WHAT'S IN A NOUN : NOUN CLASSES, GENDER, AND NOUNNESS

Alain Kihm (CNRS, UMR 7110)

With a few exceptions, Niger-Congo languages are characterized by more or less extensive *noun class* systems. I give examples from three languages of the Atlantic subfamily :

(1)(a) **na**kiej / **ba**kiej 'robber(s)' vs. **u**ndaali/**ngë**ndaali 'cat(s)' (Manjaku)
(b) sàcc **b**i / sàcc **y**i 'the robber(s)' vs. muus **m**i / muus **y**i 'the cat(s)' (Wolof)
(c) gujj**o** / wuy**Be** 'robber(s)' vs. 'ullu**ndu** / 'ullu**ɗi** 'cat(s)' (Fula)

Noun class markers are set in boldface. Manjaku looks very much like Bantu languages, with noun class markers showing up as prefixes on the root (cf. Swahili ***mw****izi* / ***we****zi* 'robber(s)' – note : /we/ = /wa + i/). In Wolof, noun class markers are prefixed to the postposed definite determiner *-i* (or suffixed to what looks like a preposed indefinite determiner – cf. *a**b** sàcc* 'a robber' / *a**y** sàcc* 'robbers'). In Fula, noun class markers are suffixed to the root. This seems to cover morphological variation in the domain.

In these languages, the nominal vocabulary is thus partitionned into *x* formally marked classes (13 in Manjaku, 10 in Wolof, 25 in Fula), which yield a partly semantic classification : humans (no sex distinction) vs. animals (no sex distinction) vs. plants, etc. We must qualify with "partly" because in all languages many nouns are assigned to classes in a seemingly arbitrary fashion : e.g., Manjaku ***u****mbanj* / ***ngë****mbanj* 'knife/knives' denoting an artifact but assigned to classes 3/4, to use the traditional numbering, although these "normally" are the classes for animals – cf. ***u****ndali* / ***ngë****ndali* in (1a) – in the sense that the majority of nouns referring to animals belong to these classes (and the majority of nouns denoting artifacts belong to classes 9/10 – cf. ***ka****to* / ***i****to* 'house(s)').

As can be seen from the examples, noun classes are *paired*, one class having singular meaning and the other having plural meaning. Pairing is mostly regular, but there are exceptions : e.g., Manjaku ***na****päts* 'child' (class 1) the paired plural of which is not /****ba***päts/ (class 2) as expected (see ***na****kiëj* / ***ba****kiëj* in (1a)), but class 3 ***u****päts* 'children'. Such (regular) singular vs. plural pairs are sometimes called *genders* (so that, e.g., ***u****ndali* / ***ngë****ndali* 'cat(s)' would belong to gender II consisting of classes 3 and 4).

Noun classes command *agreement* (cf. Manjaku ***na****kiëj* ***na****mak* /1-robber 1-big/ 'big robber', ***ngë****ndali* ***ngë****mak* /4-cat 4-big/ 'big cats' ; Fula *'ullu**ndu** maw**ndu** **ndu'u*** /cat-11 big-11 this-11/ 'this big cat'). The scope of the agreement domain varies, as it may be limited to the noun phrase (Manjaku) or extend to the verb phrase of which the noun phrase is an argument (Fula, Bantu languages).

Noun classes have a derivational function : cf. Manjaku *kiëj*, Wolof *sàcc*, Fula *wujj*, all verbs meaning 'to rob', with respect to which ***na****kiëj*, *sàcc **b**i*, and *gujj**o*** may be considered Agent nominalizations (see Mufwene 1980).

These properties should be compared with the properties of gender systems as they are found in Indo-European and Afroasiatic languages. Gender systems differ from noun class systems in the following ways : (a) there are only 2 or 3 genders ; (b) classification is based on sex which is of course only relevant for animate beings. As far as inanimates (or not obviously sexed or unfamiliar animates such as insects

and giraffes) are concerned, arbitrariness seems to reign supreme (cf. French, German, Arabic), much more so than in noun class systems.

Yet, there are similarities between the two types of systems. First, gender like noun class commands agreement in and beyond the noun phrase (cf. Italian ***Il gatto nero è uscito*** 'The black (tom)cat went out' vs. ***La gatta nera è uscita*** 'The black she-cat went out' ; Arabic *xaraja l-qiTTu l-'aswad^{u}* /exit.PF.3sgm the-cat.m the-black.m/ 'The black (tom)cat went out' vs. *xaraja**ti** l-qiTT**at**u s-sawd**aa**^{u}* /exit.PF.3sgf the-cat.f the-black.f/ 'The black she-cat went out' ; German *Der schwarze Kater / Die schwarze Katze ist ausgegangen*). Secondly, singular-plural pairings are observed (cf. Italian *gatt**o*** / *gatt**i*** vs. *gatt**a*** / *gatt**e***, Arabic *xaadim^{un}* / *xaadim**uun**^{a}* 'manservant(s)' vs. *xaadim**a**^{tun}* / *xaadim**aat**^{un}* 'maidservant(s)'), although not unexceptionally (cf. Portuguese *gat**o*** / *gat**os*** vs. *gat**a*** / *gat**as***). Thirdly, gender assignment may fulfil a derivational role, although a much more limited one than noun class assignment does : cf. French *relieur* 'binder' (a person) vs. *relieuse* 'binder' (a machine), and see Ritter 1991 about Hebrew. Finally, gender markers are always realized as suffixes in the same way as noun class markers are in some language (e.g., Fula) with the interesting difference, though, that postposed noun class markers look rather more like enclitics than suffixes. More generally, noun class languages partake of agglutinative morphology, whereas gender languages, at least of the I-E and Afroasiatic families, are fusional. The correlation is probably significant.

