

Plural Number as a Lexical Derivation (Wordformation)*

Robert Beard

glossa 16:2 (1982)

Abstract: *None of the theories of plural number as they presently stand adequately represent all number phenomena of IE languages. Of the descriptivist, syntacticist and lexicalist theories, the latter as described by Halle and Lieber holds the most promise. However, in order to explain all number phenomena—count nouns, mass nouns, plural nouns, pluralis tantum and collective nouns, we must look at pluralization as a process separate from the morphological process that assigns the suffixes which mark it. Once number is perceived as an abstract, lexical process, it becomes the same for all IE languages and all its manifestations may be described in terms of two lexical features plus a handful of rules manipulating them. The interesting implication of this is that there must be two types of lexical extension rules: one which adds lexical (semantic) features to a stem and another which merely adjusts the values of features which are already present.*

Number has traditionally been taken as an inflectional category in a class with gender and case among IE languages. For example, the most recent Soviet Academy grammar (Svedova et al. 1980: 471) claims that 'the nominal morphological category of number is an inflectional category, expressing itself in a system of two opposing forms—the singular and plural number.¹ Along with the majority of descriptivists, the grammarians of the Soviet Academy of Sciences agree that number is one of three categories reflected in nominal inflectional paradigms—number, gender and case—which they distinguish sharply from the processes of lexical (L-) derivation or word formation.

A more recent American view has been that number is a syntactic feature '... introduced by a context-free rule of the base applying to nouns', while case features are '... introduced by a rule that does not belong to the base subcomponent of the syntax at all but rather to its transformational part' (Chomsky 1965: 171-172). The most detailed exposition of this approach is probably Chapter II of Crockett's *Agreement in Contemporary Standard Russian* (1976). Both Chomsky and Crockett distinguish number from the 'inherent' lexical feature, gender, thereby producing a system consisting of (1) inherent lexical features such as gender, (2) base component features such as number, and (3) transformational features like case. Since all three of these types of categorial features are simultaneously represented by a single morpheme in inflectional languages (e.g. the Slavic FemNomSg *-a*), these thinkers must posit a separate morphological (M-) component subsequent to the transformational component which can interpret all these features before inserting the appropriate paradigmatic endings.

The third approach to number is implicit in Halle's position that inflection itself is a lexical process

much like L-derivation (Halle: 1973). Halle concludes this after noting that the same facets of semiproductivity characterizing L-derivation are found among inflectional paradigms. Lieber (1981) has extended Halle's position to conclude that number is a lexical feature of what she terms 'inflectional stem allomorphs'. Lieber would store the singular and plural stems of all nouns independently and relate them via 'morphological' rules which identify redundant lexical relations and relate these lexical stems to their common root. Halle's insight that inflectional and derivational morphemes are essentially homogeneous is crucial to understanding number in the IE languages, but it does not lead ineluctably to the conclusion that inflection is a lexical subsystem. In fact, there is evidence preventing such a conclusion which we will see further on. Before examining it, we must see why the idea of syntactically determined number must be abandoned.

One reason some theorists have been attracted to the idea of syntactically determined number is the misconception that it is highly regular and productive. Sampson (1979: 46), for example, claims that '... it seems factually untrue that rules of derivational morphology ... can be assimilated to semantically regular, productive 'rules of grammar' such as the rule pluralizing nouns; the former rules are typically not fully productive and the meanings of the words produced are not in general predictable'. Sampson must support the 'syntactic' definition of number since his assumptions are those of Chomsky (1970), who would restrict syntactic rules to 'fully productive' processes and seek explanations of less than fully productive processes in the lexicon. The critical assumption here is that pluralization is a 'fully productive' therefore syntactic process.

It is indisputable that pluralization exhibits all of the types of irregularity and semiregularity associated with lexical derivation. One finds among the plural of all IE languages both formal irregularity, e.g.

