a host of references to little accessible works, because of their age or the language in which they were written, viz. Russian or German. The present work wishes to contribute to this revival, albeit from a somewhat different perspective, as will appear.

Motives for doubting the old vulgate are indeed easy to find. Sex-based gender systems, of the English “rational” or French “arbitrary” type, are certainly not the only ones in existence. In many languages in various families (Niger-Congo, Caucasian, Sino-Tibetan, Oceanic, Australian, Amerindian of all families, etc.) nominal items are formally divided by diverse means, according to criteria that have to do either with “natural” categories such as being a human (of either sex), or a plant, or an animal, or a dangerous thing, or with descriptive properties of the denoted object, like being elongated, or flat, or liquid, and so forth.³ Noun classes and classifiers are the names for what these languages present. And, exactly as in sex-based systems, irrelevancies – that is, currently incomprehensible allotments – are widespread, although never to the point of entirely obscuring some rationale for the division.

There seem to be no good reason, therefore, for not considering gender, noun classes, and classifiers as diverse expressions of a fundamental faculty of the human mind, namely classification, understood as the spontaneous and irrepressible need we have of pigeonholing all things in the world in order simply to *see* them, what Aronoff (1994 : 181, fn. 3) describes as the “inherent human desire for order, however ungrounded the particular order may be” (also see Lakoff 1986).

It is not my intention, however, to speculate more than has already been done on the cultural and biological foundations of the classificatory faculty and activity. All I wanted to suggest by the preceding remarks is that a linguistic phenomenon so rooted in such a basic faculty of mind is hardly likely to be uninteresting to linguists who view language as a mental organ whose design is supposed to be driven by the need optimally to interface with cognitive faculties (see Chomsky 1998).

What really interests me is the *grammatical* function of Class, as I will designate the global expression of classification however it is realized, a sorely neglected subject, I feel, due perhaps to too much focusing on the extralinguistic correlates. Such neglect is all the more deplorable as Class is a particularly apt topic for comparative grammar, precisely because of its uncertain status as far as universality is concerned. Why should so many grammars avail themselves of such an often cumbersome system, while many others apparently ignore it ? Does it mean that Class as a category is indeed different from universal Number (or Tense), or is it that we do not know yet what purpose it really serves and that, as a consequence, Class’s non-universality is only an appearance ?

In one of the most extensive surveys of the issue to date, viz. Corbett (1996), the author assumes the widespread view that Class (what he calls Gender) *qua* grammatical category only exists as agreement. Portuguese or Swahili have Class because, e.g., attributive and predicative adjectives must agree for it with the noun they modify or are predicated of. Therefore, a language that formally divides its nominals according to Class criteria without triggering agreement of any sort does not have Class. Interestingly enough, such a language does not seem to exist : I know of no language where Class is overtly marked *on the noun itself* without some agreement ensuing – whereas languages where Number is marked and does not trigger agreement are easy to find.⁴ Of itself, this is an interesting observation. Furthermore, the fact (let’s assume it is one) that the theoretically conceivable Class-without-agreement type of language is not attested rather shows, it seems to me, that agreement cannot be used to define or explain Class, precisely because it is a necessary correlate of it. To conclude from entailment to identity is a logical fallacy typical of structuralist views that hold that relations (here, agreement) are more real than what they relate. Generative grammar takes the opposite view, namely that elements or features are real, and relations such as agreement are abstractions on feature matching or non-matching.⁵

I will follow a rather narrow path, but one that will take us, hopefully, onto wider grounds. As mentioned earlier, Class systems can be considered to fall out into three types : (sex-based) gender systems consisting in two or three genders, with Romance as a typical representative ; noun class or multiple (> 3) gender systems, as exemplified by many Niger-Congo languages ; and numeral classifier systems, as in Chinese (see Aikhenvald 2000 for a different taxonomy which the present one crosscuts and simplifies, but does not contradict).⁶ I will leave the latter mostly out of consideration. What I intend to do is to compare the first two systems in order to try and make sense of two empirical generalizations that I find significant, even though they have been given little notice so far.

First, there is the observation that Class exponents of gender systems are suffixes, whereas those of noun class systems may be suffixes or prefixes.⁷ Actually, the distinction one must draw is more subtle : languages with gender systems belong to the inflectional type where functional morphemes are suffixed to and fused with the roots. In contrast, noun class languages manifest the agglutinative typology where functional morphemes are merged with the roots which they may either precede or follow ;⁸ in either case, they look

more like proclitics or enclitics than affixes.⁹ This seems to be the true generalization that stands in need of an explanation (see Dixon 1986 for a study of typological correlates that largely anticipates my own efforts).¹⁰

The second empirical generalization is better known, but just as unexplained as the first one (but see Renault 1987). It is that Class exponents, both gender and noun class, are always combined with Number morphemes when overt, i.e. non-singular. For gender languages, we can take Italian *case* /cas-e/ ‘houses’ as an exemplar, and Manjaku *ito* /i-to/ ‘houses’ for noun class languages.¹¹ Both /-e/ and /i-/ are syncretic morphemes where Class (respectively feminine gender and noun class 7 – see below) and Number (plural) cannot be told apart. As for the corresponding singulars, *casa* and *kato*, it seems we may assume that /-a/ and /ka-/ are pure Class exponents without a Number feature. I will return to all this at length, as it raises many questions.¹²

Those are genuine grammatical generalizations insofar as none of the semantico-conceptual properties that may be associated with Class force them into existence. Therefore, they must proceed from a grammatical property of Class. This property, I will argue, ought to be deduced from the basic fact – all too often held to be so self-evident as to hardly merit discussion – that Class is a property of nouns as opposed to verbs, to a degree that sets it apart from the other so-called “phi-features” Number and Person.¹³ The hypothesis I will try to support, then, is that Class’s grammatical function is as a nominalizer, a *n* category similar to verbalizing *v*. More precisely, I will argue that Class is the or one possible content of the *n* head already proposed in the literature (see, e.g., Marantz 1997), and that the differences observed as to its position and degree of fusion with the root are related to how contentful that head is. Whether several *n*’s should be assumed, as is being currently envisaged for *v* (see Arad 1999 ; Doron 2000), is a related issue which I will also consider.

The study will be conducted within the global framework of Distributed Morphology (see Halle & Marantz 1993 ; Noyer 1997 ; Marantz 1997 ; Embick & Noyer 1999). That is to say, I adopt the notion that syntax operates on abstract feature bundles (roots) drawn from the lexicon, then delivering the outcome of its operations to the autonomous morphological component where vocabulary items, i.e. overt realizations of the feature bundles, are inserted and further worked upon (also see Aronoff 1994 for a different, but compatible architecture). On the other hand, *contra* Embick & Noyer (1999), I will keep to the assumption that Kayne’s (1994) Linearization Correspondence Axiom (LCA) is fully valid, crucially implying that lowering (rightward movement) is disallowed and adjunction is to the left across the board. It is not my purpose to make a general argument against these particular proposals of Embick and Noyer’s, who accept lowering and rightward adjunction in syntax as well as in morphology. I will be satisfied with showing that such anti-LCA operations are not needed to account for the empirical generalizations stated above.

These generalizations come from comparing gender systems like those of the Romance languages with noun class systems as exemplified by Manjaku. Since the latter is in all probability much less well known than the former, I will first present it in some detail.

2. Manjaku noun classes

I chose Manjaku rather than the Bantu languages usually adduced when it comes to noun classes for two reasons. First, it is a language for which I have first-hand data (see Kihm & Gomes 1988 ; Kihm 1998). Secondly, the syntactic correlate of having noun classes, i.e. agreement, is limited to the DP in this language, which simplifies the presentation. In addition to my own data, I will rely on Doneux (1967, 1975) and Buis (1990).

As noted already (see fn. 11), Manjaku belongs to the Atlantic subfamily of the Niger-Congo family, more precisely to the Bak subgroup of this subfamily (see Doneux

1975).¹⁴ Actually, Manjaku comprises a cluster of closely related languages that include, *inter alia*, Mankanya (see Trifkovic 1969), Pepel, Cur, and Bok. The variety I will deal with is Bok, spoken to the North of Bissau, which has vehicular currency among all (ethnic) Manjakus, although not among Mankanyas, the latter's language being markedly different, at least in vocabulary. I will continue to use "Manjaku" as a convenient denominator, however.

Manjaku used to be dubbed "Bantoid" because its noun class system looks very much like that of Bantu languages. It can be schematized as follows, adapting Bantuist traditional presentation :

(1)	1 na-	2 ba-
	1a a-	
	3 u-	4 ngë-
	5 bë-	6 m-
	7 ka-	8 i-
	9 pë-	10 kë-
	11 ndë-	
	12 tsë-	
	13 dë-	

In principle, each row represents a singular-plural pair. 1a is a subset of 1 that includes only kinship terms and is paired with 2 as well. 11 is used to derive diminutives from roots of any other noun class, and it has no plural. Similarly, 12 and 13, locative in meaning and extremely limited in use, have no associated plural noun classes. I now give examples for each noun class and its paired plural: *na-kiëj* (1) / *ba-kiëj* (2) 'thieve(s)'; *a-nin* (1a) / *ba-nin* (2) 'mother(s)'; *u-ndali* (3) / *ngë-ndali* (4) 'cat(s)'; *bë-calam* (5) / *m-calam* (6) 'wild mango tree(s)'; *ka-to* (7) / *i-to* (8) 'house(s)'; *pë-kës* (9) / *kë-kës* (10) 'eye(s)'; *ndë-bus* (11) 'little dog'; *tsë-ko* (12) 'place'; *d-i* (13) 'in, at'.¹⁵

These examples show an idealized view of the system where each pluralizable noun class is paired with one and the same plural noun class, and the semantic correlates are fairly transparent : 1(a)/2 are for human beings ; 3/4 for animals ; 5/6 for plants ; 7/8 for artifacts ; and 9/10 for body organs. The fact is that those correlations hold for a great number of nouns. But it is also a fact, and a well-known one, that they do not hold for another great number of nouns.¹⁶ To take but a few examples, we find nouns denoting instruments in 3/4 (e.g., *u-ndink* / *ngë-ndink* 'machete(s)'; *u-mbanj* / *ngë-mbanj* 'knife/knives'); many nouns denoting natural phenomena in 3 or 5 (e.g., *u-ru* 'smoke', *u-futs* 'wind', *u-nu* 'day', *bë-rëm* 'night', *bë-ruä* 'fire') or even 7 (e.g., *ka-nkuël* 'twilight', *ka-borar* 'dew', *ka-mpalambamb* 'thunder'); and many nouns denoting artifacts in 9 (e.g., *pë-caa* 'basket', *pë-bank* 'dyke'). In addition, 3 hosts a number of nouns referring to abstract entities (e.g., *u-bon* 'hunger', *u-lemp* 'work') as well as language names (e.g., *u-jakin* 'Wolof language') including the word for 'language' itself (*u-jipan* / *ngë-jipan*).

The pairings indicated in (1) are not fully regular, either. For instance, 1 *na-päts* 'child' has a 3 plural *u-päts* 'children', not 2 /**ba-päts*/. Nouns in 9 denoting artifacts regularly have plurals in 8 rather than 10 (e.g., *i-bank* 'dykes', *i-caa* 'baskets', *pë-ndog* / *i-ndog* 'stick(s)', *pë-ngare* / *i-ngare* 'gun(s)').¹⁷ Mass nouns, which normally have no plural, appear in several noun classes as shown by *bë-ruä* (5) 'fire', *i-tuj* (8) 'saliva', *m-lik* (6) 'water' – although the latter's membership is probably due to the fact that 6 seems to be the preferred noun class for liquids (see *m-cäm* 'palm oil', *m-kër* 'oil', *m-ne* 'pus', *m-taw* 'milk' – but *pë-ñaak* 'blood' is in 9).

Unique (or "lexical") class membership is a property of nouns.¹⁸ This means that verbs, prepositions, and adverbs lie outside the noun class system. This proposition is straightforwardly true for the first named : since agreement in Manjaku does not exceed the

boundaries of the DP, as we shall see presently, verbs in this language never bear any marking as to the noun class of their arguments – and verbs as such are immune to Class categorization, for reasons I will try to clarify later on. It must be qualified, however, for prepositions and adverbs. For one thing, basic prepositions are few in Manjaku, perhaps limited to *di* ‘in, at’ and *ni* ‘with’ (also ‘and’) – and even the first item may be analysed as /d+i/, i.e. noun class 13 plus a deictic morpheme (see below).¹⁹ Other relations for which English uses simple prepositions are expressed through PPs consisting of *di* and various nouns (e.g., *di bë-run* ‘in front, before’, *di u-fets* ‘at the back, behind’, *di ruäts* ‘on top, on’). As shown, those nouns may pertain to noun classes 5, 3, or 13, perhaps others.²⁰ Likewise, adverbs fall out into two categories, depending on whether they exhibit no noun class membership (e.g., *takël* ‘yesterday’, *ntsäri* ‘today’, *faan* ‘tomorrow’, *kotiës* ‘a little’) or whether they are nouns generally pertaining to noun class 3 (e.g., *u-cäk* ‘firstly’).²¹

Besides nouns, whose noun class is fixed, and classless items there is a third category, namely agreeing items that include adjectives, some numerals, determiners, and a particular type of pronominals. Since this section aims only to description, I will content myself with giving examples for each type. Adjectives in Manjaku do not constitute a unified syntactic category, as they are represented by agreeing items when in attributive function, but by (stative or quality) verbs when predicative :

- (2) ka-to ka-mak
7-house 7-big
the/a big house
- (2a) i-to i-mak
8-house 8-big
(the) big houses
- (3) Ka-to a mak
7-house Pro be-big
The house is big
- (3a) I-to a mak
8-house Pro be-big
The houses are big

As evidenced in (2) and (2a), attributive adjectives follow the head noun they modify and agree with it in noun class (and number).²² (3) and (3a) illustrate an already mentioned feature, namely absence of agreement beyond the DP. The *a* morpheme glossed Pro is a kind of predicate marker that is obligatorily present with 3rd person subjects (compare *M mak /you(sg) be-big/* ‘You are big’), although it does not agree with them. I will have more to say about it in the following.

Numerals from ‘one’ to ‘ten’ are also split. From ‘one’ to ‘four’ they behave like adjectives :

- (4) ka-to ka-lole
7-house 7-one
one house
- (4a) i-to i-tëb / i-wants / i-bakër
8-house 8-two / 8-three / 8-four
two / three / four houses

From ‘five’ to ‘ten’, in contrast, they are nouns with their own class membership, although they still follow the noun they quantify, as if they were in some kind of apposition to it :

- (5) i-to kà-ñän / p(ë)-aaj / k(ë)-uäs / u-ntaja²³

8-house 7a-five / 9-six / 10-eight / 3-ten
 five / six / eight / ten houses

Determiners are /-i/ ‘this’, /-un/ ‘that’, and /-lon/ ‘some, a certain’ and they appear rightmost in the DP :

(6) ka-to ka-mak k-i / k-un/ ka-lon²⁴
 7-house 7-big 7-this / 7-that / 7-some
 this / that / some big house

Finally, there is the intriguing root /ko/, probably to be analysed as a fully unspecified (pro) nominal. I use this label because /ko/ is able to enter all noun classes, thus giving rise to items that function either as kind terms or as anaphoric, classified pronouns. For instance, *na-ko* (1) means ‘he/she’ with specific reference to the person designated as the Topic ;²⁵ similarly, *u-ko* / *ngë-ko* (3/4) ‘it / they’, if the Topic is an animal or a non-material entity ; *ka-ko* / *i-ko* (7/8) if it is a concrete object. On the other hand, *u-ko* / *ngë-ko* may be used as a noun with the meaning ‘animal’ (e.g., *u-ko w-i* ‘this animal’) ; likewise, *ka-ko k-i* ‘this thing’ (also ‘this piece’), and the diminutive (class 11) *ndë-ko nd-i* ‘this small thing’. The latter, however, cannot be used as an anaphoric pronoun referring back to anything being described as small ; nor can, it seems, *bë-ko* / *m-ko* (5/6), *pë-ko* (9), and *tsë-ko* (12), which are only kind terms meaning respectively ‘tree(s)’, ‘fruit’ (a mass noun), and ‘place’. I hedge this claim, because there might be some, as yet unexplored, variation or special conditions in this area. Perhaps idealizing somewhat, then, noun class distinctions for anaphoric Topic pronouns seem to reduce to three, viz. 1/2 for humans, 3/4 for animals and non-material entities (and perhaps more if 3/4 is indeed the less specified noun class), and 7/8 for concrete objects. Naturally, given a felicitous discourse context, determined kind terms may always be used to refer to Topic entities for which there is no specific *-ko* pronoun :

(7) Man win bë-mango. Bë-ko b-i a mak
 I see 5-mango 5-KO 5-this Pro be-big
 I saw a mango tree. This tree/it is big
 (7a) Man ya ka-to Jon. Tsë-ko ts-un a wara
 I go 7-house Jon 12-KO 12-that Pro be-nice
 I went to Jon’s house. That place/it is nice

Ko-pronouns only fulfil the subject function and they do not cooccur with *a* :

(8) Man me Jon. Na-ko (*a) wara
 I know Jon 1-KO (*Pro) be-nice
 I know Jon. He is nice

which suggests they compete for the same position (say Spec vP). This accords with the fact that, even though *ko*-pronouns are discourse-linked with a current topic, their use does not imply explicit topicalization. What must be considered here is that *a* is sufficient by itself to provide a sentence with its necessary overt subject (Manjaku not being a null subject language), so that *A wara* ‘She/he/it is nice’ is a perfect sentence. In that case, however, the noun class of the intended subject is grammatically irretrievable ; all one knows is that it is 3rd person and singular or generic. *Ko*-pronouns’ main (or sole) function, therefore, seems to be disambiguation. Perhaps they can be defined as anti-obviative, insofar as they signal that a previously introduced topic has not been switched.