The similarities between noun class and gender languages are sufficient, it seems, to warrant the following leading hypothesis : noun classes and genders jointly represent the grammatical expression of a basic faculty/activity of mind, namely *classification*. Moreover, being classified constitutes the signature of nounness. Nouns are formally classified according to ontological (or encyclopaedic) properties of their referents, whereas verbs never are (see Jackendoff 1983 for a possible characterization of ontological categories). I know of no language where verbs bear special markings in keeping with intrinsic properties of the actual events they denote, such as being dangerous, pleasant, etc., or where there would be, say, one verb 'to fly' with three different morphological markings according to whether it is birds, or men, or clouds that fly. When such distinctions are maintained, then entirely different verbs are typically used, indicating that 'birds-fly' and 'men-fly' are conceived of as different events, just like a bird is a different entity than a plane (for a real example, cf. Wolof *tuur* 'to pour (water)' vs. *xelli* 'to pour (tea or coffee)'. This is not classification, but mere naming. One also finds verbs such as French *mettre bas* 'to whelp' and *accoucher* 'to give birth' which are normally predicated of certain types of entities only, female pets and women respectively. But there is nothing obligatory in this connexion, and few cat lovers in our urban culture will have qualms saying *Ma chatte a accouché de trois chatons* 'My cat gave birth to three kittens'. It is thus a matter not of grammar, but of social convention.

From this leading hypothesis the following assumption results : the set {Class}, the union of noun classes and of genders, represents the content of the grammatical category Noun. Our task is thus to provide a unitary account of the substantial and formal similarities and differences just pointed out. The morphosyntactic model we shall apply to this end aims to cumulate the results of two types of approaches : Distributed Morphology (see Halle & Marantz 1993) as far as the overall architecture of grammar is concerned ; processual models of morphology (see Anderson 1992 ; Aronoff 1994 ; Beard 1995) for the nature of morphological operations.

Linguistic forms proceed from the interaction of three components : the lexicon, syntax, and morphology (that includes a major part of phonology). The lexicon is a set (of sets) of elements (or features in the broad sense of the term meaning linguistic properties – see Chomsky 1999) which are either substantial or grammatical. Substantial elements or *roots* are endowed with encyclopaedic meaning, i.e. they have a denotation (Frege's *Bedeutung*, extension) and ultimately a referent in some mental world, including the "real" world. Grammatical elements have no denotation outside the linguistic faculty (although they may be indirectly linked to mental world referents – cf. [Plural] and multiplicity), they only have grammatical meaning (Frege's *Sinn*, intension). They may be called functional features or f-features, making it clear that lexical f-features so defined are neither coextensive with nor theoretically equivalent to the syntactic functional categories of the Principle & Parameter and Minimalist theories.

Elements, i.e. roots and f-features, combine in the lexicon and in the morphosyntactic interface (see Beard's 1995 distinction between derivation and inflection). They are "abstract" in two ways : (a) they are devoid of phonological substance, which comes through morphological association of a *feature bundle* [root + f-feature(s)] with an exponent in the vocabulary ("late insertion", syntax only manipulates features) ; (b) roots are not categorized as nouns or verbs, they acquire a syntactic category as argument (head of NP) or predicator (head of VP) by combining with the elements *n* (pro-noun) and *v* (pro-verb). At this stage, I use the cover term "element" for *n* and *v* so as not to prejudge their identification as roots or f-features.

Our assumption may thus be stated as follows : {Class} is the content of *n* (see Kihm to appear). More accurately, the lexicon includes a set {n} the members of which are classified values of *n*. In the limit case of "classless" or "genderless" languages such as Bambara or Hungarian, {n} is a singleton set with the sole value Nounness. A set {v} will also be assumed, the values of the members of which are, say, subcategorization features. For a given member n_i of {n}, combining it with a root R_i may thus a priori occur either in the lexicon, prior to syntactic insertion, or after syntactic insertion. A secondary assumption I will make is that only *n* (and *v*) directly (lexically) combine with roots ; f-features such as Number combine with *n*, hence the derivational character of pluralization (see Beard 1995) and the general syncretism of class/gender and number morphemes (see Renault 1987). Tense would combine with *v*, except for the very likely possibility that Tense is a root and is always inserted in syntax (see Chomsky 1999). I have nothing more to say on this issue. As a result, combinations of elements in the lexicon consist in (i) combining f-features with members of {n} or {v} ; (ii) combining the resulting sets {n/v + f} with roots. This suggests that sequences of lexical operations always terminate with application to a root, which we may call the *Terminal Root Principle* (TRP). (I take no firm stand here about the possibility of roots combining in the lexicon to yield, e.g., compounds or incorporations. For my present purposes, I only need to consider root–f-features combinations, and I assume that roots do not combine lexically.)

