ON HUNGARIAN MORPHOLOGY

ANDRÁS KORNAI

Abstract

The aim of this study is to provide an autosegmental description of Hungarian morphology. Chapter 1 sketches the (meta)theoretical background and summarizes the main argument. In Chapter 2 phonological prerequisites to morphological analysis are discussed. Special attention is paid to Hungarian vowel harmony. In Chapter 3 a universal theory of lexical categories is proposed, and the category system of Hungarian is described within it. The final chapter presents a detailed description of nominal and verbal inflection in Hungarian, and describes the main features of a computer implementation based on the analyses provided here.

0 Preface

This thesis was written in 1984-1986 – the first publicly circulated version (Version 1.4) was defended at the Hungarian Academy of Sciences (HAS) Institute of Linguistics in September 1986. An extended Version 2 was submitted to the HAS Scientific Qualifications Committee in August 1988, and was formally defended in September 1989. Version 2.1, the 1994 paperback edition in *Linguistica: Studia et Dissertationes* (ISBN 963 8461 73 X) was different from Version 2 in two main respects: an English version of the official Summary Kandidátusi Értekezés Tézisei replaced the old Introduction, and the section on implementation reflected the system that was running at Xerox PARC at the time of the formal defense. The current Version 3.0 (pdf only) is based on the source of the paper version with minimal reformatting (the addition of hyperrefs, and a slight revision of this preface – salient changes *italicized*).

To leave the material in the 1986 thesis in manuscript format until 1994 would have required more patience than the author can lay claim to. The abstract analysis of feature systems in chapter 2.1 forms the basis of the more detailed analysis in Kornai 1993. The analyses of Hungarian vowel harmony in 2.1 and 2.6 have been published in Kornai 1987 and Kornai 1991. Part of the material on syllable structure in 2.4 can now be found in Kornai 1990, and a more extended version of the material in 2.5 on postlexical phonology is in Kornai and Kálmán 1989. The analysis of lexical categories in 3.1 and 3.2 has been published as Kornai 1985. Except for chapter 4, which describes the morphological rules in detail and discusses a two-level implementation, almost all the material has been published elsewhere *by 1994*. What was the purpose of the paper edition in 1994 and what is the purpose of the present (2007) pdf version?

In 1994 it was hoped that by bringing the parts together in a convenient monographic format the reader could use the volume as a reference work detailing nearly all aspects of the inflectional and many aspects of the derivational morphology of Hungarian. *In this regard, the book has been clearly superseded by* Siptár and Törkenczy 2000. There are a number of places where references to this volume and to later developments could be incorporated, especially in the treatment of underspecification (see Steriade 1995), multiple specification (see Lieber 1987) and feature geometry (see Clements and Hume 1995). In general, I restricted myself to updating the references by citing the published version of works that I originally consulted in manuscript form.[1]

Since the rule system developed for the thesis has been extensively tested on over twenty thousand stems covering virtually the whole lexicon of contemporary Hungarian, it was also hoped that future models of Hungarian morphology can, if nothing else, learn from the mistakes of the 1989 model presented in the 1994 edition. *This hope was certainly fulfilled:* the underlying database has become part of the bloodstream of research in Hungarian phonology and morphology, providing one of the three main sources for the current morphdb.hu, the other two being the Dictionary of Hungarian Inflections by

[1] In this process I had to trade in philological accuracy for usability: the sections quoted from the mss may no longer be present verbatim in the published works, but the latter are generally very close to the former in spirit, and are, unlike the mss, easily accessible to the reader.

László Elekfi and László Németh's Hungarian ispell dictionary. Since a fair amount of the material here serves as motivation for later developments such as the annotation scheme used in the morphdb database, linguistically oriented users of this resource may still find Version 3.0 a relatively painless introduction to the subject.

It was also hoped that the overarching plan of this work, as presented in chapter 1, might still have some relevance for theoretical linguistics. In a fast moving field like generative grammar authors seldom get a chance to contemplate whether the topics they chose to investigate and the methods they applied a decade *(now two decades)* ago would still make sense. In the eighties the author was lucky enough to pick a problem domain, morphology, that was just about to come in from the cold, and a technical framework, autosegmental theory, that has since grown from an esoteric branch of Africanist phonology into a mainstay of generative grammar. In the nineties some of the fundamental judgement calls made in this work, in particular the unashamedly procedural mode of description and the unbridled minimalism/reductionism that are the common threads binding it all together, came under renewed attack. Whether the work stands up against these lines of criticism was supposed to be decided by the 1994 reader. *In 2007 the author feels free to draw his own conclusions:* proceduralism is dead but rule systems are alive and well. Eighties phonology was not particularly concerned with the psychological reality of intermediate stages of derivations, unless they were the output of a specific lexical level, and contemporary phonology shows the same indifference to Gen candidates thrown out by Eval, unless they matter for sympathy. (Actually, the two issues as strongly related, see McCarthy 2000, Kiparsky 2007.)

