Acta Linguistica **35** (1985): 117-131

Lexical Categories and X-Bar Features

András Kornai

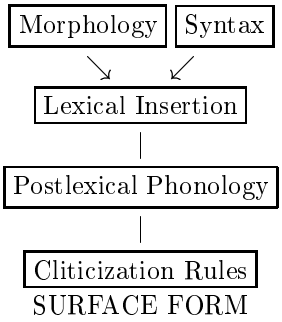
0 Introduction

With the wide-spread acceptance of the lexicalist hypothesis (Chomsky 1970) or, as it is better known today, the Lexical Integrity Hypothesis (LIH), morphology has become an integral part of generative linguistics. Yet in spite of the enormous progress made possible by the introduction of formal models, certain fundamental questions of morphology are still unresolved. One such question conserns the status of lexical categories: what do we mean by the terms 'noun', 'verb', etc.? This is the subject matter of this paper.

Another, and perhaps even more important problem area is the definition of words.The classical definition 'maximal domains between potential pauses' appeals directly to the intuition of the speakers: this can be supplemented by the investigation of the domains of various phonological processes like stress placement, vowel harmony, etc. The *phonological words* defined this way usually happen to be

(1) A) minimal free forms (Bloomfield 1926)
B) maximal stable forms (Bloch, cited in Hockett 1958:19.4)
C) maximal fixed internal order domains (see e.g. Matthews 1974:162ff)
D) maximal non-recursive domains (see e.g. Matthews 1974 loc. cit.)
E) anaphoric islands (Postal 1969)

There is no logical reason for these domains to coincide: theoretically, it should be possible to find phonological words that satisfy A), C), and E), but not B) and D); or to find constituents satisfying A)-E) that do not happen to be phonological words. But of the 64 theoretical possibilities, only five or six are attested, and with the introduction of a few supplementary concepts like compounding, cliticization, and bracket retention (tmesis), this variety can be reduced even further: the remaining types are frequently called 'morphological word', 'lexeme', etc. Now, it is precisely this 'coincidence' that makes it possible to organize the grammar in the following manner:



In what follows, it will be assumed without further argumentation that morphology supplies fully formed words (see e.g. Jensen and Strong-Jensen 1984) which are inserted into the terminal nodes

of independently generated syntactic trees. A word w can be inserted under a node n if they are of the same lexical category: in addition to this the *morphosyntactic features* of n and w must be the same. For instance, 'boy' cannot be inserted under a node V, or under a node N<+ PLURAL>.

The paper is divided into three sections. Section 1 outlines a purely morphological approach to the problem of lexical categories, which is exemplified by a brief description of Hungarian nouns. Section 2 gives a more formal proposal: this is illustrated by a detailed description of the nominal paradigm in the proposed framework. Section 3 is devoted to X-bar features: here the category system of Hungarian is analyzed.

I would like to acknowledge here my indebtedness to the participants of the Veszprém Morphology Conference and in particular to Farrell Ackerman, László Kálmán, and Ferenc Kiefer.

1 Paradigms and word classes

In describing the grammar of a language, we have to make reference to certain lexical categories. Some of these, e.g. the category 'adjective' in English, have psychological reality in the sense that speakers of the language can grasp the concept easily after the presentation of a few examples and counterexamples, and make highly consistent judgments afterwards. Other categories, though highly relevant for the grammar, are not susceptible of such psycholinguistic testing: this seems to be the case with closed (or just very small) categories like pronoun, auxiliary, etc. But even if our grammar appeals only to psychologically real categories, we do not necessarily know what characteristics of a word make the speakers assign it to one category or another. In addition to what might be called the *taxonomical* problem of classifying individual words, we also have some *ontological* problems: What is a lexical category? How many categories are there in any given language? Is there a universal set of lexical categories, and if not, how can we identify categories cross-linguistically?