For explicit topicalization (involving left-dislocation) as well as for non-subject functions, another set of 3rd person pronouns is used, based on the root /-ul/ : *n-ul* / *buk-ul* (1/2) ; *w-ul* / *ng-ul* (3/4) ; *k-ul* / *y-ul* (7/8) ; *nd-ul* (11) ; *d-ul* (13).²⁶ Here is an example showing both uses :

- (9) N-ul, na-ko me ul²⁷
 1-3SG 1-KO know 1a-3SG
 As for her/him, she/he knows her/him

It is notheworthy that the same restrictions as to noun class combinability seem to hold with *ul* as with *ko*.

3. The function of noun classes

The semantic and number pairing vagaries just reviewed, for which Manjaku is far from being a signal case, led many authors, especially in the Africanist tradition, to conclude that noun classes as such do not serve any purpose (but see Denny & Creider 1986 for a welcome exception). A function that is often granted them, however, but not “as such”, is that of marking Number. Indeed, the class prefix alternation of, e.g., *ka-to* / *i-to* expresses whether one or a plurality of the denoted object is implied – just as the /-s/ of *houses* does – and since linguists do not know and native speakers cannot tell what *ka-/i-* “means” as a classifying device, it may be good method to assume that the understandable function is the only active one, at least in the present state of the language.

Actually, it is not entirely true that native speakers cannot tell. They have a clear awareness of the blueprint of the system, so to speak, in the sense that if the prefix is *na-/ba-*, then it’s a human being, if it is *u-/ngë-*, then it is likely to be an animal, and so on. Borrowings demonstrate it, that are unfallibly assigned to the noun classes that best suit their denotations. For instance, /mango/, a borrowing from Portuguese, is entered into 5/6 (*bë-mango* / *m-mango*) if it refers to the tree, into 9 if it refers to the fruit (*pë-mango*), following a regular pattern according to which fruits or parts of trees are in 9 (see *bë-jaak* / *m-jaak* ‘palm tree’ vs. *pë-jaak* ‘palm’, *bë-jaam* / *m-jaam* ‘bamboo tree’ vs. *pë-jaam* ‘bamboo piece’). It is that knowledge, moreover, that allows language users, upon hearing, say, *Man win k-ul* ‘I saw it’, to infer that the object of seeing allusively referred to by means of the pronoun *k-ul* is most probably one concrete inorganic thing, even if they may know nothing more about it.

The problem, as is well-known, is that it is no more than a blueprint, valid at a certain level of abstraction, but replete with inconsistencies when one comes down to particulars. For instance, the plural of *pë-mango* (9) ‘mango’ is *m-mango* (6) ‘mangoes’ (and similarly for *pë-jaak*, *pë-jaam*, etc.), so that the distinction of the tree and its fruit or part is erased in the plural. Worse, there are exceptions to the rule of thumb that fruits are in 9, as shown by *bë-kuäle* / *m-kuäle* ‘cola tree(s)’ (5/6) vs. *ka-kuäle* / *i-kuäle* ‘cola nut(s)’ (7/8). In still other cases, the distinction cuts along different lines, as in *bë-rungäl* / *m-dungäl* (5/6) ‘baobab’ vs. *m-dungäl* (6) ‘baobab fruit’ vs. *pë-rungäl* (9) ‘baobab fruit seed’. Similarly, as already noted, it is not predictable whether a noun denoting a concrete object belongs to 7/8 or to 9 with a plural that is generally in 8 rather than 10.

Examples could be multiplied, but they would add nothing to the unescapable verdict : noun classes work poorly as a classifying device. Now, this should not be taken to mean that noun class systems do not express some classificatory faculty of the mind, as they do achieve some reliable classification (insofar as, e.g., Manjaku 1/2 nouns *always* denote humans). But they do so in an unsystematic, far from optimal fashion. The conclusion we shall draw from this observation, however, is not that noun classes are useless hindrances,

the linguistic equivalent of the appendicitis, but rather that their *grammatical* function is distinct from classification as such, although not disconnected from it, as will be seen.

That this function could be Number marking seems highly unlikely, though. Indeed, if it were merely that, why should Number features be required to associate with other, grammatically otiose features ? Or, if noun class is not considered a feature at all, why should Number multiply its exponents so unaccountably ? An interesting standard of comparison in this respect is given by Wolof, a neighbouring language, genetically related to Manjaku.²⁸ In Wolof, noun class exponents are consonants merged with determiners postposed to the head noun (e.g., *kër g-i* /house G-Det/ ‘the house’, where /g/ is the noun class exponent). In Standard Wolof, there are eight noun classes for singular DPs, but only two for plural DPs, one of which is limited to basic nouns denoting humans (e.g. *jigéen j-i / ñ-i* ‘the woman/women’, but *kër g-i / y-i* ‘the house(s)’ like *fas w-i / y-i* ‘the horse(s)’). In Dakar Wolof, however, noun class contrasts are definitely on the wane, so that for many speakers, with of course a lot of variation, the tendency is to use only one noun class in the singular (*b-*) and one in the plural (*y-*). If this tendency goes to completion, then clearly *b-i / y-i* will reanalyse as a postposed determiner, with the consonant alternation uniquely expressing the Number contrast. What this will amount to, then, is for a complex noun class plus Number distinction to be *replaced* by a simple singular vs. plural contrast. Such seems indeed to be the general direction of this type of historical changes (see, e.g., the evolution from Old to Modern English), which proves *a contrario* the grammatical significance of noun class (or Class) contrasts as long as they are represented.

I will now introduce more facts about the Manjaku system, that should help us asseverate the basic function of noun classes and, by extension, of Class. We saw that noun class 9 (*pë-*) includes nouns of rather various semantics : fruits or parts of trees (e.g., *pë-mango* ‘mango’, *pë-jaak* ‘palm’) and artefacts (e.g., *pë-caa* ‘basket’, *pë-ngare* ‘shotgun’), but also body parts (e.g., *pë-konj* ‘finger’, *pë-kaba* ‘buttock’, *pë-kës* ‘eye’) and natural formations (e.g., *pë-bos* ‘sandy hill’). Concerning body part nouns, something must be noted that is not entirely an aside, namely that the Number contrast associated with noun class pairing may be actually more complex than a mere singular vs. plural distinction. Take ‘fingers’, for instance : if the plural refers to a discrete number of fingers that does not usually exceed ten, i.e. the normal number a human being is endowed with, then it is expressed in noun class 10 (e.g., *kë-konj kë-wants* ‘three fingers’) ;²⁹ if it refers to an unknown and/or indefinite number, generic interpretation (fingers in general) included, it is expressed in noun class 8 (*i-konj* ‘fingers’). This shows, at the very least, how inadequate it is to consider 8 as simply being “the plural” of 7. Not only can it be paired with other noun classes (9 and also 5 – see *bë-rëk / i-rëk* ‘river(s)’), but its precise meaning depends on the root it is merged with, since *i-to*, for instance, in contrast with *i-konj*, refers to any plurality of houses.

There is another important function of noun class 9, however, and that is to nominalize verbs to produce the equivalent of English infinitives in Control constructions :

- (10) Na-kiěj a ngal pë-fäm pë-lëman³⁰
 1-thief Pro want 9-break 9-door
 The thief wants to break the door

Notice that *pë-fäm* ‘to break’ keeps its ability to theta- and Case-mark an object.³¹ Noun class 9 “infinitives” also enter a construction that is syntactically and semantically similar to the English *be – -ing* construction :

- (11) A ci tsi pë-ji
 Pro be in 9-laugh
 S/he is laughing

On the other hand, phrases like *pë-fäm pë-lëman* may be used as argument DPs in sentences like (12) :

- (12) Pë-fäm pë-lëman wara-ts^{32 33}
 9-break 9-door be-nice-Neg
 Breaking the/a door is not nice

It is impossible, however, to pluralize *pë-fäm* to denote, say, repeated breakings of the same or several doors.

Such a collection of properties suggests that noun class 9 items derived from verbal roots (or rather roots used in a verbal capacity – see below) constitute a mixed category of nominalized verbs, sharing features with English infinitives and gerunds. The main value for my purpose of this observation about noun class 9, however, is to draw attention to what I claim is the basic function of noun classes, viz. noun formation.

This function appears very clearly looking at the two arguments in sentence (10). It then appears that *pë-lëman* ‘door’ is built on a root /lëm/ meaning ‘to cover’ when used verbally (e.g., *A lëm pë-lik* ‘She/he covered the well’).³⁴ Likewise, *na-kiěj* ‘thief’ derives from /kiěj/ that shows up as a verb in, e.g., *A kiěj pë-lëman* ‘She/he stole the door’. Such a use of noun class 1/2 as an agentive morpheme is fully productive ; other examples are *lam* ‘to swim’ vs. *na-lam / ba-lam* ‘swimmer(s)’ ; *lemp* ‘to work’ vs. *na-lemp / ba-lemp* ‘worker (s)’, and so on. Actually, “agentive” is too precise a label as evidenced by such formations as *na-cën / ba-cën* ‘neighbour’ related to *cën* ‘to be neighbours’, itself associated with *pë-cën* ‘fact of being neighbours, neighbourhood’ ; or *na-fur / ba-fur* ‘madman / madmen’ related to *fur* ‘to be mad’ ; or *a-buk (1a) /ba-buk* ‘child(ren)’ related to *buk* ‘to give birth’. The real meaning of noun class 1/2 would then rather be “person involved in the event or state implied in the meaning of the root”, with much indeterminateness as to the thematic relation of that person with respect to the event or state – insofar as *a-buk* denotes the result of the event, the child, rather than its agent, the mother.

Another thing worthy of note is that, despite appearances, /pë-/ formations as in (10) and (11) cannot readily be analysed as inflectional, in the sense that an English gerund is inflectional. For one thing, noun class 9 is not the only one to form infinitive-like items, as shown by the following Mankanya examples (Trifkovic 1969 : 117, 118, 121) :

- (13) U-ñiing wo tsi u-pay
 3-hyena be in 3-climb
 The hyena is climbing
- (14) I wo tsi ngë-tsiilën
 you be in 4-lie
 You are lying
- (15) Dë bi yitir ba-jaar ba-wajents du ka-tiiban
 I Past meet 2-farmer 2-three in 7-clear the ground
 I had met three farmers clearing the ground

In (13), the root /pay/, with the general meaning of ‘climbing, going up’, is nominalized in noun class 3 which, apart from including most animals, language names, and still other things, is also the class for abstractions in the dual sense of abstract notions (e.g., *u-wejëts* ‘thought’ next to *wejëts* ‘to think’) and of logical abstraction over the extension of a concept (e.g., *u-bon* ‘hunger, famine’ next to *bon* ‘to be hungry’, *u-lemp* ‘work’ next to *lemp* ‘to work’). Apparently, there is no referential difference between (13) and the equally possible ... *wo tsi pë-pay*, in the sense that both describe the same event. In (14), *ngë-tsiilën* is a plural meaning ‘lies’, with a paired singular *m-tsiilën* (5), so that a more literal translation

might be something like ‘You are (deep) in lies’. Again, the same state of affairs could be described, with perhaps some difference in expressivity, as *I wo tsi pë-tsiilën*. Finally, *ka-tiiban*’s closest equivalent is probably French *défrichage*.³⁵ An important function of noun class 7 in all Manjaku languages is thus highlighted, which consists in building “action nouns” comparable to Arabic *maṣdar*’s (e.g., *‘ilm* ‘fact of knowing, knowledge’ on the root *‘lm*), whose semantic (and syntactic) difference from /pë-/ infinitive-like formations, although certainly real, is not easy to assess.³⁶ This point will be taken up again later on.

On the other hand, “infinitivization”, if it may be so called, is not the only function of 9. We already met with several nouns in that noun class denoting fruits (e.g. *pë-mango* ‘mango’), body parts (e.g., *pë-kës* ‘eye’), artefacts (e.g., *pë-caa* ‘basket’), or natural objects (e.g., *pë-bos* ‘sandy hill’ next to *m-bos* ‘sand’), not obviously built on roots having a verbal capacity. Even when the root does have this capacity, the result is not always something we want to describe as an infinitive : see, e.g., *pë-lik / i-lik* ‘well(s)’ vs. *(pë-)lik* ‘to draw water’. (Other formations on the same root are *m-lik* (5) ‘water’ and *ka-lik* (7) ‘fruit juice’.) Shall we consider those instances of /pë-/ different morphemes than the one in (10) ? There seems to be no compelling reason for such a move. A more interesting solution is to view /pë-/ as the same noun-forming morpheme in all cases, for whose general meaning I won’t try to devise a formula at this point. Only consider that a well is indeed something water is drawn from, thus manifesting the same thematic indetermination we already observed with noun class 1/2 (see *a-buk* with respect to *buk* above).³⁷ That an associated verb cannot always be found out, on the other hand, may be either one of those accidents of which languages are customary, or it may be simply due to the not fully explored state of the Manjaku vocabulary. What we have to explain, then, is the apparently more verb-like character of /pë-/ formations in examples like (10). The issue will be tackled in the next section.

Before we come to that, however, there are two more facts I wish to point out, as they clearly demonstrate the inherent noun-forming potential of noun class exponents. One has to do with numerals : as mentioned above, some agree with the noun they quantify, some do not. When counting from one to ten, the agreeing numerals are assigned to class 3 for ‘one’ (*u-lole* ‘1’), to class 4 for the remainder (*ngë-tëb* ‘2’, *ngë-wants* ‘3’, *ngë-bakër* ‘4’), thus quantifying over all possible nominals in accordance with the abstracting meaning of noun class 3/4 (see (5) and fn. 23 for the numerals from five to ten).

The other significant fact is the pronominal use of noun class exponents. In Manjaku (i.e., Bok), as we saw, noun class exponents in this function are merged with /-ko/ when subjects, with /-ul/ when objects. In Mankanya, however, they appear as such in the position where Bok has *a* for all subjects, the latter being restricted to human subjects of noun class 1 (a) in Mankanya. We thus find (Trifkovic 1969 : 102) :

- (16) Ba daan
 2 drink
 They (humans) drink
 (17) Ka-toh ka joot-i
 7-house 7 crumble-Perf
 The/a house crumbled

and so forth.³⁸ They also appear in object position, as in (18) (p. 105) :

- (18) Dë thuuman ka
 I fill 7
 I fill it

where *ka* refers to, e.g., *ka-kana* ‘calabash’.³⁹ Something similar is observed in Manjaku (Bok), but only in the subject position of embedded clauses – see *Man ngal u-bus w-i u*

tsëp /I want 3-dog 3-this 3 leave/ ‘I want this dog to leave’ (lit. ‘I want this dog it leaves’). In the following, I will make full use of this property of noun class exponents.

To sum up this section, evidence was presented to the effect of showing that the basic function of noun classes beyond classification itself – imperfectly ensured in any case – is to form nouns. That noun classes play a role in derivation has naturally been recognized in all studies on the subject (see in particular Mufwene 1980, whose conclusions largely foreshadow my own ; also see Katamba 1993 : 211ff.). My claim, as I hope to show below, goes farther than that, however, and, if correct, will enable us to connect what can be observed in Manjaku with a more general theory of Class. Let me just indicate for the present that there is nothing surprising in the fact that noun formation, in a sense to be clarified presently, and classification should go hand in hand. Indeed, it belongs to the inherent properties of nouns that they denote entities that can (perhaps must) be allotted to different classes of *things*, by virtue of innate and culturally informed cognitive processes, diversely expressed in languages. Events and states, in contrast, seem never to be classified in the same way : for instance, we find languages where nouns denoting dangerous objects belong to a given noun class (see Lakoff 1987) ; but I know of no description of a language where verbs referring to dangerous actions or events (e.g., ‘to wound’, ‘to climb’, ‘to fight’, etc.) bear some specific morpheme marking this character. Verbs are categorized, sometimes overtly, according to stage-setting properties of the denoted processes (whether they are open-ended or not, continuous or not, stative or dynamic, which participant they affect, and so forth), but never, it seems, according to “real” descriptive properties such as being perilous or harmless, typical of this or that entity, etc.⁴⁰ I will refrain from speculating about why it is so, but it seems to be a true and significant generalization related to what deeply separates nouns from verbs, viz. Class (see Aronoff 1994 : 66, who asks the same question in more or less my terms).

4. Noun-forming Class

My purpose in this section is to account for the Manjaku evidence using the framework of Distributed Morphology (DM), and then to try and generalize it at least to gender languages such as Romance (and secondarily Semitic). In so doing, I will also attempt to make sense of the empirical generalizations stated in the introductory section.

As already indicated, one of the basic tenets of DM is that the lexicon consists in roots which lack category as well as a phonological form, and that may be semantically underspecified to varying degrees (see, e.g., Marantz 1997). Roots acquire a category by being inserted in particular syntactic configurations, thus becoming morphemes ; morphemes are associated with phonological features in the vocabulary component of morphology, and may then be called “exponents” (see Embick & Noyer 1999).

Manjaku supports these assumptions rather straightforwardly. Take, for instance, the series *lik* ‘to draw water’, *pë-lik* / *i-lik* ‘well(s)’, *m-lik* ‘water’, *ka-lik* ‘fruit juice’. We might content ourselves with observing we have here four (or five) words that share a common root (in the traditional, descriptive sense of the term) and have some meaning in common, but are not related to each other in any meaningful way. Just four (or five) separate “lexical items” among thousands.⁴¹ If that is the truth, it is a deadend, and a dull one, and we should bow to it only if nothing else works. Let us try something else, then.

Assume that $\sqrt{\text{lik}}$ is a root in the Manjaku lexicon. (The square root sign is there to indicate that the letters/phonemes it flanks are intended as no more than a label.) On the semantic side, that is at the interface of the lexicon with language-external cognitive faculties, it seems to be related to the global concept of “water”, meaning a set of entities and events that are conceived of as bearing a direct relation to “water”, in the same way as

the Arabic root $\sqrt{\text{ktb}}$ is related to the global concept of “writing”, surfacing now as a verb (e.g., *kataba* ‘he wrote’), now as a noun (e.g., *kitaab* ‘book’, i.e. object produced by writing). A very vague characterization, to be sure, but one we have to be content with since we cannot probably do much better in the present state of our knowledge of how concepts are represented in the mind.