0.1 Textual conventions

Examples, rules, figures, and formulas are numbered consecutively in each chapter: for instance, (2:18) refers to item (18) in chapter 2. Given the availability of hyperref, the footnote/endnote distinction of Version 2.1 is no longer maintained. When there is no danger of misunderstanding, Hungarian forms are given orthographically. Glosses appear in single quotes, and phonemic transcriptions (in cases where the spelling is insufficient) in slashes. In a few cases, phonetic transcriptions are given in square brackets. Angled brackets are used with features exclusively: tree structures are often given as bracketings.

0.2 Acknowledgments

This work owes its existence to a grant of the Hungarian Academy of Sciences (HAS) that enabled me to study morphology at the HAS Institute of Linguistics (NYTI) while on leave from the HAS Computer Science and Automation Institute (SZTAKI). I would like to thank Előd Knuth†, then head of the Information Systems Department, and Tibor Vámos, then director of SZTAKI, for their generosity in providing access to the computer facilities of SZTAKI — this work would have been impossible without their support. I am deeply indebted to many linguists for their suggestions, comments, and criticisms over the years. My former colleagues at the NYTI, and in particular the erstwhile collective of Room 13, are to be thanked *in toto*.

Improvements from Version 1 to 1.4 were due largely to László Kálmán (NYTI), Ádám Nádasdy (Eötvös Loránd University), and Péter Siptár (NYTI). Version 2 benefited a great deal from the opinions of my opponents at the 1986 defense, Ferenc Papp†, and Péter Siptár (NYTI) as well as from the comments and criticisms submitted in writing or made in person at the defense by József Hermann†, István Kenesei (University of Szeged), Ádám Nádasdy (Eötvös Loránd University), and Péter Siptár (NYTI), who went over so many versions that mentioning his name a third time is still less than adequate.

Version 2.1 also benefited from the opinions of my opponents at the formal defense, Ferenc Papp, and György Szépe (NYTI), as well as the comments and criticisms of those on the committee: László Elekfi (NYTI), György Hell (Budapest Institute of Technology), László Hunyadi (University of Debrecen), István Kenesei (University of Szeged), Katalin É. Kiss (NYTI), Mihály Péter (Eötvös Loránd University), and László Pordány (University of Szeged). Kari Swingle (UCSC) is to blame for the errors that still remain.

Extending the system from 200 to 20,000 stems would have been impossible without the SZOTA1R database. I am indebted to many people who contributed to SZOTA1R, especially to the late Ferenc Papp, whose *Debrecen Thesaurus* provided the majority of stems in SZOTA1R, and Mihály Füredi, whose *Frequency Dictionary* was extensively used in testing and refining KIMMO1. I would like to thank Lauri Karttunen, whose TWOL system provided the basis of the 1989 implementation. Without his patient instruction and untiring support, I would still be working on Hungarian morphology with paper and pencil.

The 1985 lectures of Nick Clements on syllable structure at the Salzburg International Summer School have had a decisive influence on my treatment of the material in 2.4. It is a pleasure to acknowledge my indebtedness. The ideas contained in Appendix 2.6 and in particular the treatment of ternary harmony are the result of joint work with Donca Steriade (MIT) and Harry van der Hulst (University of Amsterdam). Errors and omissions are my own.

Finally, I would like to thank my advisor, Ferenc Kiefer, for his constant help and encouragement.

1 Introduction

Aside from a few isolated attempts, such as Kiefer 1970, morphology played a very limited role in the early development of generative grammar. The basic reason for this was that the standard generative model (Chomsky 1965) treated sentences as strings of morphemes: both base (rewriting) and transformational rules operated on morphemes. This one-step model (originating in the work of Harris 1946, 1951) was gradually replaced by a two-step model in which sentences are treated as strings of words, and words are treated as strings of morphemes, much as in traditional grammar.

In order to (re)introduce 'word' as an explanatory category, the class of possible models had to be delimited so that the division of labor between rules of syntax, on the one hand, and rules of phonology/morphology on the other, become clear. This is accomplished by the Lexicalist Hypothesis (Chomsky 1970), now usually called the Lexical Integrity Hypothesis (LIH). Although the LIH exists in many versions (see Scalise 1985), for our purposes it will be sufficient to state the following basic requirements, which are common to nearly all versions of the LIH:

(1) Rules of syntax (and semantics) cannot make reference to the phonological content of words.

(2) Rules of syntax (and semantics) cannot modify the phonological and morphosyntactic features of words.

Thus, in addition to forbidding rules like Affix Hopping (Chomsky 1957), the LIH also forbids the derivation of word-forms by syntactic rules. But if syntax can not derive word-forms, each and every word-form must be supplied by the lexicon. According to the traditional view of the lexicon as a list, this would mean that every (paradigmatic) form of a word must be listed. Given that such forms often number in the thousands, listing them all appears to be impractical, if not impossible.

But the traditional file-card based technology of lexicography has gradually been replaced by a computer-based technology that can handle several orders of magnitude more data, and the reason why generative morphology avoids listing all word-forms is not a practical but rather a theoretical one. It is the *Principle of Brevity*, stated by Chomsky and Halle (1968:12) as follows:

> "Regular variations (...) are not matters for the lexicon, which should contain only idiosyncratic properties of items, properties not predictable by general rule."

Listing all word-forms (and in particular, all paradigmatic forms) thus contradicts the Principle of Brevity, while the LIH, apparently, requires exactly this.