The traditional answer, generally accepted until the end of the last century, was based on the idea of *class-meanings.* But in spite of its long history and great intuitive appeal, this idea could not withstand the criticisms levelled at it by the structuralists: "The school grammar tells us, for instance, that a noun is 'the name of a person, place, or thing'. This definition presupposes more philosophical and scientific knowledge than the human race can command, and implies, further, that the form-classes of a language agree with the classifications that would be made by a philosopher os scientist. Is *fire,* for instance, a thing? For over a century, physicists have believed it to be an action or process rather than a thing: under this view, the verb *burn* is more appropriate than the noun *fire.* Our language supplies the adjective *hot,* the noun *heat,* and the verb *to heat,* for what physicists believe to be a movement of particles in a body. (...) Class meanings, like all other meanings, elude the linguist's power of definition, and in general do not coincide with the meanings of strictly defined technical terms. To accept definitions of meaning, which at best are makeshifts, in place of an identification in formal terms, is to abandon scientific discourse." (Bloomfield 1933:16.2)

The structuralists based their solution to the problem of lexical categories on the notion of distributional equivalence: two items belong to the same class if they can be substituted for each other in every context in which they appear. This definition (for a more detailed version, see e.g. Harris 1951 ch 15) is better suited for morphemes than for words, since fully formed words with different inflection usually appear in widely differing contexts. In some cases, e.g. verbs with various person-number affixes, the situation might be saved by appealing to partial morphemic similarity and complementary distribution, but in others (e.g. the finite vs. the infinitival forms of a verb) only *ad hoc* solutions can be found.

Another problem with this solution is that it provides no basis for a feature analysis of the resulting categories. The distributional regularities of the resulting classes cannot be captured in terms of (binary) features in any obvious manner, and the number of 'natural' classes of categories is too small to provide a basis for feature analysis. This point, i.e. the arbitrariness of the so-called *X-bar features* of lexical categories was noted by Kean (1976), who shows that such features cannot be motivated syntactically or semantically. Yet generative grammarians persist in using X-bar features, and this creates a need for a theory which makes the development and the comparison of such feature analyses a feasible task. This problem will be discussed in Section 3.

The central claim of this paper is that lexical categories and their feature analyses belong to the domain of morphology: in particular, the definition of lexical categories should be based on the (word-) *internal* distribution of stems and affixes rather than on the *external* distribution of words (or morphemes) within sentences. In order to make this claim more precise, and to facilitate comparison with the structuralist definition, I will try to phrase it in purely distributional terms. First it should be noted that under the LIH, sentences can be segmented into words, since lexical insertion has to operate on fully formed words. I will assume that sentences can also be segmented into morphemes (by, say, the methods outlined in chs 12-13 of Harris 1951), and that every word contains an integral number of these. Moreover, it is assumed that the phonological rules of the grammar are formulated in such a manner that every combination of morphemes, when entered to the (morpho)phonological component, will give rise to phonologically possible words, and that speakers of the language are capable of deciding whether a given combination can be a word of the language or not. (This will make it possible to employ a somewhat loose terminology that makes no distinction between words and their morphological makeup.) These assumptions are shared by the majority of existing generative morphological models, like Aronoff (1976), Lieber (1981), Kiparsky (1982), Selkirk (1983).

The first distinction to be made is between *stems* and *affixes.* Since stem + stem and stem + affix combinations can give rise to possible words, while affix + affix combinations usually can not, it is possible to classify morphemes as stems or affixes solely on the basis of their word-internal distribution.[1]

The second distinction to be made is between *derivational* and *inflectional* affixes. Although it is true that in general inflectional affixes are farther from the stem than derivational ones, this does not give us sufficient leverage to distinguish between the two, and additional criteria must also be employed. There is no need to go into details here, because the ultimate object of the enterprise is to set up *paradigms,* and in this, morphological considerations are of secondary importance. In what follows, I will make liberal use of the criterion suggested by Anderson (1982): "Inflectional morphology is what is relevant to the syntax."