What I loosely termed "lexical combination" must now be made more precise. Combination is a set-theoretic operation that may take two forms for which I will use the DM terms "fusion" and "merger", with some reinterpretation. Fusion consists in properly including a set within another set, while merger consists in forming a new set that includes two or more sets :

(2) Fusion : {A {B}}
Merger : {{A}{B}}

There is no third possibility, except noncombination, to which I return. Given the TRP combining operations always result in a (set of) f-feature(s) fusing or merging with a root, as shown in (3) :

(3) Fusion : {R {F}}
Merger : {{R}{F}}

As we shall see presently, "choosing" a particular combination mode has consequences for morphological realization. Moreover, empirical evidence suggests that, at least in this domain, merger is the unmarked option. I will assume this.

Once all lexical operations have been completed, fused or merged complex sets are projected or copied as nodes of syntactic X-bar structures. By definition, fusions or mergers always consist in one root and a set of f-features, and they are thus inserted as X° level nodes. Let us now apply all preceding assumptions to the problem at hand, beginning with gender systems such as that of Italian or Arabic.

Gender basically involves the contrast of two classes (sometimes three) traditionally labelled "feminine" and "masculine", but which we might just as well designate by numbers (1 and 2) or, following the logic of markedness (see Jakobson 1932/1966), by [+] and [–]. Given this and the other properties listed above, it is clear that gender is a semantically impoverished category, largely devoid of encyclopaedic content. It therefore qualifies as an f-feature. Insofar as it determines the values of the members of {n} – say n_f and n_m in Italian and Arabic – its function beyond assigning nounness is to distinguish formal word classes in the sense of Aronoff (1994) and Harris (1999), whose connexion to worldly properties is at best indirect, via the encyclopaedic content of the root (compare *gatta* 'she-cat' and *casa* 'house', *qiʈʈa*^{tun} 'she-cat' and *šawka*^{tun} 'fork'). From this functional character of gender and the unmarked character of merger we deduce the fact that gender-valued *n* must merge with roots. The lexical representation of Italian *casa* is then as follows :

(4) {{R_{kas}}{n_f}}

and that of its plural *case* 'houses' is as in (5) :

(5) {{R_{kas}}{n_f {Pl}}}

where the lexical f-feature [Plural] first fuses with n_f. (In these representations, subscript "kas" is a pure label identifying the root for its encyclopaedic correlates.) Note in passing that the unmarked option would be to merge [Plural] with *n* as in Portugese *casas* whose lexical representation is {{R_{kas}}{{n_f}{Pl}}} or in Turkish *evler* 'houses', lexically identical to *casas*, except that *n* has no Class value.

Both (4) or (5) are now fit to identify N° nodes – more accurately, to identify X° nodes in a syntactic structure as N-valued. The implied view of syntax is that of a template of hierarchized positions or sites – see Marandin 2000. Note that *n* having merged in the lexicon is not represented in syntax. As mentioned earlier, lexical functional elements are conceptually different from functional categories, which they render partly, but only partly otiose (see below).

Following Chomsky (1999) (but also see Bird 1995) I assume Spell Out, i.e. morphological interpretation, to be cyclic, occurring every time a feature bundle the morphological component can (and therefore must) handle has been assembled. In (4) and (5) the condition is trivially met. (Notice that head movement becomes dispensable under this view, perhaps entirely.)

In the present model, the morphological component comprises two parts : the morphosyntactic interface (MSI – cf. Anderson's 1992 Morphosyntactic Representation or MSR) and the vocabulary. The MSI's function is to linearize the feature bundles delivered by syntax, i.e. to translate the inclusion relations of (4) and (5) into left-to-right immediate precedence relations. This is necessary given the temporal nature of phonological objects. Clearly, the most straightforward translation is to have merged elements follow the elements they are merged with, which we may indicate by substituting angled for curly brackets in the MSI :

(6) $\{\{R_{kas}\}\{n_f\}\}_{MSI}\rightarrow \langle\langle R_{kas}\rangle\langle n_f\rangle\rangle$

A crucial assumption here is that Linearization sees only down to merger and bypasses fusions, with the upshot that lexical fusion always has syncretism or internal inflection (e.g., Arabic internal or "broken" plurals) as a morphological correlate. For (5) we thus get (7) at the MSI :

(7) $\{\{R_{kas}\}\{n_f\ \{Pl\}\}\}_{MSI}\rightarrow \langle\langle R_{kas}\rangle\langle\{n_f\ \{Pl\}\}\rangle\rangle$

where fused Plural is not linearized with respect to n_f. Merger, on the other hand, always correlates with suffixation.