Generative morphology resolves this contradiction by treating the lexicon not as a static list but as a dynamic (generative) component of the grammar. The LIH requires only that syntax must get fully formed words from the lexicon, but does not require that such forms be stored in the lexicon. On the contrary, the Principle of Brevity demands that no form that can be produced by a regular operation should be stored. Thus we must distinguish the output and the content of the lexicon. The output of the lexicon is the set of well-formed words. If we treat compounding or recursive derivational processes as productive, this will be an infinite set, which makes it impossible to think of the lexicon as a list. The content of the lexicon, however, will be reduced to the list of irreducible elements (morphemes) and the rules operating on them (suppletive and other irregular forms are also included here).

Therefore, the fundamental goal of generative morphology is to characterize the basic elements and operations of the lexicon — the contributions of the present dissertation to this goal are discussed in 1.2 – 1.4.

1.1 The methods of the investigation

We have seen above that the LIH and the Principle of Brevity, taken together, will naturally lead to a generative view of the lexicon. Given that traditional grammar makes as sharp a division between phonology and morphology as is made by the LIH between morphology and syntax, it seems possible to gain a better understanding of the structure of the lexical component by enforcing more principles of separation than just the LIH. This is the method of 'natural' generative phonology, where the most important principle of separation is the Morphophonemic-Allophonic Principle introduced by Koutsoudas et al. 1973. More recently, research in 'natural' morphology (Dressler 1985) attempts to isolate a third, morphophonological component between phonology and morphology.

The basic method of the present investigation is exactly the opposite of the strategy of natural phonology/morphology. My fundamental assumption is that phonology and morphology form an indivisible unit which I will call 'the lexicon' or just 'morphology'. This assumption seems to be contradicted not only by the obvious difference in the size of the basic units (phonemes vs. morphemes) but also by the different nature of the typical phonological operations (e.g. assimilation or deletion) and the typical morphological operations (e.g. affixation or compounding). Therefore it is necessary to discuss briefly why 'mainstream' generative grammar treats phonology and morphology homogeneously.

Jakobson pointed out that we find a number of rules (such as the rule of word-final devoicing in Russian, see Halle 1959) that are morphophonemic and allophonic at the same time. With the introduction of context-sensitive rules (Chomsky 1956), the homogeneous formal treatment of phonological and morphological rules became possible and necessary, since there was no separate morphological component at the time. The principle of cyclic rule application made clear that each morphological rule (affixation) triggers phonological rules such as stress shift. Furthermore, Lexical Phonology (Kiparsky 1982) provided a unified treatment of cyclic rules and the phonological processes taking place at morpheme boundaries and showed that the rules of phonology and morphology are arranged in the same stratal structure.

The most important argument in favor of a homogeneous treatment of phonology and morphology was provided by the fundamental transformation of phonology in the last decade. This transformation started with the introduction of a separate tier for tone (Leben 1973, Goldsmith 1976). This was followed by tiers for harmonizing vowel features (Clements 1976), for aspiration (Thráinsson 1978), for nasalization (Hyman 1982), for syllabicity (Clements and Keyser 1983) and so on. The multi-tiered representations thus formed made it possible to treat the infixing morphology of Semitic languages by purely phonological means (McCarthy 1979) and to reduce reduplication to concatenative affixation

(Marantz 1982). With the aid of multi-tiered representations, other processes, hitherto assumed to be purely morphological, also became amenable to a treatment in terms of phonologically motivated operations.

The methodological basis of the present dissertation is the principle of parsimony (Occam's razor). Thus, when we seek a characterization of the basic elements and operations of the lexicon, we seek an answer to the following questions. What are the fundamental units that *must* be stored in the lexicon? What are the operations which are *indispensable* for the task of generating every word-form? Starting from the phonological form of words we can argue that the words can be decomposed into syllables, the syllables can be decomposed into phonemes, and the phonemes can be decomposed into distinctive features. The distinctive features are atomic, and thus will necessarily be part of the lexicon. Starting from the meaning of words, the minimal units having both phonological and semantic content are (by definition) the morphemes, so the lexicon will have to contain these too.

However, the question whether morphemes are built from phonemes or directly from features is arguably open: for instance, in Hungarian, backness is a property not of the individual vowels, but of the whole morpheme (see Hetzron 1972). Autosegmental phonology expresses this fact by locating backness on a separate tier. For instance, the segmental content of *apa* 'father' and *epe* 'bile' will be *ApA* in both cases, where *A* is the same *a/e* archiphoneme that we find in the dative suffix *nak/nek*. In the full representation of these morphemes, both the segmental and the backness tiers are present, together with the association lines between them:

```
  +B            -B
  / \           / \
  ApA           ApA          nAk

 (apa)         (epe)        (DAT)
```

As can be seen, the dative morpheme is not associated to the +B or –B feature — the basic rule of vowel harmony says that the *A* of *nAk* has to be associated to the backness feature of the stem. Therefore association (and also delinking) will have to be listed among the fundamental operations, the more so because they appear in the description of other phonological processes (such as assimilation or compensatory lengthening) as well. Similarly, a theory of the lexicon can not do without the operations of feature insertion and deletion. The use of these operations and of concatenation has been extensively justified in autosegmental phonology already in the description of tone languages, so the present dissertation could take these to be given.