- (1) deer, moose, quail, fish, salmon
 ox-en, child-r-en
 women, men, feet, mice, lice
 live-s, wive-s, hoove-s, knive-s
 alumn-i, foc-i, embol-i, bacilli-i
 dat-a, cili-a, phenomen-a, criteri-a
 alumn-ae, antenn-ae, hyperbol-ae
 bas-es, indic-es stas-es, matric-es

as well as semantic irregularities and subregularities such as

- (2) skie-s, heaven-s, feeling-s, look-s (∅ semantics)
 oat-s, mump-s, measle-s, green-s, grit-s (∅ semantics)
 rain-s, snow-s, water-s, sand-s (amount)
 scissor-s, plier-s, tong-s, (lexical class)

Although these types of irregularity and subregularity seem to firmly establish a relation between inflectional and derivational morphology, we cannot conclude with Halle that inflectional morphology itself is a lexical process, since (1) transformations determine surface case markings to a great extent and (2) since the lexicon and morphology are handled by two different parts of the brain: the former by the posterior linguistic, or 'Wernicke's', region, the latter by the anterior linguistic, or 'Broca's', region.² The preferable theory of number will be amenable to these facts which support a separate posttransformational M-component that both assigns the affixes and adjusts allomorphically the morphemes of both inflectional and lexical derivation. I will assume such a component henceforth (see

Beard 1981 for details). Note that I also assume that meaning is not directly related to morphemes, but is grammatically related, i.e. grammar intervenes between meaning and affix. The M-component thus has nothing to do with meaning; meaning is controlled by lexical and syntactic rules such as the pluralization rule (13).

At this point I would like to dwell on the syntactic treatment of number in Chapter II of Crockett's work (Crockett 1976). Crockett notes that some nouns have 'inherent' number, i.e. lexically determined number. P But she would limit inherent number specifications to *pluralis tantum* stems corresponding to such English lexemes as

(3) Pluralis tantum

pants	(oats)	measles	pliers
shorts	(grits)	mumps	tongs
trousers	(greens)	hives	shears
slacks	(collards)	blues	scissors

Stems with variable number are still treated syntactically, following the early Chomskyan tradition. This results in a variable syntactic plurality feature in addition to the fixed lexical one. Since it is the syntactic feature which ultimately determines suffixation, Crockett posits a rule of agreement which adjusts the unmarked syntactic plurality feature [+Plurality] whenever it dominates a lexical item marked <+Plurality>. Then to finally define variable number, Crockett introduces a reference feature, {±Aggregate}, which the speaker controls during speech, and a rule of syntactic agreement which marks the syntactic plurality feature positive, [+Plural] just in those instances when the speaker perceives his referent to be {+Aggregate}.

Crockett's system is appealing in its ability to handle number and agreement in Russian with a single lexical feature, [+Plural]. rather than two, [+Singular, ±Plural]. However, her system runs afoul of singularis tantum, i.e. mass nouns, which never pluralize.

(4) *Singularis tantum noun*

air, fleece, dirt, peace, contemplation, cleanliness, waving, semantics, mahogany, hate, pork

To account for these nouns, Crockett introduces a fourth feature, another lexical one, <±Homogenous> <±Homogeneous>. Stems bearing this feature are excluded from pluralization under normal circumstances.

Although Crockett presents a most thorough analysis of number, and although she avoids the assumption of a feature [±Singular]. her explanation of all number phenomena in Russian relies on four features: two lexical, one syntactic and one, presumably, performative. By 'all number phenomena' I mean specifically (5.).

(5) (singular nouns)

(plural nouns)

count nouns

mass nouns

pluralis tantum nouns

collective nouns

For the 4 phenomena of number which must be explained (singular and plural are variations of count

nouns), Crockett has introduced 4 features at three grammatical and extragrammatical levels: not much economy. And there is also good reason to feel uneasy about the introduction of reference features into P-markers, since it is difficult to see how syntactic morphological processes, hitherto restricted to agreement rules, can provide a decision mechanism controlling number. If the decision as to whether a singular or plural form is to be used in a sentence is left to the judgment of the nonideal individual speaker, the choice of singular or plural in the grammar must be random.