Linearized MS representations are then spelled out, i.e. given phonological substance. With the empirical material we are working on it is not really crucial whether Spell Out is accomplished as in DM (see Halle & Marantz 1993 ; Halle 2000) by associating feature bundles / morphemes with phonological exponents specified for a maximal subset of the features of the morphemes they are paired with (e.g., /e/ ↔ [+Fem, +Pl]) ; or as in Anderson (1992) and Aronoff (1994) via Word Formation Rules which may have the forms of functions (see Raffelsiefen 1992). Since the actuality of so-called nonconcatenative morphology seems rather to support the latter view, I will adopt it in its functional version (see Kihm in preparation). That is to say, Spell Out is conceived of as a function that takes feature bundles as its domain and returns phonological objects. Following Lowenstamm (1996) I define a phonological object as an indefinite sequence of C and V positions such that every odd-numbered position is a C and every even-numbered position is a V, and there is at least one C and one V (CV^+). I therefore call the function the CV-function (f_{CV}). F_{CV} applies to linearized feature bundles, so that unlinearized (fused) features do not receive particular phonological expression as required, and it applies recursively to successive feature bundles. Composed application of f_{CV} to the linearized products of (6) and (7) thus returns (8) and (9) :

(8) $(\langle\langle R_{kas}\rangle \ldots\rangle (f_{CV})\rightarrow /C_kV_aC_sV/ \times \langle n_f\rangle (f_{CV})\rightarrow /CV_a/) \rightarrow /C_kV_aC_sVCV_a/$
(9) $(\langle\langle R_{kas}\rangle \ldots\rangle (f_{CV})\rightarrow /C_kV_aC_sV/ \times \langle\{n_f\ \{Pl\}\}\rangle (f_{CV})\rightarrow /CV_e/) \rightarrow /C_kV_aC_sVCV_e/$

Later phonological processes ensure that uninterpreted positions are correctly erased, yielding /kasa/ and /kase/ (pronounced [káza] and [káze]). Needless to say, subscript letters in (8) and (9) actually are abbreviations for matrices of phonological elements.

Let us now turn to noun class systems, beginning with Manjaku which may be considered the polar opposite of Italian to some extent. The properties of noun classes outlined above lead us to view them as *roots*. That is to say, {n} in Manjaku and like languages (e.g., Bantu) is a set $\{n_i, n_j, \ldots, n_n\}$ of generalized roots (pronouns) denoting natural ontological categories (human, animal, etc.) potentially able to classify every entity in all possible worlds. In fact, the classifications actually achieved never are exhaustive nor fully consistent, but this raises no theoretical problem – after all, natural languages are handled by human beings whose intellectual productions (from myths to scientific theories) are characterized by extensive *bricolage* in the sense of Claude Lévy-Strauss. That noun class systems are not optimal (compared with, say, Dewey's classification, itself falling well short of exhaustivity and full consistency) is thus to be expected. (Note, however, that at least one class pair, viz. 1/2, is consistent insofar as it only includes nouns referring to persons, even though not all nouns referring to persons belong to it.)

The assumption that we are dealing with pro-nominal roots is strongly supported by the fact that in Manjaku and most noun class languages of the Niger-Congo family, noun class exponents also appear in the function of subject or object clitic pronouns. The evidence is especially clear in Mankanya, a member of the dialect cluster Manjaku really consists in (see Trifkovic 1969). We thus find :

(10) Ba daan
2 drink
They (human) drink
(11) Ka-toh ka joot-i
7-house 7 crumble-Pf
The house crumbled
(12) Dë thuuman ka
I fill 7
I fill it (any object in class 7, e.g. ***ka****kana* 'calabash')

Similar observations can be made in Fula and in Bantu languages, even though they are not always so neat because clitic pronouns (aka subject/object affixes) often have different phonological forms than the noun class exponents to which they correspond – but cf. Fula *Colel-ngel ngel majjii* /bird-Class Class got-lost/ 'The bird got lost' (Sylla 1982 : 76) and Changana *Chi-pichi chi balek-ile* /7-cat 7 run.away-Pf/ 'The cat ran away' (Ribeiro 1965 : 443).

Being roots, class-marked members of {n} do not need to combine lexically with roots. This does not mean they cannot combine, since they actually can as we shall see, but that noncombination is the unmarked option in this case. The lexical configuration corresponding to, e.g., singular ***na****kiej* 'robber' may thus be represented as follows :

(13) $\{_{LEX}\ \{n_1\} \ldots\ \{R_{kiej}\}\}$

where $\{_{LEX}\ \}$ is the whole lexical set, and $\{n_1\}$ (i.e., the member of set {n} denoting the ontological category Person – I use traditional class numbering) and $\{R_{kiej}\}$ are two roots singled out for syntactic insertion. Recall that $\{R_{kiej}\}$ cannot be selected alone since it lacks category and cannot function as a syntactic object. The other option is to select *v* (a functional element, not a root) and fuse it with the root ($\{R_{kiej}\ \{v\}\}$), thus producing a feature bundle that is capable of identifying a V° node (cf. *A*

kiej ***u****mbanj* ‘S/he stole the/a knife’). Short of that, the only possibility is to select a member of {n} simultaneously with selecting $\{R_{kiëj}\}$. With this root, it has to be $\{n_1\}$ or $\{n_9\}$ yielding the action noun ***pë****kiej* ‘act of stealing, theft’ ; the array is wider with other roots – cf. ***m****lik* ‘water’ ($\{n_6\}$), ***ka****lik* ‘fruit juice’ ($\{n_7\}$), ***pë****lik* / ***i****lik* ($\{n_9\}$ / $\{n_{10}\}$) ‘well(s)’, and the verb *lik* ‘to draw water’. With this root, we see that the choice of a particular member of root {n} is actually crucial for assigning its encyclopaedic meaning to the word. With other roots that always appear in the same class and cannot be used as verbs, probably for semantic reasons, e.g. $\{R_{sa}\}$ referring to ‘cloud’, there must be a meaning postulate stating that selecting $\{R_{sa}\}$ entails or correlates with selecting $\{n_7\}$, hence ***ka****sa* / ***i****sa* ‘cloud(s)’. Both roots are then concurrently projected into syntax.