For the reasons sketched above, I could not take it for granted that phonemes are primitives. With the introduction of a root tier, autosegmental phonology made it possible to identify a phoneme with the set of features associated to a root node, and to treat the (traditionally problematic) affricates and diphthongs as well as the long (geminate) phonemes in a homogeneous manner together with short phonemes (Clements and Keyser 1983). Accordingly, I did not permit rules that operate on phonemes rather than directly on features. With the introduction of a morphemic tier (McCarthy 1979), parentheses, and in general segmentoid boundary markers also become eliminable (for the syllable boundary see Kahn 1976, for boundary markers of various strength see Mohanan 1984).

The central aim of the dissertation is to show that Hungarian morphology can be described with the extremely limited inventory of representations and operations outlined so far. Since the most powerful tools of the standard theory (such as transformations, curly brackets, and Greek letter variables) were already eliminated from autosegmental theory, the dissertation concentrates on the remaining two strongest tools, namely diacritic features and feature changing rules. Since my aim was to limit the number of tools available, the choice between alternative analyses was always dictated by the criterion of using the least number of *ad hoc* tools.

The method employed in the investigation of the meanings of words and morphemes was also dictated by considerations of parsimony. Although we do not know precisely what features distinguish the meaning of one morpheme (or word) from the meaning of another one (or, to put it differently, our decisions in these matters will greatly depend on the theory of lexical semantics we adopt), we can assume without further argumentation that the number of the ultimate features is finite (less than the number of elements that have to be listed in the lexicon). Since some operations that will insert features in larger structures (and further transform these structures) will be necessary for phonological purposes anyway, the dissertation employs the same operations in the investigation of meaning as well. The morphosyntactic features that transmit the information between syntax and morphology are also handled by these operations, thereby steering clear of the problem whether morphosyntactic features are to be treated as morphological (Kiparsky 1986), syntactic (Gazdar *et al* 1985), or semantic (Lapointe 1980).

1.2 Summary of new results

2.1 develops an algebraic treatment of phonological features which, through the investigation of the relation between phonological features and natural classes, provides a unified treatment of the original (Pāṇinian), the standard (SPE), and the modern (autosegmental) approaches. Here I will omit the algebraic details and provide a somewhat simplified model that employs only concepts from elementary set theory.

Let $P = \{p_1, p_2, ..., p_k\}$ be a set of phonemes and $F = \{f_1, f_2, ..., f_n\}$ be a set of binary features. The mapping C:$P \to 2^F$ will be a *feature analysis* if it satisfies the following criteria:

(3) Emicity. If $i \neq j$, then $C(p_i) \neq C(p_j)$.

(4) Compactness. If $N \subset P$ is a natural class, we can find a set of forbidden features T' and a set of required features B such that $p \in N$ iff $B \subset C(p) \subset F - T'$

In the dissertation I argue that features and phonemes can be treated uniformly in the algebraic model, and show that the standard theory of features makes two predictions concerning the set of natural classes:

(5) The number of natural classes is a (small) polynomial function of the number of phonemes ($k^{1.585}$)

(6) The set of natural classes is closed under intersection

I show that the Pāṇinian theory makes essentially the same predictions (with the function k^2 instead of $k^{1.585}$), and I develop a concept of 'natural class' for the autosegmental case that leaves (5) and (6) in force. The theory is illustrated on a feature analysis of Hungarian vowels based on the tridirectional features $\langle I, U, A \rangle$ and the model is interpreted with the aid of barycentric coordinates so that it relates phonological features to phonetic facts.

2.2 deals with the feature analysis of consonants by reformulating the traditional analysis in terms of the feature geometry suggested by Clements (1985). The investigation of consonant epenthesis undertaken here provides a new argument in favor of a proposal made by É. Kiss and Papp (1984) that in Hungarian *dz* should be treated as a sequence of two phonemes.

2.4 deals with syllable structure. Earlier results (chiefly Siptár 1979,1980) are discussed from the point of sonority, and the ID/LP model of constituency (Gazdar and Pullum 1982) is used to exclude final clusters such as *pj, kj* that appear only in inflected forms.

2.5 deals with postlexical rules. In addition to an outline of Hungarian consonant sandhi, the rules of Hungarian sentence intonation, jointly developed with László Kálmán, are also sketched.

1.3 Vowel harmony

The investigation of Hungarian vowel harmony traditionally concentrates on the binary alternations *a/e, á/é, o/ö, ó/ő, u/ü, ú/ű*, and mentions the ternary alternation *o/e/ö* only in passing. An important empirical result of the present dissertation is that it includes the quaternary alternation *a/e/o/ö* which, following Vágó 1975, is treated in the literature as if it were independent of the problem of vowel harmony. As we shall see, a unified treatment of binary, ternary, and quaternary alternation is justified not only because in the surface representation the alternant *a* (as in *házat, házak, házam*) plays the same role as the alternants *o,e,ö*, but also because this way suffix-combinations can be described by the same machinery used for single suffixes.