Now, given a set of stems and another (in the ideal case, disjoint) set of inflectional affixes, the paradigm of a stem or a word can be defined as the set containing those stem + affix(n) combinations that give rise to possible words of the language in question. If a word containing, say, a stem s and two (inflectional) affixes a and b can be subjected to further affixation, than the paradigm of $s + a + b$ is simply the intersection of the paradigm of s with those words that contain both a and b (in some order). If a paradigm of a word cannot be expressed as an intersection of this sort, it will be

[1] The actual 'discovery procedure' will be somewhat complicated, because in general more than one affix can be present in a word, and because certain elements can be affixes in one word and stems in others.

called *irreducible;* otherwise, it is *reducible* to its superset paradigm.[2] Since the traditional notion of paradigms is captured by the irreducible paradigms in this model, the adjective 'irreducible' will be dropped from now onwards. After these preparations, the definition of lexical categories becomes a trivial matter:

(2) Two stems (or words) belong to the same category if and only if their paradigms contain the same inflectional affix-combinations.

(3)

				0 = NOM
				t = ACC
	0			nak = DAT
	om	0		val = INS
	od		0	ért = CAU
	a			vá = TRA
	unk			on = SUE
	otok	ék		ra = SBL
sógor	uk		é	ról = DEL
	aim			ban = INE
	aid			ból = EAL
	ai			ba = ILL
	aink		éi	nál = ADE
	aitok			hoz = ALL
	aik			tól = ABL
	ok			ig = TER
				ként = FOR

The paradigmatic forms of the noun *sógor* 'brother-in-law' are given under (3): the same affix combinations (modulo assimilation and vowel harmony) can be used with more than 98% of the 35171 items classified as nouns in the seven-volume 'A Magyar Nyelv Értelmező Szótára' (Explanatory Dictionary of the Hungarian Language, 1959).[3] The results are even better if idiom chunks (e.g. *jotta* 'iota' in *egy jottányit sem...* 'not even a iota') are removed from the corpus. For the treatment of the remaining 'defective' nouns, see Section 2.

Since definition (2) is fully operational, it is possible to apply it to languages that fall outside the scope of the traditional Word and Paradigm model. For instance, in a purely isolating language we will have no inflectional affixes, so every word will belong to the same category. In this limiting case, the separation of morphology from syntax will not simplify the grammar at all, and the same is true for an ideally synthetic language where every sentence contains but one word. In general, the

[2] A few intuitively reducible paradigms turn out to be irreducible under this definition, and it is necessary to introduce the concept of *suppletion* at this point.

[3] All lexicographic data were obtained from a lexical data-base system that has been developed jointly by László Éltető (Computer Science and Automation Institute) and the author (see Éltető 1985). The system uses data from the Reverse-Alphabetized Dictonary of the Hungarian Language (Papp 1969) which, in turn, was based on the Explanatory Dictionary.

complexity of the category system is directly proportional to the (average) number of morphemes in words: this effect remains unexplained if the definition of lexical categories is based on syntactic or semantic considerations.

2 Morphosyntactic features

Since the paradigmatic forms of a given stem are determined (modulo phonology) by the affixes, they are usually encoded with the aid of binary features: the + value of such a *morphosyntactic feature* is usually taken to represent the presence, and the - value the absence of the corresponding affix in the given word form. For example, *fiúk* 'boys' will be represented as fiú<+ PL> rather than fiú<- SG>, and *fiú* 'boy' as fiú<- PL> rather than fiú<+ SG>. In the simplest case, the marked (+) value of a feature corresponds to some phonologically non-null affixal marker, and the unmarked value encodes the lack of surface marking.[4] Markedness (in the Prague School sense) will play an important role in what follows, and when this simple principle fails (e.g. because both members of an opposition are marked on the surface), more complex arguments will be given.

First it should be noted that in general there is no one-to-one correspondence between affixal morphemes and morphosyntactic features. The same feature, e.g. ± PL, might well correspond to different morphemes in the verbal and the nominal paradigms. Since a feature can also appear more than once in a single paradigm, it is necessary to structure morphosyntactic features in a manner somewhat different from the usual (feature-matrix) solution in phonology (but see Labov 1981:299, Mascaró 1983). Following Anderson (1982:4.1) I will suppose that morphosyntactic features are arranged into trees.[5] A similar representation is used in GPSG (see e.g. ch 2 of Gazdar et al. 1984). The node labels in morphosyntactic trees will be the features themselves, with the additional constraint that

(4) only marked values can dominate other features.