Linking a class element (functional or root) with a root, be it through lexical combination or syntactically as here, can be thought of as functional application. That is to say, $\{n_1\}$ can be viewed as a function that applies to $\{R_{kiëj}\}$ to return a noun glossable as “person implied in stealing”. (Similarly, *v* is a function that applies to the root to return a verb meaning ‘to steal’). Notice that the semantic relation thus achieved has to be left indeterminate, since it is not constant – cf. ***a****buk* / ***ba****buk* ($\{n_{1a}\}$ / $\{n_2\}$) ‘child(ren)’ related to *buk* ‘to give birth’. Here the noun denotes the result of the event, not the agent (the mother).

The syntactic instantiation of functional application is c-command. (This is implicit in all current formulations of X-bar theory.) Therefore, root {n} must identify a position that c-commands the position identified by the root. This gives us (14) as a representation for ***na****kiej* :

(14) $[_{nP}\ [\{n_1\}\ [_{XP}\ \{R_{kiej}\}]]]$

I prefer this developped structure – even though there is no real use I can think of for the specifier positions it offers – to the more minimal $[\{n_1\}\ \{R_{kiej}\}]$ because in the latter both elements c-command each other, and we may want the linearization function of the MSI to be sensitive to asymmetrical c-command when it deals with syntactically assembled material, in accordance with Kayne’s (1994) Linearization Correspondence Axiom (LCA). Note that $\{R_{kiej}\}$ has no category to assign to its projection, which I accordingly label XP. Syntactic category is assigned by $\{n_1\}$ to its own projection. A crucial step now is to consider that we have a full feature bundle in (14) – and none before it – so that (14) may, therefore must be interpreted by the morphological component as soon as it is assembled. In this way, nP turns out to be equivalent to an X° node, i.e. a morphological unit.

The MSI then converts (14) into the linearized object $<\{n_1\}\ <\{R_{kiej}\}>>$ to which f_{CV} applies to return the phonological object $/C_nV_aC_kV_{ie}C_jV/$. We thus directly derive the fact that, in the unmarked case, the exponent of root {n} surfaces as a prefix to the exponent of the root.

What about plural ***ba****kiej* ‘thieves’ ? In everything that precedes I implicitly (and standardly) assumed that singular is the default interpretation resulting from not associating {n} with the lexical functional element [Plural]. The assumption seems to be correct. The solution to plural formation is then to assume that, as in Italian, [Plural] fuses with {n}, and the fusion $\{n_1\{Pl\}\}$ is spelled out as /ba/, $\{n_3\ \{Pl\}\}$ as /ngë/ (cf. ***u****ndali* / ***ngë****ndali* ‘cat(s)’), and so forth. In other words, $\{n_2\}$, $\{n_4\}$, even-numbered class labels in general are synonymous with $\{n_1\{Pl\}\}$, etc. A more graphic representation would then be to label classes according to their basic ontological

categories, e.g. {n_P}, P = person, instead of {n_1}. I will keep to traditional numbering, however.

The well-known difficulty with plural formation in noun class systems, however, is that singular-plural pairings are not entirely regular, so that, as already mentioned, class 1 ***na****pats* 'child' has its plural in 3 (***u****pats*), which is normally the *singular* class for animals, rather than in 2. Actually, this makes us suspect that ***u****pats* is in fact a collective independently related to {R_{pats}} (cf. French *marmaille* 'brood'), which removes it from consideration for the problem at hand. More to the point, there is the case of ***bë****fetsar* 'friend' (incidentally a noun with human denotation not in 1, but in 5, normally the singular class for plants) the plural of which is ***i****fetsar* 'friends' in 8, normally paired with 7, the artefact class, rather than in 6 (cf. ***bë****calam* / ***m****calam* 'wild mango tree(s)').

What I wish to emphasize, however, is that the possibility of such discrepancies is actually predicted by the present model. Indeed, since selecting the root and selecting a particular member of {n} are two separate operations, there is no model-internal necessity that the same {n_i} be targeted each time a given root is selected. What there is is external, semantic regularity : a cat's ontology is to be an animal, and several cats are no less animals than is one, so {n_1} or {n_1{Pl}} ought to be selected whenever reference is to a cat or cats. This is why meaning postulates seem to be the right conceptual tool to capture this relation. But again, this is regularity born from a sense of ontological stability, it is not (conceptual) necessity. Deviations and exceptions which native speakers have to learn one at a time are therefore allowed and actually observed. Should they become too numerous, the system would crumble – which is apparently occurring in Wolof. In Manjaku, however, it is still functional. Children encountering ***bë****fetsar* for the first time only have to learn that (i) they must discard {n_5}'s ontological category in this particular case ; (ii) they have to abstract from the root's denotation as well as from regular pairings and select {n_7{Pl}} (= {n_8} – see above), despite {n_7}'s ontological category, when reference is to several friends. It would be interesting to know whether children acquiring Manjaku and similar languages show a tendency to regularize such items, and whether they do it according to the root's denotation (producing */**ba**fetsar/ in the plural) or according to regular pairings (producing */**m**fetsar/ in the plural). Agreement with such nouns is apparently variable : "Avec le pluriel irrégulier *upäts* (enfants) on a le choix entre : *upäts uties* et *upäts baties*" (With the irregular plural *upäts* (children) one has the choice between *upäts uties* [little children] and *upäts baties*) (Buis 1990 : 23). Unfortunately I lack more precise data on this issue. In Wolof roots denoting persons tend to pluralize in the personal plural class /ñ/ whatever class they are assigned to in the singular (cf. *jigéen* ***j****i* 'the woman' > *jigéen* ***ñ****i* 'the women', not */jigéen **y**i/, unless disparagement is intended – see Cissé 1998 : 42).