The acceptance of a posttransformational, prephonological morphological component devoid of semantics sets the stage for understanding the alternative to Halle's 'lexical inflection'. Crockett's idea of plurality as a potentially inherent lexical feature materially advances that understanding. If we assign now the M-component the capacity to insert affixes independent of lexical and inflectional derivational processes, number may be seen in isolating and inflectional languages as a purely lexical function.⁴

L-derivation now becomes a wholly abstract process which assigns meanings to new L-derivates—but not affixes—in the lexicon. That is, the conditions on L-derivation may be substantially different from the conditions on the affixes which mark it. Further, case inflection is the result of case features introduced in the base component and subjected later to T-rules which may greatly alter them (e.g. the nominative and accusative cases may be replaced by the genitive, or instrumental and genitive cases, respectively, during nominalization or gerund formation). Both the accumulation of case, gender and number features, on the one hand, and productive L-derivation features, on the other, are marked by the same M-component which must stand behind the T-component but before the P-component.⁵ Assuming such a general theory of grammar, then is number a category of L-derivation or inflection?

In inflectional languages, nouns which are susceptible to any case are susceptible to all cases. Inflectional case seems to be 'fully productive' in Chomsky's sense; thus, if pluralization is fully productive, we should suspect its being an inflectional category. But we have seen that pluralization shows all the types of irregularity which characterize L-derivation. The effects of what Karcevskij termed 'morphological asymmetry' are evident in the fact that the plural nouns of (1) reflect formal irregularity independent of the semantic irregularity of (2). That is, the formal and semantic irregularities do not coincide; we do not find semantic irregularity everywhere and only where we find formal irregularity (the definition of subregularity). The constraints on the semantics of pluralization are discrete from those on the morphology of affixation as is the case with L-derivation.

The second argument for pluralization as a lexical process is that in highly inflectional languages like Russian, Latin and Sanskrit, pluralization leads to a shift in paradigm. The plural paradigm is not the same as the singular paradigm of the same lexical item, differing only in some plural morpheme, as is the case in agglutinative languages; rather, the plural paradigms are wholly different from the singular ones. The difference, in fact, at least equals the differences between the paradigms of the various singular lexical classes, i.e. masculine, feminine, neuter, etc. Now, if we define gender, a category of purely lexical classes, in terms of the paradigm each gendered noun is subject to, to be consistent, we must agree that plural nouns belong to lexical classes different from those of their singular counterparts. In Russian, for example, gender is not even maintained in the plural; here, the plural paradigm can only be considered a fifth lexical class. Since inflectional processes are incapable of changing the lexical class of a noun, and since this is precisely what L-derivations do, we must conclude that plural, too, is an L-derivation.

The third argument for pluralization as a lexical process comes from the fact that constraints on pluralization are always lexical constraints. A lexical stem of any gender is susceptible to all the cases of

that gender's paradigm, but thousands of singular nouns cannot be pluralized at all and in every IE language there are dozens of plural nouns which cannot appear in the singular. Moreover, there are stems which cannot structurally pluralize

(6) deer, fish, sheep, moose

but which semantically do, and those which structurally pluralize but semantically do not.

(7) skie-s, feeling-s, look-s, heaven-s

Whether a stem is subject to formal or semantic pluralization or singularization is strictly a matter of the lexicon, a matter of lexemic properties."

The fourth argument comes from the historical development of IE languages. In languages otherwise as divergent as Bulgarian, English and Hindi, but where the inflectional system has completely atrophied, the genitive ending has been replaced by a preposition or postposition, e.g. Bulgarian *na*, English *of*, Hindi *ká/ke/kí*.

(8)	Bulgarian	gramatika na b"lgarskija ezik 'grammar of the-Bulgarian language'
	Hindi	paṇḍiton ká ghar 'the house of the pundits'

The genitive *-s* seems preserved as an inflectional ending in English, but it behaves more like an asyllabic postposition than an ending. For instance, like prepositions and postpositions, it governs phrases, not lexemes, e.g. *the king of England's hat*.