This formal restriction is intended to capture an important aspect of the intuitive notion of markedness, namely, that in a multiple opposition, there can be but one basic (i.e. unmarked) member. Take, for instance, the possessive affixes in Hungarian. The affixes listed in the second column of (3) serve a dual purpose: on the one hand, they mark the affixed element as being the property of someone (the possessor), and on the other hand, they spell out the person and number of the possessor. Basically the same affixes, with an infixed *-i* can be used if the possession is plural. Nouns without possessive suffixes take the plural suffix *-k*. If we add a zero for the sake of completeness, we have a 14-way opposition: theoretically, this can be described with four binary features. But using such *ad hoc* features would make the rules of agreement extremely complicated: a more revealing description has to be based on the elementary oppositions expressed by the affixes.

First, we have to separate the person/number of the possessor from the number of the affixed noun (because in the rules of subject- predicate agreement, the former is irrelevant): this latter will be encoded in the feature ± PL. This leaves us with a 7-way opposition: the possessor can be 1st sg, ...,3rd pl, or there can be no possessor at all. It is this latter case which is truly unmarked: the ramaining cases, therefore, can be subordinated to a feature POS. Condition (4) makes it impossible

[4] There are a few cases, e.g. in the English verbal paradigm, where cross-linguistic considerations might justify a feature-analysis like *eats* = eat<- ME - YOU - PL>.

[5] Anderson actually employs unlabelled bracketings, but this is largely a notational difference.

to distinguish various -POS cases (which is just the desired effect), but enables us to use the standard person/number features ± ME, ± YOU, ± PL under +POS.

The situation is somewhat complicated by the 'familiar plural' affix *-ék* 'family, friends of', the prescence of which will be encoded by the feature + FAM. Since forms in *-ék* are always plural, it is possible to treat *-ék* as a special plural affix. As the unmarked member of the singular/plural opposition is the former, the feature ± PL has to be employed (rather than the feature ± SG), and this enables us to subordinate ± FAM to + PL.

The next to last column of (3) describes a different possessive suffix, one which behaves anaphorically (see e.g. Lotz 1967): the prescence of this suffix *-é* will be encoded by the feature + ANP. If the anaphoric referent is plural, we have the suffux *-éi* :this can be captured by subordinating the feature ± PL to +ANP. For lack of space, the case system will not be discussed here in detail; (oppositions like essive/lative are briefly mentioned in Section 3) end a separate feature is used for every surface case ending.[6] The unmarked member of the case system is the nominative: every other feature in the last column of (3) will be subordinated to an abstract feature + CASE.

This gives us the nominal paradigm in the form of the following tree:

(5)

<+POS<± ME ± YOU ± PL> +PL<± FAM> +ANP<± PL> +CAS<XYZ>>

The paradigms of numerals and adjectives differ only minimally from (3): numerals have no plural forms, while adjectives can also have (comparative *-bb* and superlative *leg-) degree* affixes in addition to the suffixes listed in (3). This means that if we apply the criterion (2) mechanically, those adjectives that happen to have defective paradigms in degree must be classified as nouns. From a purely formal point of view, adjectives in comparative form cause the same problem, since if we add the comparative suffix a second time, the resulting word, though phonologically well-formed, will not be acceptable in Hungarian. In this respect, the comparative suffix is typical: suffixes that can be iterated (see (1D) above) are in the minority.

Following Aronoff (1976), I will suppose that the presence of a non-iterative suffix *blocks* the application of the corresponding suffixation rule. In the case of comparative degree this means that word forms appearing as <+ COMP> on the surface behave as <- COMP> from the morphological point of view. This fact can be described without doubling the number of morphosyntactic features if we suppose that the + value of a feature corresponds to *potential* suffixation in morphology, but *actual* suffixation in lexical insertion. For example in (5) above, blocking will rule out every 'double plural' like **sógorokok, *sógoraimok, *sógorokék,* etc. Notice, that the relevant morphemes *-k, -i-, -ék* were all encoded bu the same feature (on the second branch of (5)), and we do not have the blocking effect for ± PLs on different branches.

In what follows, the actual vs. potential interpretations of a feature will be denoted by unbroken (resp. broken) lines in the trees formed by the morphosyntactic features. This 'colouring' of the edges is but a technically convenient solution: substantially the same effect can be achieved by the introduction of negative bar-levels (see Selkirk 1983:1.2) or with a new feature <± ACTUAL>.