Although vague, however, such a characterization must at least aim to be correct. In this respect, there is a question that cannot be eschewed, given the constituency of the $\sqrt{\text{lik}}$ series : why choose “water” as a semantic correlate rather than “well” ? Why, in other terms, assume an implicit definition of “well” as something water is drawn from, rather than of “water” as something drawn from wells ? Numerous, more or less speculative answers can be given to justify the present choice, including, for instance, the greater pregnancy and universality of natural kinds such as “water” over unnecessary artefacts such as wells. A more concrete and, I think, more convincing reason, though, is the presence of *ka-lik* ‘fruit juice’ in the series. A cognitive connexion from “water” to “fruit juice”, on one hand, and to “well”, on the other hand, seems indeed natural enough, whereas if we proceeded from “well” to “water”, then we would need an independent connexion from the latter to “fruit juice”. That is to say, selecting “water” as the semantic pivot allows us to build a more tightly knit network than had we chosen “well” in this function.

The real point of this discussion, I think, is to show that roots are indeed *roots* in a very concrete sense, and that the property and its content can be empirically validated by such evidence as the Manjaku $\sqrt{\text{lik}}$ series (or paradigm). Later on I will attempt to be more precise about what roots are the root of, for which the notion “encyclopaedia” will have to be introduced.

Let us now return to more language-internal issues. Roots are selected from the lexicon to be inserted into syntactic structures, thus becoming heads as shown in (19), where XP is the (obligatory) complement of the head :

(19) [$\sqrt{\text{P}}$ $\sqrt{\text{O}}$ [XP]]

Clearly, (19) is not a well-formed syntactic object, as it has no category allowing it to function as a predicate or an argument. Suppose we want (19) to be a predicate. We can assign it the required category by merging it with a functional projection headed by the functional head ν also selected from the lexicon. Assuming ν to be of the causative variety (see Arad 1999) and the root to be $\sqrt{\text{lik}}$, we turn (19) into (20), where YP is a subject in Spec νP :

(20) [νP [YP] ν^{O} [$\sqrt{\text{P}}$ [$\sqrt{\text{O}}$ *lik*][XP]]]]

The root head then raises-adjoins to ν^{O} , thereby acquiring verbal features :

(21) [νP [YP] *lik*+ ν^{O} [$\sqrt{\text{P}}$ [$\sqrt{\text{O}}$ *t*][XP]]]]

Since ν is not associated with a morphological exponent, the final form is *lik*, i.e. the exponent (or vocabulary item) associated with the root.

The question we have to ask now is : Does (21) account entirely for the meaning of *lik*, i.e. ‘to draw water from a well’, and the fact that it can be used intransitively or with a complement that looks very much like a cognate object, namely *m-lik* ‘water’ ?⁴² The answer seems to be “no”. Given the global meaning we assumed for the root $\sqrt{\text{lik}}$, (21) yields a verb analogous to the English *to water*. Now, there is no necessary reason why *to water* should have the meaning it has in English, “to pour water on (plants)”, rather than the

one it has in Manjaku, “to draw water from (wells)”. Yet, we should be able, if not to predict, at least to formalize such diverse meanings in different languages.

A possible formalization consists, following Hale & Keyser (1993), to derive *lik* from a lexical relational structure (LRS) like (22) :⁴³

(22) [_{VP} V [_{VP} water [_{PP} from well]]]

Incorporating the PP into the abstract verb yields a verb that has the semantics of *lik*. Notice by the way that (22) also accounts for French *puiser (de l'eau)*, related to *puits* ‘well’, which happens to match *lik*’s semantic and syntactic properties quite closely.

The problem, however, is that Hale & Keyser’s theory is not *prima facie* compatible with a view of the lexicon including globally underspecified roots.⁴⁴ Neither does there seem to be room in DM for the lexical syntax Hale & Keyser assume (see, e.g., Harris’s 1999 flow chart of the grammatical organization generally assumed in DM). Moreover, (22) clearly predicts that the final form of the verb must be derived from *well* (as it indeed is in French), not from *water*, as it appears to be in Manjaku. The two theories, then, seem hardly gearable to one another (but see Bouchard 1995 for an approach to the relations of semantics to syntax that might be more immediately congenial to the present attempt).

It is not necessarily so, however. DM’s model of grammar provides for a component distinct from the lexicon and the vocabulary, called the “encyclopaedia”. As the name suggests, it is the component where terms (to use a neutral label) having a denotational relation with language-external reality are given their effective meaning (e.g., that *cat*, in addition to being a noun and a CVC syllable, refers to a purring mammal, or that you water aspidistras, but you don’t water your own body when you take a shower), thus ensuring that they are used not only grammatically, but also felicitously. In the usual model, the encyclopaedia steps in at the output of PF and LF, and it feeds semantic interpretation at the interface of language proper and “external” faculties (see Harris 1999). The assumption, then, is that only fully derived items (exponents) are susceptible to possess encyclopaedic meaning. Intuitively, that is true : only *kitaab* ‘book’ refers to something ; the root $\sqrt{\text{ktb}}$ does not (in Fregean terms, it has meaning, but no denotation). That truth, however, does not necessarily imply that the encyclopaedia must be ordered after all other grammatical operations. In a strictly derivational view of grammar, it does ; but if we adopt a grammatical organization more in terms of parallel processing (in line with, e.g., Brody 1995 and Jackendoff 1997, and perhaps Chomsky 2001), the encyclopaedia may be considered an autonomous component along with the lexicon and morphology.⁴⁵ Under such a conception, what the grammar produces are signs (in the Saussurean sense – also see Sag & Wasow 1999 : 356ff.) which are the embodiment of three interfaced processings : (i) categorization and merger of roots in the syntax ; (ii) morphophonological operations on morphemes ; (iii) semantic characterization. Provided that each component has access to the others – i.e., that the expression of the sign in one component is linked to its expressions in the other components – no ordering is required.

The inner organization of the encyclopaedia has not really been explored, as far as I know. Yet, as it is a grammatical component, we should expect it to house the same kinds of structures and operations we find in the other components. My proposal, then, is that insofar as LRSs such as (22) are formal explicitations of semantic representations, the encyclopaedia is their proper location, provided it is related to lexicon and syntax in the way just sketched. What Hale & Keyser call lexical syntax (l-syntax), I therefore call encyclopaedic syntax (e-syntax), parallel to and interfaced with syntax *tout court* (s-syntax or narrow syntax). My difference with these authors is thus that, for independent reasons, I assume the lexicon to be made up of roots rather than full-fledged lexical items, and that it is these roots s-syntax manipulates. On the other hand, the encyclopaedia must be geared to

the fact that the developed representations it includes may correspond to simple roots (such as $\sqrt{\text{lik}}$ or $\sqrt{\text{shelve}}$, to use one of Hale & Keyser's staple examples) and to simple exponents (*lik*, *shelve*). Hence the justification of e-syntax (e.g., incorporation into V in (22) resulting in compacting the whole phrase into one position) in order to allow for one-to-one association between the components.

Concretely, this means that at the same time that $\sqrt{\text{lik}}$ (the root) is projected into syntax as shown in (20) and (21), it is linked to an encyclopaedic representation like (22) that tells us that behind the verbal morpheme (sign) *lik* that we hear, there is water, a well, and a drawing of the former from the latter. But what about the remark that the surfacing exponent ought to be that of "well" (as in French)? Actually, that would only be a problem if e-syntax fed s-syntax, as Hale & Keyser's l-syntax does. But it does not, being a parallel depiction of the relations associated with the meanings of particular instantiations of the root. Moreover, this is where the global underspecification of roots in a language like Manjaku (perhaps more extreme in this regard than English or French) comes into play. Indeed, given the series that grows from $\sqrt{\text{lik}}$, all the elements that enter the encyclopaedic representation of *lik* 'to draw water (from a well)' turn out to proceed from the same root.

The foregoing discussion seems to have veered somewhat from the issue of noun classes. It hasn't really, however, since its purpose was mainly to present some plausible (I hope) concepts I intend to use in order to tackle that issue.

As we just saw, merging the root $\sqrt{\text{lik}}$ with a *v* projection in syntax and associating the resulting verbal [*lik+v*] morpheme with an exponent (*/lik/*) in morphology, on one hand, and an encyclopaedic structure, on the other hand, gives us a full representation of the Manjaku verb *lik* 'to draw water (from a well)'. Our next task, then, is to account for the other members of the $\sqrt{\text{lik}}$ paradigm, e.g. *pë-lik* / *i-lik* 'well(s)'. In fact, we have no choice of the analysis, given our assumptions: since $\sqrt{\text{lik}}$ is uncategorized, it must be merged with a noun-forming functional category, call it *n*, if it is to be used as an argument, just as it must be merged with verb-forming *v* when a predicate head; */pë-/* is one exponent of *n*; noun class morphemes in general are exponents of *n*.⁴⁶

Basically, this is the solution I will adopt (see Lecarme 1999 for a similar treatment of Somali gender). A number of issues have to be dealt with, though, before we can feel confident it is at least a plausible solution. The first one is that of the exact definition of the $\{n \leftrightarrow /pë-/ \}$ (or any other noun class prefix) item.⁴⁷ Embick & Noyer (1999) distinguish crucially between what they call "functional morphemes" (f-morphemes) and "roots" (lexical items, in more familiar terms). The rationale for the distinction is the notion that for f-morphemes 'Vocabulary Insertion is deterministic and only one choice is possible in any given context' (p. 267). Root insertion, in contrast, is free, depending on the speaker's choice.⁴⁸ The question, then, is whether $\{n \leftrightarrow /pë-/ \}$ (or *pë-* for short) is an f-morpheme or a root.

Considering that *v* is generally defined as a functional category, the answer seems straightforward: *n* is an f-morpheme. Yet, given our characterization of roots and a mere examination of the Manjaku vocabulary, there can be little doubt that *pë-* is the only element uniquely associated with the meaning "well" in *pë-lik* as compared to *m-lik* 'water' and *ka-lik* 'fruit juice', even if the content of the association is far from clear. It follows that, in the context of $\sqrt{\text{lik}}$ being selected, the choice of *pë-* rather than of *m-* or *ka-* is free, depending on whether the speaker intends to talk about a well, or water, or fruit juice. Therefore, *pë-* does not lack descriptive content, and it is a root, as are *m-*, *ka-*, etc.

Actually, what we must say to be accurate is that *n* is a root, i.e. a member of the lexicon, and that to the difference of all other roots, except *v*, it is endowed with a category, namely N. In other words, *n* is *the* noun in the lexicon (and *v* *the* verb). And we must add that *n* is a label for a set of prototypical nouns, or protonouns, each having its own meaning and exponent (and perhaps assuming a theory of prototypes as in Lakoff 1987 and Posner

1986). This points to an important property of noun class systems, evidence for which was given in the preceding section, namely that {n}, the set of protonouns, not only forms nouns from roots, but also imparts the nouns so formed with particular meanings, e.g. Agent (or rather “person involved”) as in *na-lam* ‘swimmer’. That is to say, {n} also fulfils the function of derivational morphemes such as /-er/ in English. Actually, rather than with /-er/, the significant comparison is with *man* in such forms as *madman* (*na-fur*). Each member of {n} is thus associated with a general concept.

The practical difficulty for the analysis is that this concept is not always clearly recognizable in the present state of the language. It is obvious for the *n* whose exponent is /na-/, i.e. “human being” or for the *n* realized as /ts-/, meaning “place, location” (except, probably, in the one exception *tsë-mak* ‘big brother’). In other cases, we have to acknowledge the fact that only part of the mergers of a particular member of {n} with a particular root reveals what we may consider to be the concept inherently linked to that *n*. Here, I am merely repeating what I pointed out earlier, namely that noun class systems lost their optimality in the course of time – if they ever were optimal to begin with. Nevertheless, the fact that meanings must be assigned, even if imperfectly, to the members of {n} confirms their lexical status as protonouns.

To sum up, the following overall characterization of an *n* may be given :

- (23) Lexicon : [N, human]
 ↔ Vocabulary : /na-/ ~ /a-/
 ↔ Encyclopaedia : ((\sqrt{x})(N human))
 ↔ Syntax [_{NP} n° [_{VP} \sqrt{x}]]

I take (23) to be the representation of what I called a “sign”, i.e. the summation of all dimensions defining a given linguistic unit. The encyclopaedic tier of (23) is an informal notation of the fact that the lexical item (root) consisting in the features N and [human] must enter into some semantic relation with another root. The relation is underspecified (Agent with *na-lam* ‘swimmer’, but Theme with *a-buk* ‘child’, given the meaning of $\sqrt{\text{buk}}$ ‘concept of giving birth’ – see above). The basic meaning of the *n* defined in (23), resulting from the interfacing of lexical features and encyclopaedic representation (e-syntax), is thus “person involved in some event or state of affairs”. At the same time, the syntactic specification states that this lexical item is the head of a projection taking the same root targeted in e-syntax as its complement, and the vocabulary shows its exponent to be /na-/ or /a-/, with the morphological property of being a prefix.⁴⁹ As we shall see, this property, far from being idiosyncratic, can be deduced from the characterization.

Finally, the mixed (verbal and nominal) character of /pë/-infinitives (but not, it seems, of /ka/ or /u/ verbal nouns) is easily accounted for in this framework by assuming the root to be successively merged with *v*, and then *n* (see Marantz 1997 for a similar derivation of English *destruction*-type nominalizations, an issue that will be taken up later on). The first merger accounts for the verbal properties of the item (viz. theta- and Case-marking), the second for its NP type.

5. Noun class, gender and Class

We are now equipped to embark on a comparison of noun class systems with gender systems as represented in, e.g., the Romance languages. If the basic function of what surfaces as noun class markers is indeed to make nouns out of uncategorized roots, and given the undeniable relatedness of noun class to gender, as well as the conceptual necessity of finding a function for the latter, a generalization suggests itself : gender is also basically a

noun-forming device. That is to say, the Romance lexicon is a set of uncategorized roots plus the two categorized *n* and *v* elements, just like the lexicons of Manjaku or the Semitic languages (or all languages, we assume). The /-a/ morpheme of Spanish *gata* ‘she-cat’, for instance, thus represents the morphological exponent of one member of the lexical set {*n*}, syntactically realized as the *n*^o head taking the $\sqrt{\text{gat}}$ root as its complement. Only through that merger can the root gain actual denotation and enter a proposition about some possible world.

To that extent, then, noun class and gender systems are alike, parallel instantiations of the classificatory faculty that requires entity-denoting signs (as opposed to event-denoting signs) to be stacked on this or that shelf in the library of the mind. On the other hand, both types of systems are markedly different on a number of counts, as we saw. My task, therefore, is to account for these differences in an illuminating way, if I can, that is by showing them to proceed from minimal variations in the initial, globally common conditions.

The most striking difference is revealed by the foregoing analysis of Manjaku, and it is that noun class signs, such as the one represented in (23), are not only classificatory, but they also are derivational. Gender morphemes never play such a role : in Spanish *nadador* ‘swimmer’ (compare Manjaku *na-lam*), Agent is the exclusive meaning of /-dor/ and the noun is masculine by virtue of not bearing an explicit gender morpheme ; if the swimmer is a woman, *nadadora* is used. Moreover, it never happens that /-dor/ fails to have the Agent (or some related) meaning, whereas there are many cases where /na-/ (or rather $\sqrt{\text{na}}$) does.

This points to somewhat distinct organizations of the lexicons. Actually, the apparent derivational function of noun class signs is a by-product of their being meaningful roots : as already suggested, *na-lam* is not the exact equivalent of *swimmer*, but it designates a human being ($\sqrt{\text{na}}$) involved in the event of swimming ($\sqrt{\text{lam}}$) ; similarly, *pë-lik* is a a thing ($\sqrt{\text{pë}}$) associated with the concept of water ($\sqrt{\text{lik}}$), and it is the encyclopaedia’s role to apprise us of the fact that this particular collocation is to be understood as ‘well’ ; *tsë-ko* means a location ($\sqrt{\text{tsë}}$) associated with a very general concept ($\sqrt{\text{ko}}$) perhaps translatable as ‘being’, hence ‘place’ ; and so forth. Classification, on the other hand, falls out directly from meaningfulness, even though, as noted, in an imperfect fashion, because of the fuzzy or multivalued character of the semantics of noun class signs.

Such a conflation of classification and derivation does not occur in languages like Spanish. On the one hand, we find signs such as Spanish /-dor/ or Italian /-ata/ (e.g., *ombrellata* ‘hit with an umbrella’ – see Ippolito 1999) which seem to be endowed with fairly precise meanings, but do not of themselves partition the noun set into several classes (for a reason to be soon made clear). On the other hand, there are the morphemes /-o/, /-a/, /-e/, and /∅/ acting as word class markers (see Harris 1992, 1999) that do perform such partitioning, since every noun in a Romance language (to the possible exception of French) must overtly belong to one of those classes. The question we must ask in order to further our comparative task, then, is whether these morphemes have meaning and ought to be considered roots like Manjaku noun class signs.

This is a complex and disputed issue. In a significant number of cases where sex of the *denotatum* is relevant /-a/ plainly refers to the female (e.g., *gata* ‘she-cat’, *jefa* ‘woman chief’, *nadadora* ‘woman swimmer’) whereas one of the other morphemes refers to the male or is ambivalent (*gato* ‘tomcat, cat’, *jefe* ‘chief’, *nadador* ‘swimmer’). When sex is not relevant, the same morphemes seem to play no other role than defining word classes (*libro* ‘book’, *puerta* ‘door’, *coche* ‘car’, *pared* ‘wall’). Harris (1999) thus defines three major classes labelled I (/o/), II (/a/), and III (all the rest), plus a fourth class (IV) for exceptional items, making it very clear that the binary feature masculine vs. feminine (M/F) cannot simply overlay this division, since all classes include M and F nouns, although the information about (sex-based) gender must be available in order to select the proper word

class form when there is a choice (e.g., *gato* or *gata*). A further complexity stems from the possible interaction of derivational morphemes with word classes and the M/F feature : for instance, all nouns showing the suffix /ist/ belong to class II (e.g., *linguista* ‘linguist’) irrespective of whether the denoted individual is a man or a woman (this is decided by the choice of the determiner *el* or *la*, *un* or *una* which thus expresses the M/F feature) ; in Portuguese, on the other hand, nouns in /-(a)gem/, while belonging to class III and denoting sexless entities, are all feminine as far as determiner selection and agreement are concerned (see *A linguagem é complicada* ‘Language is complicated’).