The plural marker, by contrast, remains a lexical suffix in languages which have lost inflection and behaves like an L-derivation suffix. Thus one may not say in English "**the many king of Englands*" only "*the many kings of England*". The plural in both Bulgarian and Hindi remains marked by suffix rather than by preposition or postposition.

(9)	Bulgarian	kost-i 'bones', set-a 'villages, agn-et-a 'lambs'
	Hindi	ghoṛ-oṅ 'horses', pothi-yán 'books', rát-eṅ nights'

Oriental languages like Chinese and Vietnamese show us that number is just as dispensable as case in natural languages, yet even where inflection completely disappears, number perserveres along with L-derivation, marked by the complete reduplication of the lexeme. In these languages it is impossible for number to be inflectional.

There are two more arguments supporting a lexical treatment of number in inflectional languages, bringing the total to six. The fifth argument lies in the fact that number markings may be borrowed. Endings such as *-i*, *-a*, *-ae*, *-es* of (1) are all borrowed in just the same sense that *-ive*, *-ion*, *-ment*, and *-ible* are borrowed morphemes. Since lexemes are frequently and easily traded by one language to the

next, it is predictable that L-derivational affixes may also be borrowed. However, there is no borrowing of inflectional endings: no borrowed past tense ending, say, or third person singular endings in English. Latinate verbs are treated here the same as Germanic ones. In this respect, too, the plural behaves more like an L-derivation than an inflectional category.⁷

The sixth and final argument for plural's being a lexical category is the behavior of the relative pronoun in inflectional languages. The relative pronoun in Russian, *kotorij*, for example, agrees with the noun it modifies in gender and number. It agrees in case, however, according to its position in the relative clause in which it occurs.

- (10) Ja videl sobaku, o kotoroj ty včera govoril
I saw the-dog about which you yesterday were-speaking
 'I saw the dog you were speaking about yesterday.'

For example, in (10), *kotoroj* is feminine singular locative: feminine and singular in agreement with the feminine singular accusative *sobak-u* 'dog', but locative in agreement with the preposition *o* 'about'. Number once again aligns itself with lexical gender in contrast to the syntactic category of case.

The evidence is overwhelming in its support of number being a lexical rather than inflectional category. The features it operates on are lexically determined and constrained, they follow the historical patterns of lexical rather than inflectional categories, they exhibit the idiosyncrasies characteristic of L-derivation and they are never controlled by the features of case. Since, however, number, gender and case are represented in inflectional languages by a single ending, it must be true that plural is determined by abstract rules, the results of which are combined with the results of other rules for interpretation by the M-component.

But what sort of abstract rule controls plural nouns if they are generated by an L-rule? To establish the nature of this rule, we must first determine the number and nature of the features it will operate upon. We have just established the nature of these features: they are lexical features. But is there only one, [\pm Plural], or is there a second feature, [\pm Singular], as well?

If we posit only one lexical feature, [\pm Plural], as does Crockett, we imply that all nouns must be either singular or plural. On the other hand, if we posit two features with variable values corresponding to Jakobson's concepts of marked and unmarked, i.e. [\pm Plural] and [\pm Singular], we would predict four lexical number classes in addition to nouns with variable number: (1) nouns which are singular and not plural, (2) nouns which are plural and not singular, (3) nouns which are neither singular nor plural and (4) nouns which are both singular and plural.

What do we find in the IE languages? There are certainly mass or singularis tantum nouns which are singular and never plural and p/uralis tantum nouns which are plural and never singular. But there are also nouns which are both singular and plural: collectives, whose lexical status is morphologically singular, but semantically plural.