A fundamental difference between the earlier analyses and the one proposed here is that here *a* and *á*, as well as *e* and *é*, are treated as differing only in length. In order to show that there are no differences between *a* and *á* or *e* and *é* in (underlying) height or roundedness, I analyze two quantity-changing processes, Low Vowel Lengthening (*apa/apát, epe/epét*) and Stem Shortening (*tél/telet, nyár/nyarat, tűz/tüzet, víz/vizet, nyúl/nyulat*).

For the description of a seven vowel system at least three features must be used. A feature analysis using only three features can be reconstructed from the phonological pattern of Hungarian: the feature A separates the vowels of the quaternary archiphoneme from the rest, the feature I separates the two alternants of binary alternations, and the feature U will group the vowels according to ternary alternation.

If we also take the proximal/distal alternations *itt/ott, ez/az, így/úgy* into account, we arrive at a feature chart that is equivalent to a Jakobsonian analysis: I = ⟨–grave⟩, A = ⟨–diffuse⟩, U = ⟨+flat⟩.

The next step in the analysis is to show that the traditional palatal/velar as well as the more modern front/neutral/back stem classifications are not sufficiently detailed and that in fact we need five stem classes. The stems *had, bab, hit, hölgy,* and *tök* all have to be in different classes, since no two of these will get the same alternants for every (binary, ternary, and quaternary) suffix (cf *hadat, babot, hitet, tököt*, and *hithez, hölgyhöz*). Stems of the *had* and *hölgy* type (which are treated in Vágó (1980) with the aid of a 'Minor Lowering' rule triggered by a diacritic ML) require separate classes not only because they are numerous, but also because this treatment permits all stems (including derived ones) to be unambiguously classified. Although from a diachronic point of view, these classes are closed, in a synchronic description they must be treated as productive, since stems bearing plural (or possessive) suffixes all belong here (cf **babokot, *babomot, *tökököt* etc.).

The simplest possible harmony rules are:

$$\begin{array}{ccc} I & & \\ | & \ddots & \\ V & C_0 & V \end{array} \qquad\qquad \begin{array}{ccc} U & & \\ | & \ddots & \\ V & C_0 & V \end{array}$$

How far can an analysis employing these rules be pushed? It would go against the Principle of Brevity to leave the stems selecting the *a*-alternant unmarked and mark the stems selecting the *o* alternant with a diacritic — we must suppose that for back stems the unmarked member of the opposition *a/e/o/ö* is *o*. Therefore the diacritic ML turns [+A,+U] into [+A] (in the *had* class) and turns [+A,+U,+I] into [+A,+I] (in the *hölgy* class). It follows from our basic aim of eliminating diacritics that the *ad hoc* ML has to be replaced by the phonologically motivated ⟨–U⟩.

Given the existence of such ML stems as *lyuk*, the only way to reconcile the +U feature of the stem vowel and the –U required by the exceptional harmonic behavior of the stem is to specify the same feature on different tiers. In other words, the parsimonious treatment of features results in the introduction of an extra tool such as the *core specification* (in the sense of Halle and Vergnaud 1983) that is used in the analysis given in 2.3. Thus, it seemed advisable to provide an alternative analysis based on four vowel quality features. This analysis, given in 2.7, is based on the standard features:

	a	e	i	o	u	ü	ö	a/e	o/ö	u/ü	o/e/ö	a/e/o/ö
back	+	-	-	+	+	-	-	0	0	0	0	0
high	-	-	+	-	+	+	-	-	-	+	-	-
low	+	+	-	-	-	-	-	+	-	-	-	0
round	0	0	-	+	+	+	+	0	+	+	0	0

The role of the diacritic ML is played by a (floating) ⟨+low⟩. It can be seen that the spreading of ⟨+low⟩ will narrow down the quaternary archiphoneme *a/e/o/ö* to the binary archiphoneme *a/e*. The rules of (privative) I-spread are replaced by a pair of (equipollent) spreading rules:

$$\begin{array}{ccc} B & & \\ | & \ddots & \\ V & C_0 & V \end{array} \qquad\qquad \begin{array}{ccc} F & & \\ | & \ddots & \\ V & C_0 & V \end{array}$$

and there is a separate rule for ternary harmony:

$$\begin{array}{llll} \text{backness tier} & F & & \\ & | & \ddots & \\ \text{CV tier} & V & C_0 & V\langle -H \rangle \\ & | & \iddots & \\ \text{rounding tier} & R & & \end{array}$$

While the analysis based on three features had to use feature changing rules, this analysis employs only monotonic (feature adding) rules. As a result, the spreading of ⟨+low⟩ onto the ternary archiphoneme is blocked because *o/e/ö* is underlyingly specified as ⟨–low⟩. Since the standard analysis does not distinguish between the ternary and the quaternary archiphoneme, the Minor Lowering rule of Vágó 1980 would generate the incorrect form **hölgyhez* in such cases.

1.4 Summary of conclusions

Chapter 3 discusses the notions *word* and *lexical category*. The explanatory value of the notion *word* is argued to stem from the fact that several logically independent methods of segmentation yield essentially the same word-sized units. Lexical categories are defined morphologically which has the advantage of yielding a feature analysis (X-bar theory) of lexical categories as a byproduct of morphological analysis. The (morphosyntactic) features defining lexical categories are argued to have their own 'geometry'. The basic restriction on the tree structures thus formed is that only "+" (marked) nodes can have daughters. The theory is illustrated on the category system of Hungarian, with special emphasis on the problem of defective paradigms.