Insofar as Romance word classes classify entities, then, they do so according to the criterion of sex or, as I will continue to call it, gender. Perforce such a criterion is only applicable to a subset of actually or mentally existing entities. The consequence of this state of affair, I will assume, is therefore that in Romance, contrary to Manjaku, the classificatory feature (M/F) is divorced from the categorial feature N. The latter is inherent to the word class morpheme, being its grammatical *raison d’être* : word class morphemes form nouns from roots, just like noun class signs do. The former, in contrast, is only linked when gender is relevant given the meaning of the root to which the word class morpheme is attached. Thus, *gata* ‘she-cat’ and *puerta* ‘door’ receive the following representations :

(24) $\sqrt{\text{gat}} - a$ $\sqrt{\text{puert}} - a$
 N(II) F N(II)

Gata is indeed feminine (classificatorily feminine) ; *puerta*, on the other hand, is only class II ; and the same for *gato* when it denotes a tomcat (N(I) M) and *libro* (N(I)). What remains is an entailment to the effect that class I and II nouns whose roots denote relevantly sexed kinds refer to the masculine and feminine members of that kind respectively.⁵⁰ The entailment does not go the other way, however, since not all nouns referring to classificatorily feminine or masculine entities belong to class II or I – see *mujer* ‘woman’ (N(III) F) and *jefe* ‘chief’ (N(III) M).

What about nouns like *linguista* or *acrobata* ‘acrobat’, which seem to run afoul of the entailment, as they are class II and may refer to masculine members of the kind, or *modelo* ‘model’, class I and able to refer to a woman model, as in *esa modelo finlandesa* ‘this(F) Finnish(F) model(I)’ ?⁵¹ There is a ready explanation for the first one, and it is to assume that the word class morpheme is attached not to the root, but to the derivational morpheme /ist/. Not being a denoting root, this morpheme cannot be gender relevant. Therefore, *linguista* is actually like *puerta*, i.e. only class II, not like *gata*. The same account probably extends to *acrobata* with class II also preempted by the /ata/ ending (also see *pirata* ‘pirate’). As for *modelo*, the human denotation of ‘fashion model’ is a fairly recent development from the basic, gender-irrelevant meaning of ‘model’ as an object and/or a notion. Therefore, *modelo* is like *libro*. Nothing, on the other hand, says that class I is incompatible with the F feature, should it be relevant, beyond the word, because the entailment is only valid of the primary combination of a root with a word class morpheme, and the compatibility we observe in *esa modelo* ‘this (woman) model’ is a direct consequence of the global mutual independence of word class and gender.

It follows from this that whenever gender is relevant, but word class of the noun is preempted for some reason (special suffix or ending, metaphorical usage, etc.) so the entailments $I \rightarrow M$ and $II \rightarrow F$ cannot hold, then the M/F feature shows up only on modifiers of the noun, being then divorced from the noun forming word class morpheme. *El linguista* ‘the (man) linguist’ may thus be represented as follows :

(25) $el \sqrt{\text{lingu}} [-ist-a]$
 M N(II)

The determiner in (25) includes a gender feature, and so it does whenever gender is relevant, whatever the word class of the noun (*la gata* ‘the she-cat’, *la madre* ‘the mother’, and so forth). When gender is not relevant, however, the logic of the explanation forces us to assume that no gender feature is present in the determiner and other modifiers, so that *el* is merely class III, not M, in *el libro* ‘the book’ and *la* merely class II, not F, in *la puerta* ‘the door’.⁵² This raises a question as, here as well, there is a measure of mutual independence, this time between the word class of the noun and that of the modifiers : class III nouns, with word class exponent /e/ or /Ø/, have modifiers in class III (e.g., *un mar frío* ‘a cold sea’) or II (e.g., *una pared alta* ‘a high wall’) ; and there is also the case of *la mano derecha* ‘the(II) right(II) hand(I)’. I do not feel bound to integrate these data to my account, however, as they do not seem to me to be crucial for the point I am trying to make. After all, since there are three (of four) distinct classes for nouns, but only two for modifiers, we expect an amount of shifting around in the distribution, sometimes due to discernible factors (e.g., class III nouns ending in /-ed/ are usually feminine, to use the received terminology), but sometimes unaccountable (and *la mano* is a genuine exception).

Important for my purpose, on the other hand, is the proposal I am making that in order to make sense of the Spanish system the classificatorily (i.e. semantically) valued M/F feature must be firmly dissociated from the solely grammatical word class morphemes, even though they may be morphologically fused at the exponence level, and there is a limited entailment from the one to the others. Actually, this dissociation should make us wonder whether M/F is a lexical feature at all. Since it is only linked when gender is relevant, and gender is determined on the basis of the denotation of the whole sign (see *el linguista, la modelo*), it seems more coherent to assume that M/F is in fact an encyclopaedic feature, assigned in this interface component with semantics and not at all in the lexicon.

If this is true, word class becomes a purely functional feature, indeed, whose sole function is to form nouns from roots. Let me give a broad outline of the system thus reached : the Spanish lexicon (not special in this respect) includes roots like $\sqrt{\text{gat}}$ or $\sqrt{\text{puert}}$ and the functional set {n}. This set consists in four members, call them nI, nII, nIII, and nIIIa, the exponents of which are /o/, /a/, /e/, and /Ø/.⁵³ Suppose $\sqrt{\text{puert}}$ is chosen. To be usable as a syntactic noun, it has to be merged with an nP projection having nII as its head. Here there is no choice : it is simply a fact of Spanish that $\sqrt{\text{puert}}$ “goes with” nII (recall Embick and Noyer’s 1999 definition of functional items).⁵⁴ In the encyclopaedia nothing happens as far as gender is concerned : a door is not a gender-relevant entity, therefore the M/F feature is not associated with the encyclopaedic representation of $\sqrt{\text{puert}}$. In morphology, nII is spelled out as /a/ and $\sqrt{\text{puert}}$ as /pwert/, hence the final exponent /'pwerta/ (I will return to what happens between initial syntax and morphology). Should [nII + $\sqrt{\text{puert}}$] be further merged with a DP projection, then word class agreement takes place to the effect that D spells out as /la/.

Suppose now $\sqrt{\text{gat}}$ is chosen. Cats being sexed entities, the M/F feature links to its encyclopaedic representation and a choice has to be made. Let us assume the speaker wants to talk about a she-cat and F is selected.⁵⁵ We can now think of the entailment stated above as of a specific connexion between encyclopaedia and lexicon, such that if F is selected in the former, then nII is activated in the latter, unless some factor more specific than the entailment prevents it. No such factor is present with $\sqrt{\text{gat}}$, so the ultimate result is *gata*, an ordinary class II feminine noun.

It is present, in contrast, in the case of *linguista*, inasmuch as nII associates with the derivational morpheme /ist/ rather than with the primary root $\sqrt{\text{lingw}}$. Note this is not problematic in the present account, since we concluded earlier that in Spanish, to the difference of Manjaku, derivation is distinct from Class. We can therefore consider derivational morphemes to be functional roots that, because of their functional character,

must combine with a non-functional root when they are taken from the lexicon.⁵⁶ Given Williams's (1981) Righthand Head Rule and our assumptions about the syntax of word derivation, we then expect word class to attach to the derivational morpheme (the noun's head) rather than to the non-functional root, as shown in the following initial representation of *linguista* :

(26) [_{NP} nII [_{VP} √*ist* [_{VP} √*lingw*]]]

Noun forming nII attracts the closest root, which adjoins to its left, hence [*ist*+nII] ; the complex functional head thus formed then attracts the non-functional root, hence [*lingw*+ [*ist*+nII]].⁵⁷

A further distinction has now to be made : among derivational roots, some like √*dor* are contentful enough that they have encyclopaedic representations. In its meaning "human agent", √*dor* is thus gender-relevant and a decision has to be taken as to whether M or F is attached to it.⁵⁸ If F, as in *nadadora* 'woman swimmer', the encyclopaedia–lexicon special connexion (or entailment) ensures that nII associates with √*dor*, winning over nIIIa by the Elsewhere Principle, since it is more specified as it is connected to an encyclopaedic feature, which nIIIa is not (see fn. 58). If M is chosen, on the other hand, as in *nadador* 'swimmer (male or unspecified)', the M → nI entailment (yielding ungrammatical **nadadoro*) cannot prevail, because √*dor* is globally associated with nIIIa (a fact to be learned) and nI is not more specified than nIIIa as both may refer to males or to the whole kind (see *gato*).

If the foregoing argument holds, then it implies, *a contrario*, that √*ist* has no encyclopaedic representation – or at least it is not valued as gender-relevant there. It is only associated with nII, as shown in (26). What must have an encyclopaedic representation, in contrast, is the whole sign realized as *linguista* after derivation through syntax and morphology. Recall we assume all components to be interfaced, so not only may syntax feed the encyclopaedia, but it must do so since the meaning of derived nouns is rarely, if ever, fully compositional.⁵⁹ *Linguista* being gender-relevant, which neither √*lingw* nor √*ist* are, the M/F feature links to it. The problem now is that, whatever feature value is chosen, it cannot mark the sign since word class is already assigned.⁶⁰ Therefore, it is merely appended to the sign (hence something like [*linguista*]_{MF}), manifesting itself only through agreement (as in *un linguista americano* 'an American linguist'). With a few slight adjustments, the same account holds for *modelo*.

We are now in a position to appraise the full extent of the difference between Spanish (or Romance) and Manjaku. In Spanish, as we just saw, classificatory gender (M/F) and word class are divorced, the former being (possibly) an encyclopaedic feature, and the latter a purely grammatical (or functional) noun-forming feature. In Manjaku, on the contrary, classification and noun class are not separate, but both are represented in the lexicon as features of particular roots (√*na*, √*ka*, etc.).⁶¹ Moreover, the classificatory criteria differ in intension as well as in extension. Whereas Romance resorts to the masculine vs. feminine contrast, only applicable to a subset of entities, Manjaku supports its classification on representations of natural-conceptual kinds such as humans, animals, plants, abstractions, etc., which are potentially applicable to all conceivable entities. The non-separability of classificatory meaning and functional noun class, I submit, follows precisely from this. That is to say, members of the {n} set in Manjaku are endowed with the inherent capacity to not only nominalize roots, as they do in Romance as well, but to address them to delineated sectors in semantic space. Encyclopaedic representations of such pre-sorted signs (i.e., subsequent to syntactic merger of some member of {n} with a given root) then come as a further narrowing of the signs' partly defined meanings, and also as a check.

I add the latter provision because, as we saw, mismatches will occur. To give just one more, striking example : whereas it seems to be the case that all class 1/2 nouns refer to

humans, the reverse is not entirely true, and there are at least two exceptions, one in class 5, viz. *bë-fetsar* ‘friend’, the other in class 12, otherwise reserved for locative words, viz. *tsë-mak* ‘big brother’. To make matters worse, *bë-fetsar*’s plural is in class 8 (*i-fetsar* ‘friends’).⁶² Sensibly enough, *tsë-mak*’s plural is in class 2 (*ba-mak*). Yet, such discrepancies are of a different order than those we observed in Spanish. In the latter, it is simply the case that the classificatory criterion associated with (but distinct from) noun class does not apply at all to a vast number of roots and derived nouns. In Manjaku, in contrast, the classificatory dimensions inherently linked to noun classes have the potential of always matching the meaning or one of the meanings of the roots they combine with. Sometimes they do not : a friend is not a plant (see above for what may be the basic meaning of class 5) ; nor are several friends artefacts (if that is the meaning of class 8, besides plural). Then we have a mismatch, which encyclopaedia checks when it tells us that, despite being merged with class 5, *bë-fetsar* means ‘friend’, not some plant like *bë-ko* ‘tree’ or *bë-mango* ‘mango tree’. Mismatches, however numerous, do not call the overall classificatory value of the system into question, though. They merely draw attention to the fact that, as already emphasized, the system is imperfect, from various mostly irretrievable and probably accidental causes ; and also, at a deeper level, that the meanings of noun classes like those of Manjaku (or Bantu, or Dyrbal) and their combinations with roots is not something that can be defined from mere observation of the language as we hear it. It would require full knowledge of the representational system shared by the native speakers of the language and/or by those who bequeathed it to them along generations.⁶³ Since such knowledge is unattainable, we must satisfy ourselves with acknowledging that noun classes have meaning, because assuming so much is the best and perhaps only way to make sense of the system, and proposing tentative and no doubt hopelessly partial interpretations on the basis of what we observe.

The basic difference between Manjaku and Romance appears thus to be at the same time limited and profound : limited because both language (UG) varieties share the category I term Class, i.e. the linguistic expression of the mental classificatory faculty that gives rise to the grammatical objects called “nouns” ; profound because the *lexical* expressions of the said category are meaningful roots in Manjaku, but semantically empty functional items in Romance. In other terms, Class morphemes in Romance are indeed *inflectional class* morphemes in the sense of Aronoff (1994), i.e. creatures of autonomous morphology that are linked to gender proper only through particular entailments (or implicational rules) with encyclopaedic knowledge, and to the classificational faculty insofar as they nevertheless impose an order on the entity-denoting (nominal) vocabulary, albeit a largely meaningless one.⁶⁴ Manjaku Class morphemes, in contrast, are direct, even though often obscure, expressions of encyclopaedic knowledge. The typological generalizations pointed to above follow from this, as I will now try to show.

6. The morphosyntax of Class exponents

In this section, I will examine the position of Class exponents relative to the (main) root and also with respect to other elements that may attach to the root, an issue that, although somewhat tangential at first sight, may have far-reaching theoretical consequences, as we shall see. Then I will deal with the fusion of Class and Number.

6.1. The position of Class exponents

Let us return to the syntactic representation of Spanish *gata* ‘she-cat’, viz. :

(27) [_{nP} nII [_{√P} √gat]]

In order to derive the surface form /gata/ from it we only need the minimal assumptions that (i) functional lexical items inserted into syntax are morphologically illegible unless they attach to the non-functional roots over which they have local scope ; (ii) attachment proceeds through attraction of the root by the functional item, so that the former left-adjoins to the latter in accordance with Kayne's (1994) LCA. From (27) we thus derive (28) :

(28) [_{nP} √gat+nII [_{vP} t]]

Stripped of all unpronounceable, (28) – or rather [_{vP} √gat+nII] – is then delivered over to morphology where √gat and nII are associated with their respective exponents, i.e. /gat/ and /a/. Notice there is no necessity in this account to specify /a/ as being a suffix (representing it as, say, /-a/). Suffixation follows directly from the syntactic processing required to build an object fit for morphological treatment, as a consequence of the functional nature of nII. Actually, (28) makes the very strong prediction that functional morphemes may only suffix, and that suffixes can only spell out functional morphemes. Whether it is a true prediction can obviously not be tested in the present study. Counterexamples come readily to mind, such as the Classical Arabic contrast of *'aktabu* 'I (will) write' vs. *katabtu* 'I wrote' where the same feature bundle [1SG] seems to be expressed now as a prefix (/a-), now as a suffix (-tu/). To dispel it one needs to prove that these elements have different lexical statuses – that, e.g., /a-/ is a clitic pronoun, the reduced form of the strong pronoun *'ana* 'I', whereas /-tu/ is a functional complex including Tense.⁶⁵ For my present purposes, however, it is enough to accept the weaker claim that the suffixal property of gender exponents in Romance follows directly from their functional character.

Conversely, therefore, the prefixal property of Manjaku noun class exponents must follow from the fact that they are roots, fundamentally no different from the roots they attach to but for their being categorized as N. Crucial here is the observation that noun class morphemes may be inserted as such, functioning then as pronouns, without restriction in the Mankanya dialect (see (17) *ka joot-i* 'it (e.g., the house) crumbled'), in embedded clauses in Manjaku. This implies that the syntactic structure associated with *ka* in (17) is probably the following :

(29) [_{nP} √ka [_{vP} Ø]]

That is to say, no particular root is selected to merge with the chosen member of {n}.⁶⁶ The structure of, e.g., *ka-lik* 'fruit juice' is the same :

(30) [_{nP} √ka [_{vP} √lik]]

Obviously, (30) is just like (27). What distinguishes them is not structural, but substantial : √ka is a root, whereas nII is functional. Only functional elements have the property Attract. (I take this as a self-evident generalization supported by all theoretical developments since the very inception of generative grammar.) Therefore, no syntactic operation applies to (30) which is delivered as such to morphology.⁶⁷

Naturally, the foregoing account would be circular, had we no independent grounds for our analysis of √ka and nII. This is why I took pains to demonstrate at length the root character of noun class signs in Manjaku, basing my demonstration on semantic criteria : noun class signs in Manjaku are endowed with a denotational potential that Romance gender morphemes lack entirely. If the conclusion is accepted, there is no circularity.