Notice that singular-plural discrepancies are not limited to noun class languages. In Hebrew one finds nouns like *šulxan* 'table' which, although masculine, take the feminine plural *šulxan****ot***. Agreement is then according to the singular's gender (*šulxan****ot*** *ktan****im*** 'small tables' – for an analysis, see Kihm to appear). Again, this is made possible by the sheer fact that [Plural] combines with {n} rather than directly with the root.

We thus accounted for the Manjaku (or Bantu) noun class system by assuming that the lexical elements that express Class are roots rather than functional elements. Naturally, the same analysis must hold for languages like Fula where noun classes exhibit exactly the same characters as they do in Manjaku or Bantu. The

crucial difference, however, is that noun class exponents show up as suffixes in Fula. In our model this implies we must allow for the marked possibility of roots combining in the merger mode in the lexicon, i.e. choosing what is the unmarked mode for functional elements. The lexical representations of *'ullu**ndu*** / *'ullu**ði*** 'cat(s)' are thus (15) and (16) (I use Arnott's 1970 numbering) :

(15) {{$R_{'ullu}$}{n_{11}}}
(16) {{$R_{'ullu}$}{n_{11}{Pl}}} ({n_{11}{Pl}} = class 25)

The alternative solution would be to project {n_i} in syntax along with the root, as in Manjaku, and then head-move the root to the left of {n_i}, an operation for which there is no motivation I can see, and which moreover we wish to dispense with entirely on theoretical grounds.

The question we should ask is, how did this difference between Manjaku's unmarked system and Fula's marked system arise ? Wolof may give us an answer. As we saw in (1) Wolof noun class exponents attach to the postposed determiners and demonstratives (cf. *muus **m**ii* 'this cat', *muus **m**oom**u*** 'the cat in question', etc.). There is a diachronic explanation to this state of affairs (see Doneux 1975b), and it is that the original class prefixes were absorbed by the roots, so that they now only show up on formerly agreeing modifiers, hence the frequent "rhyming" effects that are observed whenever the prefix was preserved as an initial (now meaningless) syllable – e.g., ***ji**géen **j**i* > *jigéen **j**i* 'the woman' (compare Manjaku ***u**ndaali **w**i* 'this cat'). (Quite possibly, rhyme was henceforth used to assign new roots to classes, as may be the case with *muus*.) What it means in synchronic terms is that the lexical elements interpreted as determiners or demonstratives merge with the members of {n}. The lexical representation of ***j**i* is thus (17) :

(17) {{n_j}{D}}

Here again we need a meaning postulate to tell us that when {$R_{jigéen}$} is selected {n_j} must be selected. Actually "meaning" is perhaps too precise a term insofar as Wolof noun classes are by and large semantically bleached, although some correlations can still be perceived (cf. *koko **b**i* 'the coconut' vs. *koko **g**i* 'the coconut tree') and the two classes whose exponents are /k/ and /ñ/ only go with person-denoting roots. Whatever is the case, there has to be a lexical (and mental) operation that relates the identification of the root to that of particular members of {n}.

To return to Fula now, what seems to have occurred is that lexical complexes such as (17) were reanalysed as simple members of {n}, as if Wolof ***j**i* became ***ji***, an unanalysable class marker, without changing their position with respect to the root. Support for this hypothesis comes from the fact that determiners in Fula appear with some variation as duplications of the class markers : cf. *'ullu**ndu** ndu* 'the cat' (Sylla 1982 : 43 ; Arnott 1970 considers these forms near demonstratives).

The most serious problem with Wolof is that, to the difference of Manjaku and Fula, roots can be inserted bare into syntax as nouns with indefinite or generic reference (cf. *Dama bëgg muus* /T-1sg like cat/ 'I like a cat/cats'). How is nounness assigned to roots in this case ? Note first that in the example just given *muus* cannot be interpreted as specific indefinite 'a certain cat' which must be expressed as ***am** muus* or (Dakar Wolof) *benn muus* (literally 'one cat'). The generalization is that whenever a noun is used with specific reference, be it definite or indefinite (on this see Lyons 1999), it is accompanied by a representative of {n} (which in the Dakar

Wolof variant is /b/ prefixed to the numeral – numerals and pronominals are two domains where Wolof has kept noun class prefixes, one more reason to consider {n} a set of roots as in Manjaku ; unfortunately there isn't enough time for a more detailed analysis). My proposal is then that it is the uncategorized root that is inserted in this case, and that it acquires nounness by default, by virtue of occupying an argument position. Being limited to genericity, i.e. reference to kind (whatever satisfies the description "be a cat"), is a correlate of this bare root character. Not all grammars permit this. In French, where gender is also associated with determiners, bare nouns are mostly impossible (cf. **J'aime chat*) and genericity has to be indirectly expressed via number (*J'aime les chats* 'I love cats'). English is similar in this respect (cf. **I love cat*). Perhaps this points to the fact that, as a general property, bare roots cannot be mobilized for syntactic use in I-E (and Afroasiatic) languages.