The actual/potential distinction appears to be plausible from the point of language acquisition as well. Since the number of affixes is rather small but their frequency is extremely high, affixes are highly salient perceptually. If the hearer is able to recognize the affixes, he will be able to tell that

[6] This makes it necessary to introduce a feature coocurence restriction (much the same way as in GPSG, see e.g. Gazdar et. al. 1984 ch 2.3) that permits only one case feature to have positive value in any morphosyntactic tree.

an affixed stem belongs to a lexical category where the affix in question can be used (e.g. he will classify a form with some overt case-ending as a nominal), but the lack of affixation does not enable him to draw the opposite conclusion.

Once we have set up a system of morphosyntactic features in accordance with the basic morphological oppositions obtaining in the paradigm in question, the marked/unmarked distinction can be exploited in formulating the rules more concisely. For instance, certain features might be left unspecified, and fully specified representations might be concieved of as the end-product of certain marking conventions. Alternately, morphosyntactic trees with unspecified features might be taken to represent archi-elements. The solution adopted here is closer to the Prague Circle ideas (cf. e.g Trubetzkoy 1958 ch 3) than to the SPE theory of markedness (Chomsky - Halle 1968:9.2), because the (only) marking convention employed here is not sensitive to context, and always inserts the unmarked (-) value.

This convention operates only on the leaves of trees, and if a full subtree is left unspecified, this will represent a *natural class* of paradigmatic forms. In the notation, fiú<+ POS> is the abbreviated form of fiu<+ POS<- ME -YOU - PL> - PL - ANP - CAS> = *fiúja* 'his boy', while fiú+POS stands for the class fiúm, fiúd, fiúja, fiúnk, fiútok, fiújuk =fiú<+ POS<± ME ± YOU ± PL> - PL - ANP - CAS>. With this convention, only a few subsets of paradigmatic forms will be natural classes: for instance, the set 'my house, our house' = házam, házunk must be written as ház<+ POS<+ ME ± PL>>, and this class is no more natural than ház<+ POS<± ME +PL> = házunk, házaink 'my house, my houses'.

The principle of blocking can be applied in the description of defective elements as well: if these elements appear in the lexicon with *actual* morphosyntactic features, the application of the relevant affixation rule will be blocked. Thus, we can predict the defective paradigms to be natural classes in the sense outlined above: for example, the paradigmatic forms of defective nouns in Hungarian must correspond to the archi-words that can be defined with the aid of the morphosyntactic tree in (5).

Fortunately, this prediction can be tested on independently collocted data. In his book describing the Hungarian nominal paradigm, Papp (1975) devotes a full chapter (ch 5) to defective nouns. If we remove those elements from the corpus that were classified as 'saepe' by Papp (these are not truly defective in the strict sense of the term, see ibid, 187ff), we are left with some three hundred nouns having various defects. The majority of these (662 from the original 693 including 'saepe' words) shows some defect with respect to the possessive paradigm. One class (*fia* 'his son', *másа* 'his imagé,... p 206) contains elements that have to take some possessive suffix: these are lexicalized with the (actual) feature + POS. With certain compounds, we have obligatory infixation, rather than suffixation (*atyjafia* 'his fellow man', *hazámfia* 'my compatriot', *szavajárása* 'his manner of speaking',... loc cit), but this need not concern us here. Another class (*bá* 'old man, as in old man Harper..', *dádá* 'spanking', *spicc* 'tipsiness',... p 201) is truly defective: these are lexicalized with the feature - POS. Some of the - POS elements are simply possessive constructions that became lexicalized (*búcsúfia* 'souvenir, lit. son-of-fiestá, *napkelte* 'dawn, lit. waking-of-sun',... p 200) – with these, the blocking effect is quite transparent. The same is true for *pluralia tantum* (*naturáliák* 'allowances in kind', *üzelmek* 'immoral dealings',... p 199), since these are always lexicalized with plural morphology in Hungarian, and there are no nouns of the 'cattle, policé type. There is one word lexicalized with the familiar plural (*katonáék* 'the army'), and there are several elements with more than one lexicalized feature. For <- POS - PL> we have *énekbeszéd* 'recitation', *ófelnémet* 'old high german', ... (p 202); for <- POS + PL> *jelenvoltak* 'those who were present', *légutak*