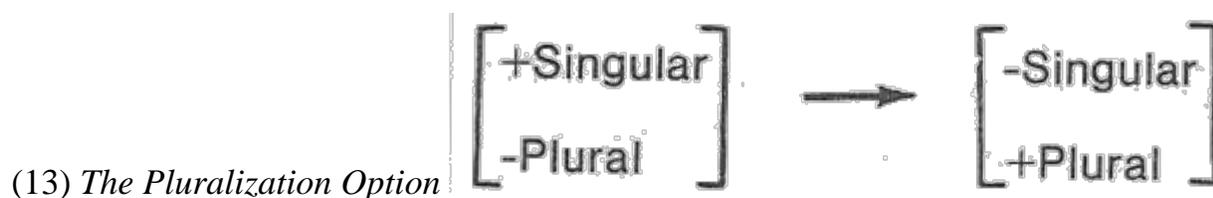
- (11) the rich, the poor, the undernourished, the peasantry, humanity, officialdom, the wounded

In the Slavic languages, where collectives are much more productive, not only are collectives

morphologically singular, they govern singular agreement (syntactic singular). The collective nouns in the following table of Serbocroatian number phenomena all govern singular verbs and adjectives, while the plural forms govern the syntactic plural.

(12)	<i>Singular</i>	<i>Plural</i>	<i>Collective</i>	<i>Gloss</i>
	dugm-e	dugm-et-a	dugm-ad	'button'
	bur-e	bur-et-a	bur-ad	'barrel'
	tel-e	tel-ic-i	tel-ad	'calf'
	gran-a	gran-e	gran-j-e	'limb'
	snop	snop-ov-i	snop-lj-e	'bundle'
	drv-o	drv-et-a	drv-eć-e (drv-et-j-e)	'tree'
	list	list-ov-i	lišć-e (list-j-e)	'leaf'

There is no question but that the collectives are used to refer to aggregate referents, just as the plural form is; however, agreement is in the singular as is the morphology of the noun. The simplest explanation of this is that these stems are marked [+Singular, +Plural). This implies a convention whereby any stem marked [+Plural] must be interpreted semantically as referring to plurality or 'aggregation'. However, the morphological component must then contain a convention which provides that a stem be marked with a plural ending or paradigm only if it is both [-Singular] and [+Plural]; otherwise, it should be marked singular. This seems inevitable regardless of the nature of our theory, but it presents fewer problems for our system since ours allows independent constraints on derivation and affixation.



Constraints will be performative ones, i.e. (13) may be used only when the speaker wishes to refer to more than one instance of the set of objects which the lexical item in question names. Pluralis and singularis tantum nouns, on the other hand, permit no such option; thus, they will not be allowed access to this rule, although they may be susceptible to other L-rules.⁸

It is interesting to note that (13) differs from previous conceptions of L-rules in that it does not involve any change in meaning, only a shift of value among the lexical features. In comparison to the possessional adjective derivation, e.g. (14):

(14)	beard	bearded
	two + head	two-headed

tinge	tinged
grass	grassy
moss	mossy

which adds the meaning of possession and shifts the lexical class of the nouns it operates on, the pluralization rule much like feminization adds no obvious intensional meaning. It does change the range of the objects the lexeme might meaningfully refer to, as well as its subclass.

The generative lexical component, we must conclude, consists of at least two qualitatively distinct types of rules, in addition to any performative rules such as the one mentioned earlier: one type adds semantic content, presumably lexical features; the other type merely adjusts the value of features already contained in the lexeme. It is interesting to note that affixation (the M-component) does not distinguish between the two types of rules, neither in terms of the types of morphemes it adds to the derivation nor in terms of productivity. Pluralization is relatively productive in all IE languages, so is the possessional adjective. Feminization is highly productive in Slavic languages, so is the agentive.

We must conclude that the same component which derives lexemes must also handle pluralization; however, it does not follow from this that the same component assigns the appropriate lexical or inflectional affix. The *-S* which productively marks plural in English also marks the 3rd person singular of verbs and is a possessive enclitic. But if the possessive is transformationally derived in gerundives like 'John's proving the theorem', then assignment of this suffix must be delayed until after all transformations are performed. The alternative is to postulate two or more spurious morphemes, $-S_1, -S_2, \dots -S_n$, inserted at various levels of the grammar. Considering the fact that homophonic morphemes would have the same allomorphic variations regardless of function or level, the greatest generalization is accomplished by assuming them to be a single morpheme inserted by a component which stores it, inserts and adjusts it immediately before it is phonologically interpreted. This way its Insertion may mark the inflectional category of third person singular, the plural lexical derivation and the possessional syntactic case, while its essential unity and uniqueness is preserved.