The question we want to ask now is what happens to (27) and (30) once they enter the morphological component. The answer, apparently, is 'Next to nothing'. What morphology receives from syntax, apart from the feature bundles to which exponents get

associated (viz. /gat/ and /a/, /ka/ and /lik/), is the specification that this particular collocation of features constitutes one projection. That is to say, inner brackets are erased (or not read) and outer brackets are interpreted as word boundaries.⁶⁸ (27) and (30) are thus converted to #gat+a# and #ka+lik# respectively and submitted to whatever word forming and phonological processes are required by the language – not many in the present case.⁶⁹

In fact, given the root character of the constituents, the morphological formation of *ka-lik* does not look essentially different from that of compounds such as *pez-espada* ‘swordfish’ (lit. ‘fish sword’) in Spanish, where the first root indicates the general category of the entity, more or less as *ka* does in Manjaku, while the second root specifies the identity. Of course there are differences inasmuch as (i) $\sqrt{\text{pez}}$ is semantically richer than $\sqrt{\text{ka}}$; (ii) *ka-lik* is more word-like than *pez-espada* because *ka* has a clitic property that *pez* has not. Now, clitichood may be a phonological property related to the availability of a CV slot to the left of any root *qua* CV sequence (see Lowenstamm 1996 who presents empirical and theoretical arguments in favour of such a slot as a universal feature of phonological form). *Ka* and all other noun class exponents, being CV objects, will naturally “slide” into this slot.⁷⁰

The “agglutinative” typology of noun class exponence as opposed to the “inflectional” character of Romance gender marking also falls out rather naturally from these considerations. Being non-functional roots trivially merged with other roots to which they cliticize by virtue of general phonological properties, noun class exponents cannot but keep their distinctivity. Recall moreover that both types of roots exist autonomously. Romance gender morphemes, in contrast, because they are functional roots, must fuse with non-functional roots that may not be associated with legitimate word-forms in the language (as is the case with Spanish $\sqrt{\text{gat}}$ which cannot surface unless it is combined with something else).⁷¹

Finally, something must be said of an issue that was mentioned earlier in passing (see fn. 62). It has to do with the existence in Manjaku of nouns such as *bë-fetsar* ‘friend’, showing that noun class roots may combine not only with simple roots, as in *ka-lik*, but also with derived ones, since *fetsar* is the applicative derivation of a root whose associated vocabulary item *fets* means ‘to follow’. Other examples are *pë-lëman* ‘door’, *lëman* ‘cause x to be covered’ a causative of *lëm* ‘to cover’; *na-rukand* ‘heir’, *dukand* ‘to bequeath’ a causative of *duk* ‘to leave’; etc. Forms like /-ar/, /-an/, and /-and/ are members of a rich paradigm of verbal derivational morphemes with meanings such as causative, applicative, directional (e.g., *pëni* ‘to come out’ from *pën* ‘to go out’), reciprocal (e.g., *telar* ‘to understand each other’ from *te* ‘to hear’), and so forth (see Buis 1990 : 46-49). They are all suffixes in accordance with the functional character of the roots they spell out, and, as mentioned, they are uniquely verb-forming (perhaps members of {v}), as the derivational function for nouns is entirely taken over by noun classes. Given this, a possible syntactic representation of, e.g., *fetsar* ‘to follow for/in favour of x’ is (31) :

(31) [_{vP} $\sqrt{\text{ar}}$ [_{vP} $\sqrt{\text{fets}}$]]

from which we derive (32) :

(32) [_{vP} $\sqrt{\text{fets}}$ + $\sqrt{\text{ar}}$ [_{vP} t]]

Merging (32) with an nP projection headed by n° = Class 5 would then yield *bë-fetsar* ‘friend’. Given the absence of verbal derivation of this type in Romance, the nearest equivalent seems to be nominalizations similar to English *destruction* (Spanish *destrucción*). Following Marantz (1997 – also see Pesetsky 1995 : 69ff.), such forms are syntactically built on a structure like (33) :

(33) [_{nP} √tion [_{vP} v [_{vP} √DESTROY]]]

The inclusion of *v* in the representation captures the fact that these forms, despite being nouns, have the argument structure of the verb they nominalize (see *the enemy's destruction of the city*, *la destrucción de la ciudad por el enemigo* – and recall our analysis of /pè/ infinitives). At the same time it shows how different Manjaku *bë-fetsar* and like forms are, since they inherit nothing of the argument structure of the corresponding verb. For instance, *fetsar* as an applicative verb requires a beneficiary and a direct object in a Double Object Construction (see *Man fetsar ul u-ndali* /I follow+Appl her/him 3-cat/ 'I followed the cat for her/him'); *bë-fetsar* 'friend' requires nothing of the sort. To put it tersely, *bë-fetsar* is a noun, it is not a nominalization. The derivation sketched above cannot therefore be the right one.

We are thus led to assume a rather strong version of our interfacing model : not only is the lexicon connected to syntax, but syntax in its turn may feed the lexicon by returning to it the complex roots that it builds, e.g. √√fets√ar. As a complex root √√fetsar (for short) is voided of the categorial features syntax assigns it and is therefore available for (re)insertion into syntax, this time under an nP projection, hence *bë-fetsar* 'friend' where the verbal character of the main root is entirely erased from memory, so to speak.

That said, I am quite willing to admit that such an explanation is certainly not the last word on the subject, if only because the doing-undoing mechanism it involves has something definitely awkward to it. Perhaps the Manjaku data should be taken as support to the notion that the function syntactic head movement fulfils would be better entrusted to some intra-lexical root combining operation. There would be no contradiction between such a move and the overall view of the lexicon defended here (see fn. 67). On the other hand, nominalizations such as (33), the end results of which keep the memory of their formative process, show that we cannot dispense altogether with a syntactic component or, more accurately, with a component that manipulates and categorizes heads according to syntactic principles such as the LCA and the V vs. N feature contrast (for an early extension of Move- α to the lexicon, see Keyser & Roeper 1984 ; also see Roeper 1993). How this component is to be called and where exactly it is located (inside or outside the lexicon or as an interface of the lexicon with morphology) are relatively secondary matters. I will say no more on the subject, however.

To return to our main topic, it seems therefore that the opposed linearizations (and distinct "visibilities") of Class exponents in Manjaku and in Romance can be completely deduced from their different lexical statuses. Let us now proceed to the Class–Number issue.

6.2. The fusion of Class and Number

The empirical generalization that Class and Number (i.e., plural) morphemes combine to be spelled out as syncretic exponents seems quite solid. Typical examples are Italian *gatta* /√gatt+nII/ vs. *gatte* /√gatt+nII.Pl/ 'she-cat(s)', Modern Hebrew *tmuna* /√tmun+F/ vs. *tmunot* /√tmun+F.pl/ 'picture(s)', and of course Manjaku *ka-to* vs. *i-to* 'house(s)'. To the difference of Class, Number is a functional element unequivocally, at least in those languages where it is obligatorily expressed as a bound morpheme.⁷² A plausible assumption within the present framework, then, is that Number is a functional root meaning 'Plurality' (1 < x).⁷³ Singularity has no grammatical representative in this view : it is the interpretative result of not combining Plurality with a root denoting a countable entity. (If the denoted entity is non-countable, the result is massivity.) Plurality is syntactically expressed as a head (Num^o) c-commanding the noun (see Ritter 1991). The syntactic representation of *gatte* 'she-cats' is thus the following :

(34) $[_{\text{NumP}} \text{Num}^\circ [_{\text{nP}} \text{nII} [_{\sqrt{\text{P}}} \sqrt{\text{gatt}}]]]$

It is easy to see that, given our basic assumptions, Number-Class syncretism follows straightforwardly from such a representation : nII attracts the root as already explained, hence :

(35) $[_{\text{NumP}} \text{Num}^\circ [_{\text{nP}} \sqrt{\text{gatt+nII}} [_{\sqrt{\text{P}}} \text{t}]]]$

(This, note, must indeed be the first step, because, under the assumption that Attract is local, Num[°] could only target nII, but it would “gain” nothing by so doing since nII is also functional.) Then Num[°] attracts the complex (non-functional) head $\sqrt{\text{gatt+nII}}$, hence :

(36) $[_{\text{NumP}} \sqrt{\text{gatt+nII+Num}^\circ} [_{\text{nP}} \text{t} [_{\sqrt{\text{P}}} \text{t}]]]$

Class and Number are thus necessarily adjoined, and in that order. It is precisely this that my analysis of Class and Number predicts. The outcome of the adjunction then depends on the morphophonological form of the exponents. In Italian, we may assume the exponent of Plurality to be /l/, i.e. a high vowel which gives an /e/ when fused with the /a/ exponent of nII.⁷⁴ In Spanish, on the other hand, Plurality spells out as /s/, a consonant that can only be merged (in the sense of added to, not fused) with the Class exponent, hence *gatas* ‘she-cats’.⁷⁵

Can this analysis be extended to Manjaku ? Let us take (34) as also representing what surfaces as *i-to* ‘houses’ :

(37) $[_{\text{NumP}} \text{Num}^\circ [_{\text{nP}} \text{n}^\circ [_{\sqrt{\text{P}}} \sqrt{\text{to}}]]]$

Here, functional attraction is a property of Num[°] only, and it targets the closest non-functional category, viz. n[°]. What enters morphology is therefore :

(38) $[_{\text{NumP}} \text{n}^\circ + \text{Num}^\circ [_{\text{nP}} \text{t} [_{\sqrt{\text{P}}} \sqrt{\text{to}}]]]$

Class-Number adjunction is again predicted, plus the fact that the resulting complex will appear as a prefix to the root. Accounting for the form of this complex, on the other hand, is not so obvious. It depends crucially on the choice we make for the identity of n[°] in (37)-(38). Suppose we consider it to be the syntactic projection of Class 7, that is the noun class lexically associated with the root $\sqrt{\text{to}}$ when Plurality is not activated. Since the exponent of Class 7 is /ka/, and it is not morphophonologically related to /i/ as the exponent of /n[°] = 7 + Num/, we are led to assume generalized suppletion : /ka + Num/ = /i/, /na + Num/ = /ba/, /u + Num/ = /ngë/, etc. In fact, it is the assumption underlying the common notion that *i-to* “is the plural of” *ka-to*. Nor is suppletive pluralization an unheard of phenomenon (see French *œil* vs. *yeux* ‘eye(s)’, Modern Hebrew *iša* vs. *našim* ‘woman/women’, Xârâcùù *kâmûrû* vs. *pââdo* ‘man/men’).⁷⁶

The problem with this account is thus not suppletion *per se*, but rather the usual conundrums of noun class systems, namely that singular–plural pairings are not one-to-one, as in the idealized chart (1), in a significant number of cases – see *pë-bank* / *i-bank* (not **kë-bank*) ‘dyke(s)’, *bë-fetsar* / *i-fetsar* (not **m-fetsar*) ‘friend(s)’ – and that it is even possible for a root to be pluralized in a class that is not normally associated with Plurality – see *na-pats* / *u-pats* ‘child/children’ (compare *u-ndali* ‘cat’). In other words, the form of the “plural” exponent is not predictable from that of the non-plural exponent through simple suppletive equations such as the ones given above, and it also depends on the identity of the

root in the sense that speakers must know that, if they choose to refer to ‘children’, then the proper exponent to prefix to the denoting root is the one that spells out $n = \text{Class } 3$.

Therefore, an explanation involving the syntactic steps (37) and (38), followed by suppletive exponence, although it works in a fair number of cases, does not seem adequate for Manjaku. True, we could repair it by postulating multiple suppletion, i.e. by positing disjunctive sets of suppletive equations such as $\{/p\ddot{e}+\text{Num}/ = /k\ddot{e}/ \vee /p\ddot{e}+\text{Num}/ = /i/\}$ and indexing roots for the member of the disjunction they merge with (e.g. $\sqrt{\text{bank}} \supset /p\ddot{e}+\text{Num}/ = /i/$). This is a quite *ad hoc* solution, however, which is not even descriptively adequate since, as we saw, more than one type of plural must be accommodated for some roots (e.g., *p\ddot{e}-konj* ‘finger’, *k\ddot{e}-konj* *k\ddot{e}-wants* ‘three fingers’, *i-konj* ‘fingers’).

In fact, the difficulty with explaining Manjaku plural formation comes from the necessity to provide for two different aspects at the same time : on one hand, it seems to be the case that the lexical association between a root and a Class root varies independently according to whether the association denotes one or several tokens of the referred entity. In other words, the association $\{\text{Class}_i + \sqrt{x}\}$ with singular denotation is independent from the association $\{\text{Class}_j + \sqrt{x}\}$ with plural denotation. Non-predictability of plural Class membership with respect to the singular (and the other way around) depends on this. On the other hand, it is also true that Class signs (i.e. Class roots and their exponents) are inherently specified for Number, as it is never the case that *ka*- \sqrt{x} , for instance, may refer to several tokens of the entity denoted by \sqrt{x} , or *i*- \sqrt{x} to one token of the same entity.⁷⁷ Because of the former fact, as already mentioned, we cannot simply derive *i-bank* from *p\ddot{e}-bank* as we do *gatas* from *gata*.⁷⁸ The latter fact, for its part, is what prevents us from adopting another, *a priori* plausible solution, namely assuming that roots merge with Plurality first – i.e., reversing the order of nP and NumP in (37). If we did that (and only that), and since singular and plural lexical associations are independent from one another, how could we make sure that pluralized roots do not combine with *any* Class roots, rather than the subset they actually associate with ?⁷⁹

Recall now that we entertained the minimal possibility (minimal in term of changes in the model) that what is attained through syntactic head movement as in (37)-(38) may also be realized through intra-lexical combination. I will then make the following proposal : in Manjaku-type languages, the functional lexical item Plurality is not mapped into syntax, but it merges with Class roots within the lexicon.⁸⁰ To put it more formally : to each member of $\{n\}$, i.e. to each $\{\{n_i\}\}$, there corresponds another singleton set consisting in the intersection of $\{\{n_i\}\}$ with $\{\text{Plurality}\}$, viz. $\{\{n_i\} \cap \{\text{Plurality}\}\}$. Members of both sets have different exponents (e.g., */na/* and */ba/*). Given a particular root not belonging to these sets (e.g., $\sqrt{\text{bank}}$), there must be a statement of the fact that this root can associate with a particular member of $\{n\}$ – i.e., a piece of knowledge that language learners have to acquire and that is inherently part of what it means to know that root. And there must be another, independent statement of the fact that the same root can associate with a particular member of $\{\{n\} \cap \{\text{Plurality}\}\}$. As a consequence, there is no NumP projection in syntax, and the n° head includes both Class and Number features.

Because roots and Class roots have meaning, and associations are not haphazard semantically, the same n_i will be mentioned in both statements in a fair number of cases, so that deductive generalizations can indeed be drawn, such as “If a root denotes a human being, then it associates with $\{n_i = /na/\}$, on the one hand, and with $\{n_i \cap \text{Plurality} = /ba/\}$, on the other hand”.⁸¹ But, and that is the gist of the proposal, those are no more than factual generalizations, as nothing in the system forces the same n_i to be targeted in the two statements. In other words, because the selections are independent, there is no necessary connexion (just a frequent one) between the semantic contents of the n_i ’s that associate with a root in or out of intersection with Plurality – just as no connexion is observed in a

significant number of cases between the basic semantics of the Class signs and the meanings of the roots they attach to.

Actually, the deep implication of the present account is that, in languages like Manjaku, Plurality's primary association is not with denoting roots, as in English or Romance, but with Class roots understood as pronouns (see above). It is Class roots that are pluralized *and* associated with denoting roots which, of themselves, have no specification for Number. And it is the global mutual independence (not precluding significant predictability) of $\{n\}$ and $\{n \cap \text{Plurality}\}$ selection that justifies conceiving of the association as a set-theoretic operation within the lexicon rather than as syntactic merger. In Romance-type languages, roots are unspecified for Number as well, but Plurality does not gravitate towards Class. It is autonomously mapped into syntax as a Num^o head, and Number-Class merger is a morphological operation. In other words, Manjaku *ngë-ndali* 'cats' is not the plural of *u-ndali* in the sense that Spanish *gatas* is the plural of *gata*; rather, $\{\{n\} \cap \{\text{Plurality}\}\} = /ngë/$ is an inherently plural Class sign semantically paired with $\{\{n\}\} = /u/$, that may (because of this pairing) but need not be associated with the same denoting roots $\{\{n\}\} = /u/$ attaches to (see *ngë-nana* as the numbered plural paired with *bë-nana* 'banana tree').

Admittedly, such a conclusion does not wander far from what has long been received wisdom in noun class studies of Bantu and so-called "Bantoid" languages. What I claim to contribute, beyond formalization, is the discovery of a basic relation between this state of affairs and the root character of Class signs in these languages, as opposed to the purely functional nature of gender morphemes in Romance languages. I will refrain from further predictions at this stage, however, because Class systems are complicated and I prefer not to overstep the comparative domain I chose. Let me just underline a few points in connexion with the preceding considerations.

First, the present account is fully compatible with the existence of systems such as that of Standard Wolof, outlined above, where the number of distinguished classes in the plural is drastically reduced relative to the singular. Our interpretation of this fact is simply that, whereas $\{n\}$ consists of eight members, $\{\{n\} \cap \{\text{Plurality}\}\}$ includes only two, one for some humans and one for all the rest, including humans – a dramatic illustration of the possible unconnectedness of *n*'s semantic contents in the two sets. (The limiting case where $\{\{n\} \cap \{\text{Plurality}\}\}$ is further reduced to one member implies that *n* is semantically void in this set, i.e. it has no features except N.⁸²)

Secondly – and this is to be taken as a tentative opening onto further research – it might be that the contrast root vs. functional element, clearly exposed by the comparison of Manjaku with Romance, admits of intermediate cases. One such case would be Modern Hebrew. It is well known that in this language the intersection of gender and plural marking yields a number of idiosyncratic formations: masculine nouns (as shown by agreement) may take the feminine plural ending (e.g., *šulxan* / *šulxanot* 'table(s)') while some feminine nouns take the masculine plural ending (e.g., *dvora* / *dvorim* 'bee(s)', *šibolet* / *šibolim* '(corn) ear').⁸³ Moreover, the true assertion that masculine nouns never bear an overt gender morpheme is not reversible, as a number of feminine nouns are also unmarked (e.g., *kos mala* 'a full cup' / *kosot malot* 'full cups').⁸⁴ Such facts led Ritter (1991) to assume that Hebrew gender, in addition to being an inherent lexical property of nouns, is morphologically derivational (see Ritter 1991: 50ff. for evidence to that effect, based on work by Bat-El 1986), whereas Number is inflectional, i.e. carried by a functional projection NumP that is unspecified for gender. Agreement facts such as *šulxanot ktanim* 'small tables', showing that the adjective agrees in gender with the (masculine) base rather than with the (theoretically feminine) plural ending, bear out the latter claim.

Yet, the fact that *šulxan* has *šulxanot* for a plural instead of expected **/šulxanim/* (and *dvora*, *dvorim* instead of **/dvorot/*) receives no explanation in Ritter's account. On the

other hand, her observation that gender fulfils clearly derivational functions in Hebrew (as in her example of *amud* ‘page’ vs. *amuda* ‘column’ – p. 51) makes Hebrew gender look similar, in this respect at least, to Manjaku noun classes, whose derivational potential was shown to be crucial for their analysis. Hence a suggestion that the complexities of Hebrew gender-number marking might at least be sorted out, if not fully disentangled or clarified, by means of the same analytical tools as we used for Manjaku (and for Romance).