4.1 and 4.2 provides a detailed description of the Hungarian verbal and nominal paradigms. The diacritic ML was eliminated in Chapter 2 — here two other diacritics, governing Stem Shortening (as in *nyár/nyarat*) and Vowel Drop (as in *cukor/cukrot*) are eliminated in favor of more motivated features. In order to describe the 52 verbal forms discussed, we need 26 morphemes (ordered by the Elsewhere Principle). This is more than twice as many as the 12 morphemes that would be necessary in a fully agglutinative system, but only half of the 52 that would be needed in a purely inflecting system. Roughly the same degree of agglutination is found in the nominal (possessive) paradigm.

In sum, the investigation of diacritics in Hungarian arguably supports the thesis that it is often the arrangement, and not the substance of the atomic units that gives rise to different behaviors, much like in the case of graphite and diamond. The investigation of feature-changing rules, however, does not seem to yield the same kind of 'conservation laws' that are common in the physical sciences. Although Hungarian vowel harmony is amenable to a feature-adding treatment, there are places both in the verbal paradigm (such as the 3rd singular present definite *ja/i*) and in the nominal paradigm (such as the plural possessive *i*) that require the use of feature-changing rules under any analysis of vowel harmony.

2 Phonology

The aim of this chapter is twofold: on the one hand, I will discuss some phonological preliminaries to the morphological analysis developed in subsequent chapters, such as the feature analysis of vowels (2.1), consonants (2.2), and a treatment of vowel harmony (2.3). On the other hand, I will show that the proposed framework gives us insights that go well beyond what is strictly necessary for morphological analysis. To that end, I will outline a treatment of some phonotactic constraints (2.4), and a brief discussion of postlexical phonology (2.5).

Given the preparatory nature of this chapter, it did not seem necessary to pursue all these developments to the same level of descriptive detail. Vowel harmony, being probably the most pervasive, and certainly the most researched, feature of Hungarian phonology, receives a detailed description and not one but two analyses are provided for it.[2] Consonantal rules are discussed only in order to exemplify the basic autosegmental mechanisms one would want to employ in a truly detailed description: on a number of occasions the details are left as an exercise to the reader in the hope that he will be interested in 'executing' the 'leading ideas' of autosegmental phonology on the wealth of material provided by Hungarian.

2.1 The feature system: vowels

The feature analysis of Hungarian is by no means an uncontroversial question. The solution offered here differs from more or less traditional proposals not only in the choice of underlying feature set, but also in the manner in which the feature composition of a given phoneme is interpreted. Since the proposal to be made here departs from the conventional *linear* mode of representation, as exemplified by SPE (Chomsky and Halle 1968:8.2) and exploits the possibilities of the *autosegmental* approach to a considerable degree, it is extremely hard to 'linearize' it in order to facilitate comparison with more standard solutions.

Therefore, it will be necessary to give a formal rendering of the 'feature analysis' problem in well-defined mathematical terms: this way it will be possible to expose the metatheory of features. The less mathematically minded reader should not worry, for the formalization will be a light one: the algebra and the geometry employed here is conveniently summarized in chapters 9 and 12 of Birkhoff and Bartee 1970, where the definitions of vectors, fields, linear spaces, and lattices are given in detail; and chapter 13 of Coxeter 1961, where barycentric coordinates are defined rigorously, or Steinhaus 1960:119, where they are treated more intuitively. (In fact, the mathematical rigor of these concepts is not essential in what comes later.) Moreover, each step will be illustrated on a concrete example, that of the Hungarian vowel system. The consonant system is discussed in 2.2.

In general, there is very little disagreement concerning the phoneme inventory — for Hungarian, it is almost universally accepted that the (surface) phonemes are the following:

[2]Since the second of these is not wholly the author's intellectual property, it is relegated to an Appendix to chapter 2.

(1A)

p	b			t	d	ty [c]	gy [ɟ]			k	g	
		f	v	sz[s]	z			s [š]	zs [ž]			
				c[t^s]	dz [d^s]			cs[č]	dzs [ǰ]			
	m				n		ny[ɲ]					
					r l							
						j [y]						h

(1B)	a [ɔ]	e [ɛ]	i	o	u	ü	ö
	á [a:]	é [e:]	í[i:]	ó [o:]	ú [u:]	ű [ű:]	ő [ö:]