NOTES

*This is a revised version of a paper delivered before the annual meeting of the LSA, December 27-29, 1981, New York, N.Y. I am grateful to Mark Aronoff, Gerald Greenberg, D. Robert Ladd, and John Nerbonne for comments which have strengthened it.

¹The translation is my own. The original reads: '*Morfologičeskaja kategorija čisla suščestvitel'nyx—èta slovoizmenitel'naja kategorija, vyražajuščajasja v sisteme dvux protivopostavlennyx rjadov form—edinstvennogo i množestvennogo čisla.*'

²More precisely, Wernicke's area seems to coordinate lexemic vocabulary during sentence-making. Lexemes or lexemic information may be stored in various parts of the brain, but not in Broca's area. The anterior linguistic region specifically stores grammatical morphemes.

³Kurylowicz (1964: 21) notes that number, like adjectival degree and verbal aspect, is 'semantic', by which he means 'paradigmatic' rather than 'syntactic'. Nominal case, verbal tense and adjectival declension are the latter. But Kurylowicz is apparently distinguishing paradigmatic categories from category markings acquired by agreement transformation. Kurylowicz does not distinguish, as does Crockett, between fixed (inherent) and variable (syntactic) plurals.

⁴Halle himself (Halle 1974) has posited such a posttransformational, prephonological M-component, but only with the powers of allomorphic adjustment. His theory of lexical inflection does not save grammatical theory a component (morphological), but simply

redistributes componential functions.

⁵Notice that I am speaking only of productive processes and assume that all idiomatic information is stored in some 'permanent lexicon' which may be accessed by behavioral rules other than L-rules. See Beard 1981 for details.

⁶A matter of lexemic properties as opposed to a matter of semantic properties. *Sand*, for example, logically should not pluralize. But if it is possible to pluralize *sand* formally or semantically, it is logical that this possibility would extend to *dirt*, *soil*, *gravel*. That *sand*, *sky*, *water*, *heaven* have the capacity to pluralize while *gravel*, *dirt*, *soil*, *air*, *atmosphere* apparently do not (at least in the same sense as *sand*, etc.) is a purely lexical fortuity.

⁸There will be other such rules. For example, in languages where collectives like those of (12.) are prevalent, there will be a rule on the order of



The constraints on this rule will no doubt be more numerous and more restrictive than those on the pluralization rule (13), but it is normal for L-rules to vary greatly in their productivity.

REFERENCES

- Beard, R. 1981. *The Indo-European Lexicon: A Full Synchronic Theory*. North-Holland Linguistic Series 44. North-Holland. Amsterdam.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press. Cambridge, Mass.
- Chomsky, N. 1970. 'Remarks on Nominalization.' In Jacobs and Rosenbaum. 1970. 184-221.
- Crockett, D. 1976. *Agreement in Contemporary Standard Russian*. Slavica. Columbus. Ohio.
- Flier, M. (ed.) 1974. *Slavic Forum: Essays in Linguistics and Literature*. Mouton. The Hague.
- Halle, M. 1973. 'Prolegomena to a Theory of Word-Formation.' *Linguistic Inquiry* 4.3-16.
- Halle, M. 1974. 'Remarks on Slavic Accentology.' In Flier. 1974. 17-41.
- Jacobs, R. and P. Rosenbaum (eds). 1970. *Readings in English Transformational Grammar*. Ginn and Company. Waltham, Mass.
- Karcevskij, S. 1929. 'Du dualism asymmetrique du signe linguistique.' *Travaux du Cercle Linguistique de Prague* 1. 88-93.
- Kurylowicz, J. 1964. *The Inflectional Categories of Indo-European*. Carl Winter. Heidelberg.
- Lieber, R. 1981. *On the Organization of the Lexicon*. Indiana University Linguistics Club. Bloomington, Ind.
- Sampson, G. 1979. 'The Indivisibility of Words.' *Journal of Linguistics* 39-47
- Svedova, N. et al. 1980. *Russkaja grammatika, tom I*. Nauka. Moscow.