Let me now conclude on Wolof. First, there is the question of singular-plural pairings. To the eight singular classes having /b/, /g/, /j/, /l/, /m/, /s/, and /w/ as their respective exponents, there corresponds one plural class spelled out as /y/. The noun class spelled out as /k/ can be defined as personal even though it includes the very ill-fitting *këf* 'thing', since the other noun in its purview is *nit* 'person', and it always denotes persons when pronominal elements merge with it (cf. ***k**an* 'who ?' vs. ***l**an* 'what ?'). Plural /ñ/ is personal throughout as it only goes with person-denoting roots, although not all, including *nit* (cf. *nit* ***ñ**i* 'the persons') and excluding *këf* (cf. *yëf* ***y**i* 'the things', where initial /y/ in *yëf* is another remnant of noun class prefixes).

These complex data suggest the following solution. As already indicated, the lexicon must be provided with meaning postulates of the form "if A, then B", e.g. "if $\{R_{jigéen}\}$, then $\{n_j\}$". Semantic properties of the root may drive the meaning postulate, but they are often obscure (see Cissé 1998). Phonological properties of the exponent, i.e. the shape of the initial syllable, also play a role. (Note : the fact that roots have no PF in the lexicon is no problem, since ours is basically a parallel representation model, so that the lexicon and the vocabulary are simultaneously available and information from the one can be used in the other.) This is how it works when [Plural] is not activated. When [Plural] is selected and fuses with {n}, on the other hand, the meaning postulate changes, and it now says, "If $R \in \{P\}$, then $\{n_P\{Pl\}\}$; if $R \in \{\neg P\}$, then $\{n_{\neg P}\{Pl\}\}$". Here {P} is a set of designated roots having the common property of denoting persons (but not all person-denoting roots belong to {P}), and $\{\neg P\}$ is the complementary set in $\{_{LEX}\}$; $\{n_P\{Pl\}\}$ is the feature bundle that has /ñ/ for an exponent, whereas $\{n_{\neg P}\{Pl\}\}$ spells out as /y/. Whatever value of {n} is pointed to by the meaning postulate when [Plural] is not selected is irrelevant. Of course, *nit* belongs to {P}. Wolof thus makes a rather sharp application of the principle we saw at work in Manjaku, namely that {n} selection with and without concomitant [Plural] selection are independent operations.

The other issue we must deal with is that of the syntactic ordering of the lexical merger $\{\{n_i\}\{D\}\}$ or $\{\{n_i\{Pl\}\}\{D\}\}$ and the root. A crucial observation here is that those mergers are not ordered relative to the root itself, but relative to the noun phrase headed by the root. Not only is this the case on theoretical grounds – since a determiner can only take an NP as its complement – but it is an empirical fact as shown by relative constructions such as (18) where the determiner follows the NP *jigéen ju dem* 'woman who left' :

(18) jigéen **j**u dem **j**i
woman J-Rel leave J-D
the woman who left

Final ordering is therefore syntactically determined. Let us assume the syntactic structure of *jigéen (**j**u dem) **j**i* to be as in (19) :

(19) $[_{nP/DP(/Pl)}\ [_{n/D(/Pl)}\ \{\ \}][_{NP}\ [_{N°}\ R_{jigéen}]\ \ldots\ ]]$

Here the root is assigned its features, i.e. nounness, definiteness, and possibly number, through c-command. All we have to assume, then, is that for some reason I will not try to assess here NP has to raise to Spec nP/DP (or perhaps to a higher Spec of what would count as a proxy projection – see Nash & Rouveret 1995). Once this has taken place, there is a feature-complete syntactic domain (perhaps a "phase" in the sense of Chomsky 1999) which can and must be linearized and spelled out by the morphological component.

Conclusion

Genders and noun classes represent two instantiations of a general property I call Class. Class is the semantic content of the lexical element *n* that assigns nounness to roots. Being classified and being a noun are two descriptions for the same reality. The lexical element *n* is therefore a set {n} the members of which are so many values of Class. In gender systems with two or at most three values of Class, *n*'s content is impoverished, so the element functions as a functional feature that merges with roots in the unmarked case. In the marked case, if fuses. Merger and fusion are viewed as set-theoretic operations : merger consists in including two or more sets into a new set ; fusion in including a set within another set. In noun class systems with more values of Class, *n*'s content is rich and members of *n* function as generalized roots (pro-nouns) inserted into syntax along with designated roots in the unmarked case (but lexically merged with roots in the marked case). The relations between particular members of *n*, be they functional or root, and particular roots are governed by meaning postulates. Other features modifiying nouns, e.g. [Plural], fuse or merge with *n*, not directly with the root. This explains why singular-plural relations are sometimes skewed, because the meaning postulate from $\{n_i\}$ to a given root may be different from the meaning postulate from $\{n_i\{Pl\}\}$ to this same root, although the normal state of affairs is that they are identical.