This gives us a 14-member vowel system. Phonetically, there is little discernible difference between the quality of *i* and *í*; *o* and *ó*; *u* and *ú*; *ü* and *ű*; or *ö* and *ő*. However, the long vowel *á* is low, central, and unrounded, while the short vowel *a* is mid-low, back, and rounded. The long vowel *é* and the short vowel *e* do not differ in backness or rounding (they are both front unrounded), but they differ in height: *é* is mid-high and *e* is mid-low. On the basis of these differences it has been suggested (e.g. in Abondolo 1988:1.1.9.2) that 'marginal' phonemes [a] (and [e]) also belong to the vowel inventory. Since the minimal pairs

vért/vért	[vert/ve:rt]	'blood-ACC/armor'
párt/párt	[part/pa:rt]	'pair-ACC/party'
Móczárt/Mozart	[mot^sa:rt/mot^sart]	'Móczár-ACC/Mozart'

suggested by Endre Tálos (pc) can*not* be told apart by native speakers, I will use the traditional 'systematic' inventory, in spite of the fact that certain other minimal pairs (based on the short variants) e.g

arra/ara	[ɔ:rɔ/ɔrɔ]	'that way/bride'
erre/ere	[ɛ:rɛ/ɛrɛ]	'this way/his vein'

as suggested by Ádám Nádasdy (pc), apparently can be. The problem of marginal vowels is clearly linked to the existence of phonemic *ë* [e] in a number of dialects.[3]

The nearly unequivocal position of grammarians with respect to the *cardinality* of the set of vowels and their phonetic properties is in sharp contrast to the diversity of the feature analyses they propose. There is no agreement concerning the *names* of the individual features (e.g. ⟨grave⟩ vs. ⟨back⟩), the *arity* of certain features (e.g. the ones for vowel height), the *values* a feature has for a given vowel, the *number* of features to be used, etc. For the reader's convenience, let me tabulate here the two main proposals (Vágó 1980, Becker-Makkai 1970):

[3]The 14-member vowel system is taken for granted in nearly every grammar dealing with the standard (Budapest) dialect of Hungarian: the list includes Lotz 1939, Hall 1944, Majtinskaja 1955, Tompa (ed) 1961, Abondolo 1988. The most important exception is Szépe 1969, which argues for the recognition of underlying *ë* in the standard dialect as well.

(2)	a	á	e	é	i	í	o	ó	u	ú	ü	ű	ö	ő
back	+	+	-	-	-	-	+	+	+	+	-	-	-	-
high	-	-	-	-	+	+	-	-	+	+	+	+	-	-
low	+	+	+	-	-	-	-	-	-	-	-	-	-	-
round	+	-	-	-	-	-	+	+	+	+	+	+	+	+
long	-	+	-	+	-	+	-	+	-	+	-	+	-	+

(3)	a	á	e	é	i	i	o	ó	u	ú	ü	ű	ö	ő
diffuse	-	-	-	-	+	+	-	-	+	+	+	+	-	-
flat	-	-	-	-	-	-	+	+	+	+	+	+	+	+
tense	-	+	-	+	-	+	-	+	-	+	-	+	-	+
grave	+	+	-	-	-	-	+	+	+	+	-	-	-	-

Originally, (2) was suggested in Szépe 1969:399 as part of the first attempt at developing a generative phonology of several Hungarian dialects. Vágó 1975 introduced the idea of taking *a* as underlyingly ⟨–round⟩, and started to use SPE feature names. In the subsequent literature, (2) has been adopted almost universally without any attempt to justify it. The emergence of autosegmental phonology did not change this situation: the ‘standard’ system was retained (with a separate tier for front/back) essentially in the same form as in (2) until quite recently, when the interest in feature geometry and underspecification called for a re-evaluation of feature systems.

The most important attempt at developing a different feature system is that of Becker-Makkai 1970:639, who presents her system, given in (3), in Jakobsonian terms. Unfortunately, this paper went largely unnoticed, in spite of the fact that it contains some of the quaternary data ignored in mainstream treatments. In the original system, the ⟨grave⟩ values for *i* and *í* are underspecified[4], and I have added the value ‘–’ in (3) in order to make it comparable to the fully specified SPE system in (2).

In spite of the different feature names, the two systems are not dissimilar: ⟨tense⟩ is the same as ⟨long⟩, ⟨diffuse⟩ is the same as ⟨high⟩, and ⟨grave⟩ is the same as ⟨back⟩ (cf SPE 4.2.1). Moreover, although Becker-Makkai notes that the ⟨–flat⟩ specification for *a* “could be disputed on articulatory grounds”, she opts for an abstract solution which makes her ⟨flat⟩ equivalent to Vágó’s underlying ⟨round⟩. The main difference between the two systems is in the treatment of short and long *a* and *e*. In (3), the only difference is in tenseness (=length), while in (2) there are other differences as well.

Working in the framework of SPE, which permits an underlying as well as a surface inventory, Vágó did not have to make a decision that flies in the face of phonetic facts in a case such as the roundness of *a*. Nevertheless, there are distinctions in Vágó’s analysis which are completely arbitrary from a phonetic point of view. Nádasdy (1985) describes the phonetic height difference between *a* and *á* as mid-low vs. low: Vágó treats both as ⟨+low⟩. Nádasdy describes the phonetic height difference between *e* and *é* is mid-low vs. mid-high: Vágó treats *e* as ⟨+low⟩ but *é* as ⟨–low⟩.

[4] “The two blank spaces represent redundancies which do not need to be marked”

But if phonetic facts cannot serve as an infallible guide in establishing a feature analysis, what is the motivation for the systems given in (2) and (3) above? What does it mean to assign a '+' or a '–' as the value of some feature for a phoneme? In the remainder of this section I will discuss this question from a general perspective — the phonological argumentation in favor of (3) will be deferred until 2.3.