A first necessary step, I think, is to rid ourselves of the traditional, but obscuring categories “masculine” and “feminine” in favour of word classes as in Harris’s analysis of Spanish, keeping them only as an encyclopaedic classificational criterion applicable to a small subset of the total lexicon (see above). Assume then that Hebrew roots are nominalized into 4 word classes which I shall designate as nI (corresponding to old masculine), nII, nIII, and nIV (these latter three corresponding to old feminine). The exponents are as follows : nI = /∅/ (e.g., *xatul* ‘cat’) ; nII = /a/ (e.g., *xatula* ‘she-cat’) ; nIII = /Vt/ (e.g., *igeret* ‘letter’, *balšanut* ‘linguistics’, *mexonit* ‘car’) ; nIV = /∅/ (e.g., *kos* ‘cup’).⁸⁵ These four (or three) elements constitute the membership of the {n} set in Hebrew. I will not speculate about their semantics.

Assume further than in Hebrew as in Manjaku there is a second, disjoint set formed by intersecting {n} with Plurality. (For simplicity, I will use nXPL as a notation, and I abstract from Dual.) One member of this set is nIPL = /im/ (e.g., *xatulim* ‘cats’). Another member is nIPLa = /ot/ (e.g., *šulxanot* ‘tables’). Empirical evidence shows that the roots lexically associated with this class must be listed, since they are in finite number (no new nouns enter this class) and no natural generalization over them, touching their form or meaning, seems feasible (see Cohen & Zafrani 1968 : 172ff.). By a simple application of the Elsewhere Principle, all roots not in the list are associated with class nIPL if they are also associated with class nI. Similarly, we have classes nIIPL = /ot/ (e.g., *xatulot* ‘she-cats’, *igarot* ‘letters’, *mexoniyot* ‘cars’) and nIIPLa = /im/ (e.g., *dvorim* ‘bees’, *šibolim* ‘(corn) ears’) where roots associated with nIIPLa also have to be listed. Note that classes nII, nIII, and nIV can be collapsed when intersected with Plurality, even though it seems to be the case that no noun in class nIV (e.g., *kos* ‘cup’) has its plural in class nIIPLa.⁸⁶ Four plural classes have thus to be distinguished as well.

What we see, then, is that {n} and {nPL} are not independent in Hebrew in the sense they are in Manjaku, since all nI-associated roots associate with nIPL or nIPLa, and all nII/III/IV-associated roots associate with nIIPL or nIIPLa. There is a measure of unpredictability unknown in Romance, however, insofar as given an nI root, for instance, learners of the language cannot associate it with nIPL or nIPLa “before consulting the list”. It is in that sense that Hebrew may be said to have no inflectional classes as argued by Aronoff (1994 : 75ff.). Moreover, it is possible for a root to be associated with {nPL} without being associated with {n} (so-called *pluralia tantum*, e.g. *mayim* ‘water’, *šulayim* ‘margin’, etc.). Because of these complexifying factors, therefore, a Manjaku-like solution, with Class and Plurality lexically combined rather than syntactically merged, seems adequate – all the more so since, as we saw, Hebrew gender has derivational features like Manjaku noun classes, although to a much lesser extent.

Nevertheless, Hebrew gender or gender-number morphemes are suffixes like their Romance equivalents. In our terms, this means that members of {n} and {nPL} in Hebrew are not denoting roots, but functional roots, in correlation with the semantic narrowness of the classificational criterion they invoke, as well as their limited derivational role as compared with Manjaku noun classes. In this way, then, Hebrew does represent an intermediate case between Manjaku and Romance : it is like the former insofar as Class and Plurality combine in the lexicon, thus precluding the syntactic insertion of a Num^o head below DP *contra* Ritter (1991), an option that UG must allow, I submit. It is like Romance, on the other hand, because Class, being only weakly derivational and limited in terms of

classificatory relevance, is functional, with the consequence that the n(PL)^o head that expresses it in syntax must attract the root it nominalizes, therefore appearing as a suffix at the output of morphology.⁸⁷

7. Conclusion (with an outlook on classifiers beyond gender and noun classes)

If all the foregoing makes some kind of sense, there is in the universal lexicon an element I call Class, the unvarying property of which is to convert roots into nouns. Correlated to this grammatical property, Class is endowed with a classificatory (i.e., semantic) content of varying richness or relevance with respect to the universe of concepts each language's particular lexicon expresses. The lexical status of Class items as denoting or functional roots, and the morphosyntactic effects that ensue, depend on this content. This is, I claim, the significant result of the present enquiry. I will now mention a few issues that follow from or are connected with it and have not been dealt with so far.

Given such a universality assumption, what about languages like Turkish, where Class has no overt manifestation whatsoever, or classifier languages such as Chinese? As for the former, the simplest assumption is that they represent the limiting case where Class, i.e. {n}, has no content at all, being purely noun-forming (and functional).⁸⁸ Note the existence of such a case is expected in principle, since Class's content is a gradient by definition, and there are numerous examples showing how a contentful Class system, of the Romance or Bantu/Bantoid type, may evolve into a less contentful or a contentless Class system analogous to Turkish – see, e.g., the cases of Wolof or Turkana mentioned above, and the Romance or Bantu-based creole or pidgin languages (see, e.g., Manessy 1977).

It is not possible, on the other hand, to do justice to numeral or possessive classifier systems, to use Aikhenvald's (2000) characterizations, in such a cursory fashion, and examining them in detail would inflate this article out of all proportion. Let me just suggest a few directions.

As their name indicates, numeral classifiers only occur in conjunction with quantifying expressions, typically numerals. The most widely known such system is that of Chinese, but I will rather give an example from Ejagham, a Benue-Congo, Southern Bantoid language (see Watters 1981 : 310, quoted in Aikhenvald 2000 : 99) :⁸⁹

(39) à-múg ʼ í- ʒkúɗ á-bá'
 6-CL:SMALL.ROUND Gen6 19-orange.seed 6-two
 two orange seeds

Ejagham, by no means exceptional in this respect, is especially interesting as it shows that a language may have recourse to more than one Class system, here noun classes concurrently with numeral classifiers. Moreover, it dramatically demonstrates the non-functional, root nature of numeral classifiers, since the classifier *à-múg* is marked for noun class (6, a plural class) like any noun, and the numeral meaning 'two' agrees with it. The discontinuous noun phrase [à-múg ... á-bá'] then forms a genitive construction (whose marker is the floating low tone glossed Gen, also associated with class 6) with the noun *í- ʒkúɗ* which belongs to a different noun class than the classifier, viz. 19, a *singular* class.⁹⁰ It seems we have every reason, therefore, to consider /múg/ a pronoun with the meaning "small round object" morphologically realized as a free form – and likewise Vietnamese *hòn* 'rock', *búc* 'flat, rectangular object', and so forth (see Thompson 1987). What will be the difference, then, between such a free pronoun and a full noun such as *í- ʒkúɗ* 'orange seed', on one hand, and a bound pronoun as noun class markers are, on the other hand?

With respect to the latter difference, it clearly lies in the fact that bound pronouns such as Manjaku /na/, /ka/, etc. *are* instantiations of {n}, whereas free pronouns must combine with (members of) {n}, visibly as in Ejagham or invisibly as in Vietnamese. Ejagham *n-múg* (the class 1 singular correspondent of *à-múg*) or Vietnamese *hòn* are thus independent lexemes, which do not form one projection with the noun they bear on, to the difference of Manjaku /na/, /ka/, etc. which, despite being denoting roots, must merge with another root because of the noun forming property of {n} that they spell out. This is enough to preclude agreement from NP in languages with numeral classifiers of such a sort. In the best of worlds, we should be able to correlate this morphological difference with the varying semantic richness of the items, since Ejagham *n-múg* or Vietnamese *hòn* clearly have more specific meanings than Manjaku /na/ or /ka/. We are not allowed such a move, though, because of the existence of languages like the Papuan language Nasioi (see Foley 1986), the lexicon of which contains more than 100 Class elements as specific semantically as the numeral classifiers of Ejagham or Vietnamese, but with agreement within the DP as the exponent of the chosen element, an enclitic attached to the head noun, is copied onto each component (also as an enclitic).⁹¹

It does not mean there is no correlation between semantic content, i.e. classificatory potential, and morphological status, but it is a complex correlation. What we have, indeed, is a classificatory potential continuum that goes from poor (genders) to general (noun classes of the Manjaku type) to specific (numeral classifiers or noun classes of the Nasioi type), and which projects onto a three-way alternative for exponence : suffix, clitic, or free form, corresponding to the three options of fusion, merger or no operation, in a hierarchy of increasing morphological freeness. From this, it seems we can draw the following observational generalizations : poor classificatory potential is never expressed as free forms, although it may be realized as clitics as in Turkana ; that is to say, it is always a value of {n}, whose contentfulness varies (and may become impoverished). Specific classification, at the other end of the continuum, can be realized as free forms or clitics, but never, it seems, as suffixes, meaning that the rich content of {n} then required prevents it from being fully functional. Finally, there do not seem to be cases where general classification of the Manjaku type, in the middle of the continuum, is realized as either suffixes or free forms. All this seems to be consistent with our basic assumptions, although, I repeat, it is no more than generalizations still far from real explanations (see Aikhenvald 2000 for more facts and for the grammaticalization issues involved).

As for the difference between numeral classifiers and “ordinary” nouns, there does not seem to be anything intrinsic in it, if only because in many languages numeral classifiers are open lexical classes, and one and the same item may be used now as a numeral classifier, now as an ordinary noun (see English *head*, with the intriguing property that it does not pluralize when used as a classifier).⁹² The crucial property would rather be external to Class proper, and it would have to do with the fact that all nouns in numeral classifier languages are basically Mass nouns that need the classifier to be quantified (see Chierchia 1998). This would be the case of Vietnamese, for instance. In Ejagham, it is noteworthy that the quantified noun in examples like (39) is singular, probably interpreted as generic.⁹³

To conclude, I would like to say a few words about agreement which, as a matter of fact, has been almost entirely left out in this study. This apparent oversight reflects to some extent my initial reaction against treatments which consider agreement the be-all and end-all of Class phenomena. But it is also due to the fact that agreement is not really problematic for my account and largely tangential to it. Take noun-adjective agreement, the only form Romance and Manjaku have in common. For Romance, we have Cinque’s (1994) theory according to which adjectives are generated preminally in the specifiers of dedicated functional projections, and N° raises over them through the heads of these functional projections, thereby creating the necessary Spec-Head configuration for agreement.

According to the present analysis, N° is not basically present in the syntax, but there is a maximal projection $[_{nP} n^\circ [_{\sqrt{P}} \sqrt{^\circ}]]$. Once n° has attracted the root head, however, a complex head $[\sqrt{^\circ}+n^\circ]$ is created, which is no more (and no less) than the decomposition of N° . The derivation can then proceed unaltered.

Cinque's account can probably be extended to Manjaku where attributive adjectives also follow the head superficially (see *u-ndali u-faacal* /3-cat 3-white/ 'white cat').⁹⁴ A problem, though, is that root heads do not adjoin to n° in Manjaku, so that syntax has to deal with a maximal nP projection until Spell Out (i.e., delivery to morphology). This certainly stands in the way of N° passing through intermediate functional heads. A possible solution is to adopt (and adapt) Cinque's suggestion (p. 106) for prenominal adjectives in Romance, i.e. to assume that nP raises overtly to a higher position whereas the root head raises at LF (perhaps to D if it is projected in Manjaku). Spec-Head agreement would ensue, since roots are lexically associated with noun classes. Number would still be a problem, however, under our assumption that Plurality and Class combine lexically in $\{n\}$ and there is no NumP projection, with the consequence that, if the root head moves by itself, LF may "know" that it belongs to noun class 3, but it may not know whether it is pluralized or not. Moreover, justifying nP raising is no easy task.

Another, much simpler solution is available, however. Consider that so-called adjectives are not different from nouns in Manjaku except for the fact that their class membership is not lexically specified. It means that /x-faacal/ 'white' (x = any noun class) has the same basic structure as *u-ndali* 'cat', viz. $[_{nP} n^\circ [_{\sqrt{P}} \sqrt{^\circ}]]$, but for the fact that n° is unspecified (beyond the N feature). In the spirit of Cinque's analysis, the basic structure of what surfaces as *u-ndali u-faacal* 'white cat' is therefore the following, with dots indicating whatever structure may be present between the two nP (perhaps none) :

(40) $[_{nP} n^\circ_x [_{\sqrt{P}} \sqrt{^\circ} = \text{faacal} \dots [_{nP} n^\circ_3 [_{\sqrt{P}} \sqrt{^\circ} = \text{ndali}]]]]]$

In fact, such a configuration is strikingly similar to that of Semitic genitive phrases (so-called Construct States) headed by an adjective (see Standard Arabic *Hasanu l-wajhi* / handsome the-face/ 'handsome of face'). Given the form of Manjaku genitive phrases (see *u-ndali na-cën* /3-cat 1-neighbour/ 'the cat of the neighbour') and the noun-adjective indistinction just mentioned, such a parallel appears quite sound. The main point of (40), however, is that it provides us with a reason why the lower nP must raise, namely that the higher nP's head n°_x is unspecified and must get a class value before it may pass the morphological threshold. Therefore, the lower nP raises to the specifier of the higher one, thus creating an agreement configuration such that n°_x gets the class value of n°_3 .⁹⁵

That said, there is a still simpler solution, and that is to assume that nothing particular happens in the syntax, and that Agree (in the sense of Chomsky 1999) is an operation of the interface, either between the lexicon and syntax, or even between the lexicon and morphology (if word order is taken care of by the latter, and syntax is pure LF). That would be a really minimal account. My proposal that Class and Plurality combine lexicon-internally in Manjaku certainly points towards that direction. On the other hand, as also mentioned, *destruction*-type nominalizations seem to indicate that not all morphology-relevant operations can be done "in the privacy of your own lexicon", to use Marantz's expression (but see Roeper 1993). As such speculations open some sort of a Pandora's box, the contents of which far exceed the aims of this work, I leave them at that for the moment.

References

Abney, Steven. 1987. "The English Noun Phrase in its Sentential Aspects." Ph.D.diss., MIT.

- Aikhenvald, Alexandra Y. 2000. *Classifiers : A Typology of Noun Categorization Devices*. Oxford : Oxford University Press.
- Allan, K. 1977. "Classifiers", *Language* 53 : 284-310.
- Arad, Maya. 1999. "On 'little v'". In Karlos Arregi et al., eds., *Papers on Morphology and Syntax, Cycle One. MIT Working Papers in Linguistics* 33, pp. 1-25.
- Aronoff, Mark. 1992. "Noun Classes in Arapesh." In Geert Booij & Jaap van Marle, eds., *Yearbook of Morphology 1991*. Kluwer Academic Publishers, Dordrecht, pp. 21-32.
- . 1994. *Morphology by Itself : Stems and Inflectional Classes*. MIT Press, Cambridge, Mass.
- Baker, Mark C. 1992. "Morphological Classes and Grammatical Organization." In Geert Booij & Jaap van Marle, eds., *Yearbook of Morphology 1991*. Kluwer Academic Publishers, Dordrecht, pp. 89-106.
- Bat-El, Outi. 1986. "Extraction in Modern Hebrew Morphology." M.A.diss., UCLA.
- Benmamoun, Abbas. 2000. "Lexical Relations and the Role of the Imperfective Template in Arabic Morphology." Unpublished ms., University of Illinois, Urbana-Champaign.
- Borer, Hagit. 2001. "One is the Loneliest Number". Paper presented to the Conference on the Syntax and Semantics of Semitic Languages, University of Southern California, Los Angeles, May 4-5, 2001.
- Brody, Michael. 1995. *Lexico-Logical Form : A Radically Minimalist Theory*. MIT Press, Cambridge, Mass.
- Bouchard, Denis. 1995. *The Semantics of Syntax : A Minimalist Approach to Grammar*. The University of Chicago Press, Chicago.
- Buis, Pierre. 1990. *Essai sur la langue manjako de la zone de Bassarel*. Instituto nacional de Estudos e Pesquisas, Bissau.
- Bybee, Joan L. 1985. *Morphology : A Study of the Relation Between Meaning and Form*. John Benjamins, Amsterdam.
- Chametzky, Robert A. 2000. *Phrase Structure : From GB to Minimalism*. Blackwell, Oxford.
- Chierchia, Gennaro. 1998. "Plurality of Mass Nouns and the Notion of 'Semantic Parameter'." In Susan Rothstein, ed., *Events and grammar*. Kluwer, Dordrecht, pp. 53-103
- Chomsky, Noam. 1995. *The Minimalist Program*. MIT Press, Cambridge, Mass.
- . 1998. "Minimalist Inquiries : The Framework." *MIT Occasional Papers in Linguistics* 15.
- . 1999. "Derivation by Phase." Unpublished ms., MIT.
- . 2001. "Beyond Explanatory Adequacy." Unpublished ms., MIT.
- Cinque, Guglielmo. 1994. "On the evidence for partial N-movement in the Romance DP." In Guglielmo Cinque et al., eds., *Paths Towards Universal Grammar : Studies in Honor of Richard S. Kayne*. Georgetown University Press, Washington, D.C., pp. 85-110.
- . 2000. "On Greenberg's Universal 20 and the Semitic DP", Unpublished ms., University of Venice.
- Cohen, David & Haïm Zafrani. 1968. *Grammaire de l'hébreu vivant*. Presses universitaires de France, Paris.
- Corbett, Greville. 1996. *Gender*. Cambridge University Press, Cambridge.
- Craig, Colette, ed. 1986. *Noun Classes and Categorization*. John Benjamins, Amsterdam.
- Croft, W. 1994. "Semantic Universals in Classifier Systems", *Word* 45 : 145-171.
- Denny, J. Peter & Chet A. Creider. 1986. "The Semantics of Noun Classes in Proto Bantu." In Colette Craig, ed., *Noun Classes and Categorization*. John Benjamins, Amsterdam, pp. 217-239.
- Dimmendaal, Geert I. 1983. *The Turkana Language*. Foris, Dordrecht.
- Dixon, R.M.W. 1972. *The Dyirbal Language of North Queensland*. Cambridge University Press, Cambridge.