'respiratory tracts',... (p 205); for <+ POS - PL> *napa* 'his mother in law', *holta* 'his death',... (p 207); and for <+ POS + PL> there is *elei* 'his forefathers' (p 209). Bot not every combination is attested: for instance, there are no nouns with lexicalized ANP feature, and there are no *singularia tantum.*

In general, the prediction that defective paradigms are natural classes of paradigmatic forms is borne out by the data: at least for nouns in Hungarian, every defective paradigm can be generated with the aid of lexicalized features that block further affixation.

3 X-bar features

The person/number features used in the description of the nominal paradigm are highly relevant in the verbal paradigm as well: in Hungarian, even the infinitival forms of a verb can indicate the person and number of the subject. This is not true for the anaphora possessiva (ANP) feature: Hungarian verbs do not take the suffix *-é*, or anything with a similar function. These facts have to be part of every description of Hungarian morphology, and in general, the grammar of any language must make it clear which affixes can, and which cannot be used with any given stem or word. If we accept definition (2) in Section 1 above, this means that categorial information need not be stored separately in the lexical entries, since the categorial status of an item can be inferred from its inflectional possibilities. Therefore ;.kp on (6) the morphosyntactic features employed in the description of paradigmatic forms can be used as X-bar features as well. ;.kp off For instance, we might say that certain stems are stored in the lexicon with the features - CASE, + TENSE, + MOOD etc. Since these must be verb stems, we need not stipulate that they also carry some *ad hoc* features like + Subj, + Obj (as in Jackendoff 1977 ch 3.2) or + V, - N (as in Chomsky 1970).

In addition to the resulting conceptual simplification of lexical entries, proposal (6) above provides a 'canonical' feature analysis for the category system of every language, and thus makes it possible to contrast these systems with each other. The situation is almost exactly the same as in phonology: while a feature analysis based on the surface phonemic contrasts of a given language is impossible to translate into other languages, it is (perhaps) possible to define a truly universal set of phonological features that can be applied with equal success to every language. The coherence between the phonological feature systems of various languages is due to their common phonetic basis; the coherence between the X-bar features (and the category systems) of natural languages might well be due to their common semantic basis.

Before describing the category system of Hungarian in terms of X-bar features, I will first list the major semantic categories which seem to be relevant in the inflectional morphology of natural languages.

1A Person (1st, 2nd, ...
1B Number (singular, dual, ...
2A Location (here, there, near, ...
2B Direction (to, from, ...
3A Gender (feminine, definite, animate, valuable, round-shaped, ...
3B Topic (familiar, known, ...
4A Tense (past, present, ...
4B Aspect (perfect, habitual, ...
5A Case (subject, object, ...

5B Voice (active, benefactive, ...
6A Degree (comparative, superlative, ...
6B Mood (interrogative, negative, ...

Although categories under the same number are usually intertwined, it seems that inclusive/exclusive should be classified with person, rather than number; honorifics with topic, rather than gender; possession with case, rather than voice; evidentials with voice, rather than case; etc. In general the situation is far more complicated than in phonology, and the elusiveness of these semantic categories might necessitate a language-by-language strategy for a long time to come. A further complication is that even those categories which are encoded by affixes in the language in question need not be inflectional: for instance, Hungarian has 'aspectual' suffixes like *-gat* (frequentative-repetitive) and 'modal' suffixes like *-hat* (conditional-permissive), but the former is derivational and the latter is inflectional.

In the category 'person' we have a 3-way opposition in Hungarian: this will be described by the supposedly universal features ± ME ± YOU. The facts of Hungarian are perhaps better described by a feature system in which ± ME is subordinated to the + value of the feature PARTICIPANT: another advantage of this system is that the feature coocurrence restriction *[+ME +YOU] can be dropped. But in order to facilitate cross-linguistical comparisons, I will stick to the feature system used this far. For number, we have the singular/plural opposition (encoded by the feature ± PL) – these categories are clearly inflectional in Hungarian.