Finally, I ought to say a word on agreement, limiting myself to noun-modifier agreement within NP in NC languages – but my proposal is probably extendable to argument-predicate agreement in the same languages. It seems to me all we need is the very simple, and actually traditional, assumption that modifiers, i.e. basically determiners and adjectives, are functional elements or roots lexically or syntactically combined with {n}, just like nouns, except that {n} is not valued in their case. What it means is that determiners and adjectives mean or denote properties ("being definite", "being red", etc.) that are not (and cannot be) a priori assigned to particular classes of entities. A root such as, e.g., Manjaku $\{R_{mak}\}$ denoting bigness thus enters the MSI as $\{\{n_x\}\{R_{mak}\}\}$, with a variable for a value of {n}. In Manjaku, contrary to, say, English, {n} must be valued, because if it is not, it cannot be spelled out, and if it is not spelled out, the root itself cannot be spelled out. Agreement then occurs to rescue this state of affairs and – that is my proposal – it is no more than feature copy *in the MSI* from the member of {n} combined with the modified root. Much more would obviously have to be added to do justice to the full complexity of the phenomenon, but such, I trust, are the basics.

Anderson, Stephen, R. 1992. *A-Morphous Morphology*. Cambridge : Cambridge University Press.
Arnott, D.W. 1970. *The Nominal and Verbal Systems of Fula*. Oxford : Clarendon Press.
Aronoff, Mark. 1994. *Morphology by Itself : Stems and Inflectional Classes*. MIT Press, Cambridge, Mass.
Beard, Robert. 1995. *Lexeme-Morpheme Base Morphology : A General Theory of Inflection and Word Formation*. Albany : State University of New York Press.
Buis, Pierre. 1990. *Essai sur la langue manjako de la zone de Bassarel*. Bissau : INEP.
Chomsky, Noam. 1999. "Derivation by Phase." Unpublished ms., MIT.
Cissé, Mamadou. 1998. *Dictionnaire français-wolof*. Paris : L'Asiathèque.
Corbett, Greville. 1996. *Gender*. Cambridge University Press, Cambridge.
Doneux, Jean-Léonce. 1975a. *Lexique manjaku*. Dakar : CLAD.
– 1975b. "Hypothèses pour la comparative des langues atlantiques", *Africana Linguistica VI*, 88, 41-129.
Halle, Morris. 2000. "Distributed Morphology : impoverishment and fission". In J. Lecarme, J. Lowenstamm, & U. Shlonsky (eds), *Research in Afroasiatic Grammar*, 125-149. Amsterdam : John Benjamins.
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. 1999. "Nasal Depalatization *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.
Jackendoff, Ray. 1983. *Semantics and Cognition*. MIT Press, Cambridge, Mass.
Jakobson, Roman. 1932/1966. Zur Struktur des russischen Verbums. In E.P. Hamp et al. (eds), *Readings in Linguistics II*, 22-30. Chicago : University of Chicago Press.
Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge (Mass.) : MIT Press.
Kihm, Alain. To appear. "Noun classes, gender, and the lexicon-syntax-morphology interfaces." In G. Cinque & R.S. Kayne (eds.), *Handbook of Comparative Syntax*. Oxford : Oxford University Press.
– In preparation. "The internal constituency of internal plurals". Ms, CNRS, Laboratoire de Linguistique formelle (UMR 7110).
Lowenstamm, Jean. 1996. "CV as the only syllable type". In J. Durand & B. Laks (eds), *Current Trends in Phonology : Models and Methods*, 419-442. Oxford : Oxford University Press.
Lyons, Christopher. 1999. *Definiteness*. Cambridge : Cambridge University Press.
Marandin, Jean-Marie. 2000. "L'hypothèse des sites en syntaxe". In J.-M. Marandin (ed.) *Cahier Jean-Claude Milner*, 175-227. Paris : Verdier.
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.
Nash, Léa & Alain Rouveret. 1995.
Raffelsiefen, Renate. 1992. "A nonconfigurational approach to morphology". In M. Aronoff (ed), *Morphology Now*, 133-162. Albany : State University of New York Press.
Renault, Richard. 1987. "Genre grammatical et typologie linguistique." *Bulletin de la Société de Linguistique de Paris* 82 : 69-117.
Ribeiro, Armando. 1965. *Gramática changana (Tsonga)*. Caniçado : Editorial "Evangelizar".
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.
Sylla, Yèro. 1982. *Grammaire moderne du pulaar*. Dakar : Les Nouvelles Editions africaines.

Trifkovic, Mirjana. 1969. *Le mancagne : étude phonologique et morphologique*. Dakar : IFAN.

Alain Kihm
CNRS, Laboratoire de Linguistique formelle
UFR de Linguistique, Université Paris 7
2, place Jussieu 75005 Paris
alain.kihm@linguist.jussieu.fr