- . 1986. “Noun Classes and Noun Classification in Typological Perspective.” In Colette Craig, ed., *Noun Classes and Categorization*. John Benjamins, Amsterdam, pp. 105-112.
- Doneux, Jean-Léonce. 1967. “Le manjaku, classes nominales et questions sur l’alternance consonantique.” In Gabriel Manessy, ed., *La classification nominale dans les langues négro-africaines*. Editions du CNRS, Paris, pp. 261-276.
- . 1975. *Lexique manjaku*. Centre de Linguistique appliquée de Dakar, Dakar.
- Doron, Edit. 2000. “Transitivity Alternations in the Semitic Template System.” Unpublished ms., The Hebrew University of Jerusalem.
- Dressler, Wolfgang U. 1989. “Prototypical Differences between Inflection and Derivation.” *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 42 : 3-10.
- Embick, David & Rolf Noyer. 1999. “Locality in Post-Syntactic Operations.” In Vivian Lin et al., eds., *Papers on Morphology and Syntax, Cycle Two. MIT Working Papers in Linguistics* 34, pp. 265-317.
- Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge University Press, Cambridge.
- Grandi, Nicola. 2002. *Morfologie in contatto : Le costruzioni valutative nelle lingue del Mediterraneo*. Pavia : Francoangeli.
- Hale, Kenneth & Samuel J. Keyser. 1993. “On Argument Structure and the Lexical Expression of Syntactic Relations.” In Kenneth Hale & Samuel J. Keyser, eds., *The View from Building 20 : Essays in Linguistics in Honor of Sylvain Bromberger*. MIT Press, Cambridge, Mass., pp. 53-109.
- Halefom, Girma & John S. Lumsden. 2000. “Decomposing the Perfective Conjugation.” Unpublished ms., UQAM.
- Halle, Morris & Alec Marantz. 1993. “Distributed Morphology and the Pieces of Inflection.” In Kenneth Hale & Samuel J. Keyser, eds., *The View from Building 20 : Essays in Linguistics in Honor of Sylvain Bromberger*. MIT Press, Cambridge, Mass., pp. 111-176.
- Harris, James. 1992. “The Form Classes of Spanish Substantives.” In Geert Booij & Jaap van Marle, eds., *Yearbook of Morphology 1991*. Kluwer Academic Publishers, Dordrecht, pp. 65-88.
- . 1999. “Nasal Depalatalization *no*, Morphological Wellformedness *sí* ; The Structure of Spanish Word Classes.” In Karlos Arregi et al., eds., *Papers on Morphology and Syntax, Cycle One. MIT Working Papers in Linguistics* 33, pp. 47-82.
- Ippolito, Michela M. 1999. “On the Past Participle Morphology in Italian.” In Karlos Arregi et al., eds., *Papers on Morphology and Syntax, Cycle One. MIT Working Papers in Linguistics* 33, pp. 111-137.
- Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. MIT Press, Cambridge, Mass.
- Katamba, Francis. 1993. *Morphology*. London : Macmillan.
- Kaye, Jonathan, Jean Lowenstamm & Jean-Roger Vergnaud. 1985. “The Internal Structure of Phonological Elements : A Theory of Charm and Government.” *Phonology Yearbook* 2 : 305-328.
- Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. MIT Press, Cambridge, Mass.
- Keyser, Samuel J. & Thomas Roeper. 1984. “On the Middle and Ergative Constructions in English.” *Linguistic Inquiry* 15 : 381-416.
- Kihm, Alain. 1998. “Propositions relatives en manjaku et théorie du déplacement comme copie.” *Linguistique africaine* 20, 75-99.
- . 1999. “Wolof NP and the Fusion vs. Merger Contrast.” *Actes du Colloque « Langues et Grammaires », II Syntaxe*. Presses Universitaires de Vincennes, Saint-Denis.
- . & Aristide Gomes. 1988. “Quelques points de syntaxe du manjaku.” *Langues et grammaire* 1, 21-56.

- Lakoff, George. 1986. "Classifiers as a Reflection of Mind." In Colette Craig, ed., *Noun Classes and Categorization*. John Benjamins, Amsterdam, pp. 13-51.
- . 1987. *Women, Fire, and Dangerous Things : What Categories Reveal About the Mind*. The University of Chicago Press, Chicago.
- Lecarme, Jacqueline. 1999. "Gender 'Polarity' : Theoretical Aspects of Somali Nominal Morphology". Unpublished ms., CNRS (URA 1028).
- Lowenstamm, Jean. 1996. "CV as the Only Syllable Type." *Current Trends in Phonology* : 419-441.
- Manessy, Gabriel. 1977. "Processes of Pidginization in African Languages." In Albert Valdman, ed., *Pidgin and Creole Linguistics*. Indiana University Press, Bloomington, pp. 129-154
- Marantz, Alec. 1997. "No Escape from Syntax : Don't Try Morphological Analysis in the Privacy of Your Own Lexicon." In A. Dimitriadis et al., eds., *Proceedings of the 21st Penn Linguistics Colloquium. UPenn Working Papers in Linguistics*, Philadelphia, pp. 201-225.
- Matthews, P.H. 1974. *Morphology : An Introduction to the Theory of Word-Structure*. Cambridge University Press, Cambridge.
- Miller, G.A. & P.N. Johnson-Laird. 1976. *Language and Perception*. Cambridge : Cambridge University Press.
- Moyse-Faurie , Claire. 1995. *Le xârâcùù, langue de Thio-Canala (Nouvelle-Calédonie) : éléments de syntaxe*. Editions Peeters-SELAF, Paris.
- Mufwene, Salikoko S. 1980. "Bantu Class Prefixes : Inflectional or Derivational ?" In Jody Kreiman & Almerindo E. Ojeda, eds., *Papers from the Sixteenth Regional Meeting of the Chicago Linguistic Society*, Chicago Linguistic Society, Chicago, pp. 246-258
- Noyer, Rolf. 1997. *Features, Positions, and Affixes in Autonomous Morphological Structure*. Garland, New York.
- Pasch, Helma. 1985. "Possession and possessive classifiers in 'Dongo-ko', *Afrika und Übersee* 68, 69-85.
- Pesetsky, David. 1995. *Zero Syntax : Experiencers and Cascades*. MIT Press, Cambridge, Mass.
- Posner, Michael. 1986. "Empirical Studies of Prototypes." In Colette Craig, ed., *Noun Classes and Categorization*. John Benjamins, Amsterdam, pp. 53-61.
- Pozdnyakov, Konstantin I. 1993. *Sravnitel'naja grammatika atlanticeskix jazykov : imennye klassy i fono-morfologija [A Comparative Grammar of the Atlantic Languages : Noun Classes and Morphophonology]*. Moskva : Nauka.
- Renault, Richard. 1987. "Genre grammatical et typologie linguistique." *Bulletin de la Société de Linguistique de Paris* 82 : 69-117.
- Ritter, Elizabeth. 1991. "Two Functional Categories in Noun Phrases : Evidence from Modern Hebrew." In Susan D. Rothstein, ed., *Perspectives on Phrase Structure : Heads and Licensing*. Vol. 25 of *Syntax and Semantics*. Academic Press, New York, pp. 37-62.
- Roeper, Thomas. 1993. "Explicit Syntax in the Lexicon : The Representation of Nominalizations." In James Pustejovsky, ed., *Semantics and the Lexicon*. Kluwer, Dordrecht, pp. 185-220.
- Sag, Ivan & Thomas Wasow. 1999. *Syntactic Theory : A Formal Introduction*. CSLI Publications, Stanford.
- Sauvageot, Serge. 1965. *Description synchronique d'un dialecte wolof : le parler du Dyolof*. Institut fondamental d'Afrique noire, Dakar.
- Scalise, Sergio. 1988. "Inflection and Derivation." *Linguistics* 26 : 561-581.
- Serzisko, F. 1982. "Gender, Noun Class and Numeral Classification : A Scale of Classificatory Techniques". In R. Dirven & G. Radden, eds, *Issues in the Theory of Universal Grammar*. Tübingen, Gunther Narr, pp. 95-123.

- Shlonsky, Ur. 2000. "The Form of Semitic Noun Phrases. An Antisymmetric, non N-Movement Approach." Unpublished ms., University of Geneva.
- Sylla, Yèro. 1982. *Grammaire moderne du pulaar*. Les Nouvelles Editions africaines, Dakar.
- Thompson, Laurence C. 1987. *A Vietnamese Reference Grammar*. University of Hawaii Press.
- Trifkovic, Mirjana. 1969. *Le mancagne : étude phonologique et morphologique*. Institut fondamental d'Afrique noire, Dakar.
- Watters, J. 1981. A Phonology and Morphology of Ejagham – With Notes on Dialect Variation. PhD Thesis, University of California, Los Angeles.
- Williams, Edwin. 1981. "On the Notions 'Lexically Related' and 'Head of a Word'." *Linguistic Inquiry* 12 : 234-274.

¹ To the well-known exception of noble vehicles like ships, cars, and locomotives, and country names, all optionally feminine. As a matter of fact, neuter is always an option for nouns denoting non-humans (including babies).

² Even there irrelevancies abound, such as French *sentinelle* ‘sentry’, feminine despite the fact that, up to a recent period, sentries were obligatorily males.

³ Classical studies for such systems are Dixon (1972) and Lakoff (1987). See Miller & Johnson-Laird (1976) for arguments to the effect that shape is the most basic and stable descriptive property (also see Grandi 2002).

⁴ Here I am talking about subject-predicate and DP-internal agreement. Pronoun concord (i.e. the fact that *girl* is coreferred to by *she*, but *syntax* by *it*) is quite probably a different matter, which is why English is such a tricky object for reflecting on Class. Fact is that Class in English is an overt grammatical category for pronouns only, and the relation of *she* to *girl* (or to *ship*) – that is, of the form *she* rather than *he* or *it* to that particular word – is not a matter for syntax (*I know him* proffered when referring to a girl is not ungrammatical, but inappropriate... or mean). Since English pronouns are DPs to themselves and cannot be modified (see **nice she*, **nice who*), and subject-predicate agreement is much less criterial than DP-internal agreement, the English state of affairs does not contradict the claim just made; it merely voids it of its relevance in the particular case. Numeral classifier systems of the Chinese type (see Aikhenvald 2000, Chapter 4), which do not involve agreement, are no counterevidence, as we shall see.

⁵ Even more so if agreement is viewed as a relation (“Agree”) motivated by the need to remove uninterpretable features from syntax (see Chomsky 1999, 2001).

⁶ Aikhenvald collapses gender and noun class systems, contrasting them with numeral classifiers, classifiers in possessive constructions and so-called verbal classifiers (see below). Given her typological-cognitive perspective, she is quite right to do so. For my own purposes, however, gender and noun class must be kept distinct at the grammatical level.

⁷ By “Class exponents” I mean *overt* Class morphemes in whatever system.

⁸ I am giving “fusion” and “merger” the sense they have in Distributed Morphology as presented in Halle & Marantz 1993.

⁹ A clear example of a language with enclitic noun class morphemes is Fulfulde (see Sylla 1982). Wolof, to which I will return, where noun class morphemes also seem to be enclitics is actually a more complex case. Turkana, an East-Nilotic language with three genders, masculine, feminine and neuter, marked with prefixes would seem to ruin the generalization (see Dimmendaal 1983). Yet, considering the way it functions, the Turkana gender system looks more like a very reduced noun class system than a gender system of the Indo-European or Semitic type. Moreover, we know that gender and noun class are not separate categories as shown by Aikhenvald (2000), so the generalizations I am trying to make can only be fully valid for the ends of the continuum, although perhaps not for midway cases like Turkana.

¹⁰ At least, it does look true in the language families of which I have some first-hand knowledge, viz. Indo-European, Niger-Congo, and Afroasiatic, and I am not aware of any evidence from other families that would contradict it. Needless to say, I am expecting such evidence any minute (but see preceding footnote). I also want to make it clear the generalization is intended only for Class exponents – that is to say, I am not implying that noun class languages, for instance, are globally agglutinative (if that makes sense), but only that noun class morphemes are always attached to roots agglutinatively rather than inflectionally, and the reverse for gender morphemes, which, if true, is unexpected and therefore interesting.

¹¹ Manjaku, a Niger-Congo language of the Atlantic subfamily spoken in Guinea-Bissau and South Senegal, will be presented in more detail below.

¹² For instance, the obvious apparent counterexample of Iberian Romance *casas* ‘houses’, where Class and Number seem indeed to be expressed through separate morphemes (see Harris 1992).

¹³ Highly relevant here is Bybee’s (1985) observation that Class agreement on verbs is far less frequent than Number and Person agreement.

¹⁴ *Bak* after the protoform of the class 2 prefix in all these languages.

¹⁵ For clarity’s sake, I separate the noun class prefix from the root with a hyphen. Note that a Manjaku dictionary (e.g., Doneux 1975) is best organized according to the alphabetical order of roots. As for transcription, /ə/ is a schwa, that can be deleted under certain conditions; /ä/ is a centralized /a/; /c/ and /j/ are palatal obstruents. For simplicity, I ignore the contrast of [+ATR] and [–ATR] (or tense and lax) vowels.

¹⁶ Whether actually greater or no is something I will not try to assess, if only because no complete dictionary of Manjaku is available. For an overview of noun class systems in the Atlantic languages, see Pozdnyakov (1993).

¹⁷ The latter item is probably a borrowing of Portuguese *espingarda* ‘gun’, the [p] being reanalysed as the noun class 9 prefix /p(ə)-/. This shows the productivity of the pattern.

¹⁸ A few nouns belong to more than one noun class without change in reference (e.g., besides *pë-ndog* (9) / *i-ndog* (8) ‘stick(s)’, there is *u-ndog* (3) / *ngë-ndog* (4) with the same meaning). This seems rather exceptional, however.

¹⁹ *Te* ‘until’ is a borrowing of Portuguese Creole *te*, itself from Portuguese *até*.

²⁰ [d] and [r] are allophones in Manjaku, [d] being found word-initially and after a nasal, [r] elsewhere. Interestingly, the fact that [r] shows up in *di ruäts* (underlyingly /Di D-uäts/) demonstrates that the phrase is treated as a phonological word (a morphosyntactic word or M-word in Embick & Noyer’s 1999 terminology).

²¹ *Kotiës* actually is a fused NP consisting in *ko* ‘entity’, an item to which we shall return, and *tiës* ‘small’. Here again, it must be noted that what English expresses through adverbials is often expressed by other means in Manjaku, in particular modal verbs such as *kak* ‘do again’ (as in, e.g., *Jon a kak pë-ro w-ul* /Jon Pro do-again 9-do 3-3SG/ ‘Jon did it again’ – see below for a glossed Pro, *pë-ro*, and *w-ul*).

²² Failure to agree is of course not sufficient to assign adjective predicates to a different category than attributive adjectives (see, e.g., German). More convincing is the fact that adjective predicates do not include a copula, although Manjaku is equipped with such an element, that is obligatorily used when the predicate is a noun: compare *A* *(ci) mak* ‘She/he/it is big’ with *A* *(ci) nasiën* ‘He is the/a chief’. Actually, as noted in Buis (1990: 23), agreeing adjectives may be used as predicates as a ‘less common’ alternative to using quality verbs (e.g., *A ci na-war* /Pro Cop 1-good/ ‘S/he is good’ vs. *A wara* ‘S/he (is) good’). The fact that the copula *ci* is then obligatory suggests that *A ci na-war* ought rather to be translated as ‘S/he is a good person’, which, if true, would come as support for the analysis of noun class prefixes I will propose below. As shown by this example, quality verbs are sometimes morphologically distinct from the corresponding adjective (but the status of final /a/ in *wara* is unclear). In addition, attributive adjectives may be derived from quality verbs through a suffix /-al/ (see *U-ndali a faac* ‘The cat is white’ vs. *u-ndali u-faacal* ‘the/a white cat’). Also shown is the absence of overt “simple” determiners in Manjaku, such that, out of context, (2) and (2a) can be interpreted as definite, indefinite, or generic.

²³ *Kà-nän* ‘five’ belongs to a so-called “derived” noun class where the prefix is followed by a low tone /à/ morpheme (perhaps the same as shows up in 1a). Morphophonological processes then apply so that /ka+à/, /ngë+à/, /bë+à/, /m+à/, /pë+à/, and /kë+à/ are realized /kà/, /ngà/, /bà/, /màn/, /pà/, and /kà/ respectively (see Doneux 1975 for the analysis; also Pozdnjakov 1993). Only noun classes 3, 5, 6, 7, 9, and 10 seem to be accompanied with derived classes. Their origin, distribution and function are unclear. (Note that *kà-nän* (7) / *i-nän* (8) means ‘hand’.) The numerals for ‘six’ and ‘eight’ are always pronounced /paaj/ and /kuäs/. Since the supposed roots /aaj/ and /uäs/ are in fact unidentifiable, noun class membership is only hypothetical (whereas *u-ntaja* has a plural *ngë-ntaja* meaning ‘tens’). There are no simple numerals for ‘seven’ and ‘nine’, which are expressed respectively as ‘six and one’ and ‘eight and one’ (see *ngë-pi paaj ni u-lole* /4-goat six and 3-one/ ‘seven goats’, *ba-nän kuäs ni na-lole* /2-man eight and 1-one/ ‘nine men’).

²⁴ With noun class 2 there is a special form *bik-i* and *buk-un* (*ba-kiëj bik-i* ‘these thieves’, *ba-kiëj buk-un* ‘those thieves’) where /bi/uk-/ is probably the same morpheme as appears in the pronoun *bukul* ‘they’ (see below). There is vowel harmony in Manjaku.

²⁵ *Na-ko*’s irregular plural is *buk-ul* ‘they’ (humans only), consisting in the /bi/uk/ allomorph of the noun class 2 prefix /ba-/ (see fn. 24) and the 3rd person singular clitic pronoun /-ul/ (for which, see below).