The relevant opposition in the category 'location' is near/far: this will be encoded by the feature ± NEAR. Phonologically, this feature is spelled out by a form of vocal symbolism:

igy/úgy	'this way/that way'
itt/ott	'here/there'
ide/oda	'to here/to there'
innen/onnan	'from here/from there'
ilyen/olyan	'like this/like that'
ekkora/akkora	'this size/that size'
ennyi/annyi	'this much/that much'
ez/az	'this/that'

The phonological description of these forms must make use of a rule of *rounding harmony* (which can be independently motivated, see Kornai in prep.) that operates in tandem with the well-known backness harmony rules of Hungarian.[7]

In the category 'direction' we have a 3-way opposition to/at/from with 'at' as the unmarked member: this will be described by the feature ± OUT subordinated to the + value of the feature LATIVE.

(7A)	
alá/alatt/alól	'to under/under/from under'
elé/előtt/elől	'to the front of/in front of/ from the front of'
felé/—-/felől	'toward/—-/from'
fölé/fölött/fölűl	'to over/over/from over'

To a certain extent, this opposition is relevant for the case system,

[7] For a comparable situation in Montanes Spanish see McCarthy 1984

(7B)

-ba/-ban/-ból	'into/in/from in'
-ra/-on/-ról	'onto/on/from on'
-hoz/-nál/tól	'to/at/from'

and for certain adjectival pro-forms:

(7C)

ide/itt/innen	'to here/here/from here'
oda/ott/onnan	'to there/there/from there'
hova/hol/honnan	'where to/where/where from'

In the category 'gender' (topic is not inflectional in Hungarian), the relevant opposition is definite/indefinite, with the first member being the marked one. This will be encoded by the feature ± D. This feature, like gender features in general, is *silent* on nouns, i.e definitness can be an inherent property of the stem (cf. Anderson 1982: 1.3, Cooper 1983). Since verbs must agree with their objects in definitness, ± D is not silent in the verbal paradigm.

The most important silent features of verbs are the *government features:* these encode information about the cases a stem governs. In general, subcategorization features need not be different from other morphosyntactic features in their intrinsic content: the only difference is that they have to be subordinated to the + value of the feature GOVT.

Tense and mood are intertwined in Hungarian. Morphologically there is a 4-way opposition between the unmarked present (indicative); past (indicative); (present/future) imperative; and (present) conditional/permissive. Aspect and voice are not inflectional in Hungarian. The case system is rather complex: we might add terminative *-ig* to the directional/locative cases listed in (7B), and there are accusative *(-t),* dative *(-nak),* instrumental *(-val),* causative *(-ért),* transitive *(-vá),* and formal *(-ként)* cases as well. The features denoting them are given in the last column of (3) in Section 1. The degree system of Hungarian is open: from the superlative *leg-...-bb,* ultra-superlative forms can be derived with the prefix *leges-* which can be iterated, so e.g. *legeslegeslegeslegnagyobb* 'the very very very greatest'.

These features give rise to the following feature analysis for majoir

(8)

PERS	LOC	D	TENSE	CASE	DEG	GOVT	
−	−	−	−	−	−	−	Adverbial
+	−	+	−	+	−	−	Noun
+	−	−	−	+	+	−	Adjective
+	−	+	+	−	−	+	Verb
+	+	−	−	−	−	+	Postposition
+	−	−	−	−	−	+	Infinitive
−	−	−	+	−	−	+	Participle

In addition to these, there is a great number of minor lexical categories. For instance, members of the postpositional class *következtében* 'because of', *múlva* 'after',... have no directional or personal forms, and govern the nominative, while elements of an otherwise identical postpositional class (*képest* 'compared tó, *kifolyólag* 'because of',...) govern oblique cases.[8] There are several open

[8] This means that postpositions can not be classified in such a simple manner as suggested by Marácz (1983). If we also take syntactic criteria into account (e.g. whether a postposition can precede its complement noun or not), the classification will become even more complicated.

classes not listed in (8): numerals, for instance, take all the possessive suffixes, but have no number or definitness. (Under the marking convention, this means that they are singular and indefinite.) Defective words are often the only members in their class (see e.g. *katonáék, elei* in Section 2), but there are one-member classes that can hardly be called 'defectivé. For instance, the reciprocal pronoun *egymás* 'each other' is potentially + CASE + ANP, and actually + D - POS, and there are no other words with exactly these inflectional possibilities.