²⁶ *Buk-ul* is thus ambiguous, serving both as the plural of *na-ko* and of *n-ul*. Noun class 3 *w-ul* is actually realized as [wël]. Noun class 13 *d-ul* is a locative pronoun similar to French *y* (see *Man ya d-ul* ‘I went there / J’y suis allé’). Noun class 12 does not combine with /ul/, but it does with the determiners /i/ and /un/ to build the locative adverbials *ts-i* ‘here’ and *ts-un* ‘there’.

²⁷ In object function, *n-ul* generally (always?) takes on the form *ul*, which I propose to explain by assuming it is then assigned to noun class 1a, whose exponent /a-/ would normally coalesce with the initial vowel. In (9), *na-ko* may of course be replaced by *a* or any 3rd person DP denoting humans without modifying the intended reference.

²⁸ Wolof, the main language of Senegal spoken by several millions, is also classified as Atlantic, but the structural differences with Manjaku are rather enormous. On Wolof in general, see Sauvageot (1965), and Kihm (1999) for a particular study of the noun phrase.

²⁹ I say “usually” because the criterion is not so much anatomy as the fact that the number be determined, explicitly or not. Doneux’s (1975) term for this is *pluriel dénombré* ‘numbered plural’.

³⁰ Compare *Na-kiëj a fäm pë-lëman* ‘The thief broke the door’.

³¹ Whether the Case is Accusative or Genitive may be in debate, since Manjaku Genitive phrases are overtly similar to verb–object constructions (see *pë-lëman na-kiëj* ‘the/a door (of) the/a thief’). The issue may be decided by looking at the Mankanya dialect where, next to the “direct” Genitive construction of Manjaku, there is a construction implying an intermediate determiner agreeing in noun class with the head noun, as in *ka-bats k-i na-pots* /7-ear 7-Det 1-child/ ‘the child’s ear’ (Trifkovic 1969: 88 – I modify her transcription somewhat). This construction, however, seems never to be used when the head is a noun class 9 “infinitive”, as in *nga bi tsiñan pë-jaar u-lugar* /4 Past decide 9-cultivate 3-field/ ‘they (the animals) had decided to cultivate a field’ (p. 121) – not **pë-jaar p-i u-lugar*.

³² Pro does not appear before negative predicates (see Kihm & Gomes 1988).

³³ Unfortunately, I lack data on whether equivalents of “his breaking the door” are constructible in Manjaku. I guess not, as “His breaking my door was not nice” would most certainly be expressed as “He broke my door (and) that was not nice”. More research is clearly necessary, though.

³⁴ The /-an/ suffix is a causative morpheme (see *col* ‘to stand’ vs. *colan* ‘to put in an upright position’ – Buis 1990: 46).

³⁵ Trifkovic translates *J’avais rencontré trois cultivateurs (qui étaient) en train de défricher*. Note, however, that ... *trois cultivateurs en plein défrichage* conveys exactly the same idea of the situation.

³⁶ Also worthy of note are the “adverbial” usage of reduplicating /kë/-prefixed verbal roots (e.g., *man par kë-par* /I pass 10-pass/ ‘I’m just passing by’ – Buis 1990: 46) and expressions like *u-bandi inji* /3-arrive me/ ‘(on) my arrival, when I

arrive(d)'.

³⁷ The indetermination is dramatically exposed by the correspondence of *u-kas* / *ngë-kas* 'male animal(s)' with *kas* 'to castrate'!

³⁸ Recall that *Ka jooti* means 'It crumbled', with a noun class 7 unmentioned subject.

³⁹ Bok equivalent: *Man cuman k-ul*.

⁴⁰ Possible counterexamples are verbs like French *mettre bas* 'to whelp', normally said only of animals, whereas *accoucher* 'to give birth' is only for women. However, it is always possible to be gross and apply the former to women, and few pet lovers have qualms saying 'Ma chatte a accouché'. Such things are therefore a matter of social convention, not of inherent categorization. As for the so-called "verbal classifiers" studied at length by Aikhenvald (2000, Chapter 6), they are actually verb-internal agreement morphemes cross-referencing the Class of the verb's arguments. Verbal classifiers is thus something of a misnomer, since they do not classify the verb itself.

⁴¹ In a strictly lexicalist model, even *pë-lik* and *i-lik* should count as two separate lexical items.

⁴² I assume the intransitive (unergative) use of *lik* to result from incorporation of the cognate object (see Hale & Keyser 1993).

⁴³ I use English words for convenience.

⁴⁴ Note that such a view is not a necessary tenet of DM. Marantz (1997), for instance, assumes otherwise. It seems to me, however, to be both empirically justified and a natural extension of DM's basic hypotheses, in particular the notion that all so-called "lexical" items are actually constructs.

⁴⁵ For a distinct, but compatible, view of the encyclopaedia, see Borer (2001).

⁴⁶ Having exponents thus sets *n* apart from *v* in Manjaku. On the other hand, there is in the language a rich array of derivational verbal morphemes with meanings such as causative, reciprocal, applicative, illative, etc. which may be viewed as being exponents of *v* plus special meanings. The fact that these (suffixal) morphemes only attach to stems that are already verbs by themselves make the identification doubtful, however. I must leave this issue unsettled here. Note that in the Romance languages the theme vowels of verbs could be considered exponents of *v* (see Aronoff 1994: 45).

⁴⁷ I abstract away the issue of Number for the moment.

⁴⁸ Embick and Noyer's definition of functional categories continues Abney's (1987) crucial notion that functional categories are devoid of "descriptive content" (also see Chametzky 2000: 17ff.). As we shall see, however, such a two-way contrast may not be sufficient.

⁴⁹ As indicated in the text, (23) does not constitute a "package", as traditional lexical matrices do, but it represents the interfacing of the several analytical dimensions that contribute to defining the sign. The fact that the syntactic specification is the same for all members of {*n*} thus raises no economy of representation issue.

⁵⁰ What I call "entailment" is what Aronoff (1994: 63) calls "implicational rules".

⁵¹ All examples are from Harris (1999).

⁵² I assign *el* to class III since it does not include an overt class morpheme, contrary to *la* (/l+a/). The plural counterpart *los* (/l+o+s/), in contrast, is class I. I abstract from "neutral" *lo*, which does not precede nouns in Spanish, as well as from phonological alternations such as *el agua* 'the water', where selection of *el* rather than *la* seems to be due to hiatus avoidance (but see Baker 1992)..

⁵³ I follow Harris (1999) in considering /e/ and /Ø/ allomorphs for class III, hence my labels nIII and nIIIa (avowedly inspired from Bantuist descriptive practice).

⁵⁴ Naturally, the root $\sqrt{\text{puert}}$ that appears in *puerto* 'port' has to be counted as a homonym, etymology notwithstanding.

⁵⁵ I leave it as a moot point whether, when one does not wish to specify the cat's sex or does not know it, M is selected, having the whole kind as a submeaning, or M/F is not associated at all, $\sqrt{\text{gat}}$ being then assigned to class I by default. Somehow, the first alternative strikes me as more plausible and more in line with the present model.

⁵⁶ Naturally, derivational morphemes may link to each other as in Italian *linguistucolo* 'poor linguist' (I am grateful to Guglielmo Cinque, p.c., for this example), and that is even one of the (relatively) distinguishing criteria between derivational and inflectional morphology (see Scalise 1988; Dressler 1989). For this we may content ourselves with the standard explanation, viz. that $\sqrt{\text{vist}}$ (the first morpheme generally) having attached to the non-functional root, a new non-functional root $\sqrt{\text{lingvist}}$ is created to which further derivational morphemes may attach.

⁵⁷ Interpreting movement in terms of attraction is crucial here. Otherwise, moving first $\sqrt{\text{vist}}$ to nII would be countercyclic, and in order to comply with cyclicity we would have first to raise $\sqrt{\text{lingw}}$ to $\sqrt{\text{vist}}$, and then the complex $\sqrt{\text{lingw+ist}}$ to nII, thus distorting the, I think, important notion that *linguista*'s gender is /-ist/'s gender. If, on the other hand, $\sqrt{\text{vist}}$ and nII pertain to different cycles (or constitute distinct phases in Chomsky's 1999 terminology), then the attract vs. raise distinction might indeed be moot. (I thank Guglielmo Cinque for drawing my attention to this issue.) See Harris (1992: 74-75) for a discussion concerning the non-syllabicity of /u/ in the root I note as $\sqrt{\text{lingw}}$.

⁵⁸ This is obviously not the case when the meaning is "instrument" as in *colador* 'strainer'. /dor/ is then only associated with nIIIa in the lexicon.

⁵⁹ “Language specialist”, as the compositional meaning of $\sqrt{\text{lingw}}$ plus $\sqrt{\text{ist}}$ would sound, does not suffice by far to understand what a linguist is. In 17th century Portuguese (perhaps Spanish as well) a language specialist, a *linguista*, was an interpreter.

⁶⁰ “Piling up” leading, e.g., to the ungrammatical (but sensible) **linguistaa* for ‘woman linguist’ is probably made impossible in morphology.

⁶¹ Naturally, classification is also represented in the encyclopaedia, as the lexical features of, e.g., $\sqrt{\text{ka}}$, are not enough to determine the meaning of, say, *ka-lik* ‘fruit juice’.

⁶² Note that *fetsar* is an applicative derivation on the root $\sqrt{\text{fets}}$ ‘follow’, thus meaning ‘follow for/in favour of’. This raises an issue to which I will return.

⁶³ A similar argument has sometimes been made for gender, saying that perhaps at some time a door was conceived of as a feminine entity (metaphors can be imagined) or that what is now the privileged exponent of feminine gender (viz. /-a/) was not so at earlier stages. The latter is true, as Indo-European studies revealed long ago; the former is at best conceivable in a few cases, hopeless in many more, given the narrowness of the classification compared with what Manjaku and similar languages achieve. Be it as it may, the difference between modern Romance languages and modern Manjaku in terms of applicability of the classificatory criteria remains in its entirety.

⁶⁴ This relative divorce is what makes it possible for membership in a given inflectional class or gender to be a function of purely formal properties such as phonological form, e.g. in the Papuan languages Arapesh and Yimas (see Aronoff 1992, 1994) or perhaps French (see Corbett 1991: 57ff.).

⁶⁵ A clue in that direction is given by the contrast between *taktubiina* ‘you(Fsg) write’ and *katabti* ‘you(Fsg) wrote’: in the former, second person (i.e., the purely pronominal component of meaning) is spelled out through the prefix /ta-/ (compare *taktubu* ‘you(Msg) write’) while marked gender (F) and Number are expressed by the suffix /-iina/ (compare *taktubuuna* ‘you(Mpl) write’ and *taktubna* ‘you(Fpl) write’); in the latter, in contrast, person, gender, and Number – and presumably Tense as well – are all bundled together in the suffix /-ti/ (compare *katabta* ‘you(Msg) wrote’, *katabtum* ‘you(Mpl) wrote’, and *katabtunna* ‘you(Fpl) wrote’). See Benmamoun (2000) and Halefom & Lumsden (2000).

⁶⁶ We might assume (29) to be the basic structure of 3rd person pronouns in a language like English as well, with *he/she/it* the lexical expressions of *n* (thus accounting for *she-cat* and similar forms). I won’t pursue this matter.

⁶⁷ I do not consider here recent proposals by Chomsky (1999) to the effect that head movement (or Attract Head) should be entirely eliminated from syntax. If Chomsky’s arguments bear weight, and we wish to stand by our notion of the lexicon and its relationships to the other components of grammar, then we will have to enrich the lexicon with processes that mimic what head movement does in syntax. I am not sure it would be such a sweeping change in the model.

⁶⁸ Recall that “word” is a morphological unit that has no relevance outside morphology.

⁶⁹ The ‘+’ signs are mere indications of the fact that different exponents are merged or fused together. They do not necessarily correspond to syntactic brackets.

⁷⁰ It is remarkable that, not taking various “expansions” into account, prefixed noun class exponents in Niger-Congo always consist in one CV syllable. A further argument in favour of that solution is that noun class signs, although derivational, cannot stack: there is no such word as **/na-pë-lik/* meaning ‘someone having to do with wells’ (‘well-digger’, for instance). Yet, no general constraint on derivation prevents it (see, e.g., Matthews 1974: 164ff.).

⁷¹ Whatever explains it, word-form legitimacy seems to be a parochial property (compare, e.g., Spanish *gato* / *gata* with Catalan *gat* / *gata*).

⁷² That is to the exception of Chinese and similar languages where Number is not a necessary category within DP (see Chierchia 1998).

⁷³ As a minimum. There may also be a functional root meaning ‘Duality’ and, in some grammars, still another one meaning ‘Triality’.

⁷⁴ See Kaye, Lowenstamm & Vergnaud (1985) for the underlying theory of phonological segments I am assuming. If Class is nI (e.g., *gatto* ‘cat’) or nIII (e.g., *cane* ‘dog’), /I/ does not fuse with the Class exponent, but replaces it (*gatti*, *cani*). I leave this problem pending.

⁷⁵ I assume the /e/ of *paredes* ‘walls’ (from *pared* ‘wall’) to be epenthetical.

⁷⁶ Xârâcùù is a New-Caledonian language (see Moysse-Faurie 1995). Historically, there is a phonological relation between *œil* (Latin *oculum*) and *yeux* (Latin *oculos*). For present-day French speakers, however, the two roots are not phonologically related.

⁷⁷ The isolated case of *u-pats* ‘children’ ceases to be a problem if we view it as a collective (see French *marmaille* ‘gang of kids, brood’), which accords well with the abstracting meaning of /u/.

⁷⁸ We can, actually, but to a price, viz. having to perform meaningless morphological tricks.

⁷⁹ Consider indeed that knowing that *ka-√* has plural meaning with such and such roots would not be more complex than knowing that *bë-fetsar* denotes a human despite /bë/’s basic meaning(s).

⁸⁰ I abstract from the distinction between numbered and unnumbered Plurality, because it is not well explored enough. Taking it into account, I trust, would not modify the reasoning significantly.

⁸¹ Such generalizations are exploited, for instance for integrating loanwords.

⁸² The most widespread dialect of present-day Wolof seems to be one where class distinctions are fairly well preserved in the singular, whereas plural presents us with the contrast of one root associated with the Class sign $\{n \cap \text{Plurality}\} = /ñ/$ (*nit ñ-i* ‘the persons’) vs. all others associated with $\{n \cap \text{Plurality}\} = /y/$.

⁸³ In both cases, agreeing items preserve the “real” gender (see *šulxanot ktanim* ‘small tables’, *dvorim yafot* ‘nice bees’).

⁸⁴ Compare masculine *bakbuk male* / *bakbukim malim* ‘full bottle(s)’.

⁸⁵ Actually, nII and nIII share the same exponent $/Vt/$, except that $/t/$ is only realized in the so-called Construct State in nII (i.e., when V is $/a/$ – see *xatulat ha-šaxen* ‘the neighbour’s she-cat’), whereas it is stable in nIII (i.e., when V is another vowel than $/a/$). The two classes could therefore perhaps be united (but see Aronoff 1994: 184, fn. 32 for a sceptical assessment). I also assume that nouns (not numerous) like *sade* / *sadot* ‘field(s)’ constitute a subset of nII. Let me emphasize again the tentative nature of all this.

⁸⁶ Meaning simply that no nIV noun stands in the list (supposing it to be true).

⁸⁷ It is important not to lose sight of the fact that we are here operating at the infra-D level. The present analysis has thus no bearing on the discussion of whether N or (remnant) NP raises to D in the Semitic and Romance languages, except insofar as it shows that what moves to D, if anything, must be a maximal projection at some level. For recent developments, see Cinque (2000) and Shlonsky (2000).

⁸⁸ English, as already noted, is a somewhat complex case where Class is contentful (taking the form of gender) but is only manifest on singular 3rd person pronouns. What we might say, then, is that English is like Spanish, but Class has a null ($/\emptyset/$) exponent across the board, except when the merged root is pronominal and Number does not intersect, in which case there is suppletion: $/she/$ or $/he/$ if sex is relevant (encyclopaedically potent), $/it/$ otherwise.

⁸⁹ I am grateful to Jacqueline Leroy for putting Watter’s dissertation at my disposal and discussing the data with me.

⁹⁰ (39) may thus be compared with, e.g., Portuguese *duas cabeças de gado* ‘two head of cattle’, where the numeral agrees in gender with the “classifier” *cabeça(s)*. Naturally, differences are at least as important as similarities, hence my use of scare quotes.

⁹¹ Note that, while the present framework explains why fully functional Class elements such as Spanish word classes must surface as suffixes, it merely predicts that denoting Class elements can be proclitics or enclitics, but it does not explain why they are the one or the other in particular languages. More research is obviously required.

⁹² According to Foley (1986), this is also the case in Naisioi, implying that the same root may function as either a value of $\{n\}$ or an argument of $\{n\}$.

⁹³ I have little to say about classifiers in possessive constructions, except to remark that they rather look like the misnamed “verbal classifiers” in that they cross-reference features of one argument in the construction, generally the possessed noun (as in the one language in Africa having that type of object, viz. the Mba, Ubangi language ‘Dongo-ko – see Pasch 1985 ; Aikhenvald 2000, Chapter 5).

⁹⁴ To the difference of Romance, attributive adjectives *never* precede the noun in Manjaku; and recall there are no predicative adjectives, but quality verbs, sometimes formally different from the corresponding adjectives (see *U-ndali a fac* /3-cat Pro be-white/ ‘The cat is white’). As for the derived status of *faacal*, see above the discussion about *bë-fetsar*.

⁹⁵ Lest readers should resent it that this account offends Greed (see Chomsky 1995: 201ff.), it can easily be recast in the following terms: n°_x needs a Class feature; it targets the closest one, viz. n°_3 , and attracts it, pied-piping the whole nP (for some reason). In the Arabic example, note, the adjective (*Hasan*) does not agree with the noun *wajh*, but with the noun *Hasanu l-wajhi* is predicated of. That is to say, the Arabic phrase spells out a derived structure, whereas (39) is basic.