In Hungarian, as in almost every language, the list of minor categories can be extended almost indefinitely. In more detailed descriptions X-bar features are indispensable in keeping track of the category system. But if the feature system has no substantive connections with the lexical information which has to be stored with the individual items, minor categories can be accomodated only at the price of some *ad hoc* assignment of feature values. The system proposed here avoids this pitfall: every information (excluding, of course, semantic representation and phonological form) is stored in the lexicon in the form of morphosyntactic features, and the X-bar features of lexical categories are a (not necessarily proper) subset of these. The general conceptual schema proposed here is the same for every language, but as the morphosyntactic features are always based on the inflectional affixes of the language in question, the category system, and in particular the minor categories might vary from one language to the next.

References

Anderson, S. R. 1982: Where is morphology? Linguistic Inquiry 13, 571-612

Aronoff, M. 1976: Word formation in generative grammar. MIT Press, Camebridge, Mass.

Bloomfield, L. 1926: A set of postulates for the science of language. Language 2, 153-164

Bloomfield, L. 1933: Language. George Allen and Unwin, London

Chomsky, N. 1970: Remarks on nominalization. In: Jacobs-Rosenbaum (eds): Readings in English Transformational Grammar. Ginn, Waltham, Mass.184- 221

Chomsky, N.- M. Halle 1968: The Sound Pattern of English. Harper and Row, New York

Cooper, R. 1983: Swedish and the Head Feature Convention. Paper presented at the Workshop on Scandinavian Syntax and Theory of Grammar, Trondheim, June 3-5, 1982

Éltető L. 1985: Uj adatbáziskezelő rendszer VM/CMS alatt. Információ – Elektronika, in press

Gazdar, G., Klein, E., Pullum, G., and Sag, I. 1984: Generalized Phrase Structure Grammar. ms.

Harris, Z. 1951: Methods in Structural Linguistics. University of Chicago Press

Hockett, Ch. 1958: A course in modern linguistics. Macmillan, New York

Jackendoff, R. 1977: X-bar Syntax. MIT Press, Cambridge, Mass.

Jensen, J. and Strong-Jensen, M. 1984: Morphology is in the lexicon! Linguistic Inquiry.

Kean, M.l. 1976: Some features of features. GLOW-paper

Kiparsky, P. 1982: From cyclic phonology to lexical phonology. In: v.d.Hulst-Smith (eds): The Structure of Phonological Representation. Foris, Dordrecht 131-176

Labov, W. 1981: Resolving the Neogrammarian Controversy. Language 57, 267-308

Lieber, R. 1981: On the Organization of the Lexicon. IULC

Lotz, J. 1967: Egy nyelvtani modell (Két fejezet a magyar nyelvtanból. Magyar Nyelv 63, 394-408

Marácz L. 1983: Dressed or Naked: The Case of the PP in Hungarian. In: Linguistic Flowers from the Botanic Gardens. Studies Dedicated to Frank Heny. Groningen University Press.

Mascaró, J. 1983: Phonological levels and assimilatory processes. GLOW paper

Matthews, P.H. 1974: Morphology. Cambridge University Press

McCarthy, J. 1984: Theoretical consequences of Montanes vowel harmony. Linguistic Inquiry 15, 291-318.

Papp F. 1969: A Magyar Nyelv Szóvégmutató Szótára. Akadémiai

Papp F. 1975: A magyar főnév paradigmatikus rendszere. Akadémiai

Postal, P. 1969: Anaphoric Islands. Papers from the fifth regional meeting of the Chicago Linguistic Society 205-239

Selkirk, E. 1983: The Syntax of Words. MIT Press, Cambridge, Mass.

Trubetzkoy, N.S. 1958: Grundzüge der Phonologie. Vandenhoeck & Ruprecht, Göttingen