A feature analysis has essentially two aims: one is to encode the phonemes with the aid of the features, and the other is to embed this analysis in the universal theory of features. In the case at hand, the aim is to choose certain features from a finite and fixed set, and describe the vowels of Hungarian in terms of these in such a manner that the following conditions are met:

(4A) Different phonemes have different feature matrices (Emicity),

(4B) Natural classes of phonemes are expressed by the combinations of these features (Compactness).

In addition to the above criteria, one would prefer a feature analysis in which the assignment of feature values is somehow "natural": for example, the contrast *o/ó* should involve only one feature, and the marked value of this feature should be assigned to *ó*. To see what is involved here, it will be necessary to give a formal rendering of feature assignment. For the sake of simplicity, I will suppose that the features involved are all binary. It will be convenient to use the values "0" and "1" instead of "–" and "+". These numbers, together with the ordinary rules of modulo 2 arithmetics (in which 1+1 is 0, but everything else happens in the usual manner), will give us the finite field GF(2), and questions of feature assignment can be translated into problems involving finite dimensional linear spaces over this simplest of finite fields.

Given a finite set F of purportedly universal features f_1, f_2, ..., f_n, these features form a basis in the n-dimensional linear space GF(2,n) over GF(2). In other words, it is possible to treat the (fully specified) feature matrix of a phoneme as the vector given by its coordinates. Formally, *feature assignment* can be defined as a mapping C from a given set P of phonemes p_1, p_2, ...,p_k *into* the linear space GF(2,n). The points of this space are in one-to-one correspondence to the vertices of a (hyper-) cube in ordinary (Euclidean) n-dimensional space.[5] Naturally, not every vector corresponds to the feature matrix of some phoneme (in some language): the image of P under C will be denoted by Q. Condition (4A) above guarantees only that C must map different phonemes on different vectors, i.e. that C(p) and C(q) must be different if the phonemes p and q are different.

More precisely, feature assignments C_i for the (possibly overlapping) phonemic inventories of various languages and dialects should be treated as a family (set) of mappings of the above sort. The strongest possible condition on a universal feature system is that phonemes having similar (acoustic) properties should have the same feature matrix regardless of the language (dialect) to which they belong: in our formalism, this would mean that whenever phoneme p of language i is (acoustically) similar to phoneme q of language j, we must have $C_i(p) = C_j(q)$.

A weaker, but more realistic, condition is that the feature assignment must be *monotonic:* whenever we have two languages i and j with phonemes p_i, q_i, p_j, and q_j differing only in a single feature F, if

[5]For the earliest analysis of features along these lines, see Cherry 1956, 1957:3.4.

p_i is to the right of q_i on the phonetic scale corresponding to F, and p_i and p_j are both (un)marked with respect to F, then p_j must also be to the right of q_j. One effect of this condition would be to reduce the number of those vectors that appear as the image of some phoneme under some C_i. But even without this condition it is quite clear that certain feature combinations do not correspond to anything attested (or even possible) in natural languages.[6]

It should be emphasized here that the parameters that have to be fixed in this formal model reflect the choices that must be made by the linguist, rather than the choices confronting the language learner. To quote Wheeler (1972:88): "...it may be an important task of a descriptive linguist to decide which features are appropriate in the case of the language he is studying". For instance, the linguist describing the vowel system of Hungarian has to choose a particular set of features: he can adopt the system proposed by Jakobson, Fant, and Halle 1952, Halle and Clements 1983, or any of the innumerable proposals in between. Becker-Makkai 1970 opts for the former, and Vágó 1980 employs the SPE system. The language learner, on the other hand, is supposed to know the 'true' feature system, which is only approximated by the proposals in the literature.

The algebraic structure of linear spaces does not correspond naturally to the linguistic structure of phoneme inventories. The (algebraic) sum of two or more phonemes will only accidentally belong to the inventory. In other words, inventories do not form subspaces. Nevertheless, the sum operation (which gives a vector containing 1 on those positions where the summands were different, and 0 on those where they were identical) can be useful. To see this, we will first identify the universal features with certain vectors. This identification is the obvious one: the feature f_k will correspond to the vector having coordinates 0 for l ($l \neq k$), and coordinate 1 on position k. This way, features can be treated on a par with phonemes. In certain cases (e.g. the feature ⟨vocalic⟩ and the English phoneme *e* in the SPE system), features and phonemes will be identical. In such cases I will say that the phoneme "stands for" the feature in question (and, symmetrically, the feature "stands for" the phoneme). Typically, however, features cannot be represented with the aid of the phoneme inventory in such a direct manner. The following definition is suggested by the minimal pair technique:

A feature f is *witnessed* by a pair of phonemes (p_i, p_j) iff $p_i + p_j = f$.

For example, in the SPE analysis of English segments, the feature ⟨voice⟩ is witnessed by the pairs *(b,p)*, *(f,v)*, etc.; the feature ⟨nasal⟩ by the pairs *(n,d)*, *(b,m)*, etc.; and so on. Clearly, in the feature analysis of a given language only those features are relevant which are witnessed by at least one pair of phonemes in the language. In 2.2, as well as in most feature analyses, this principle is applied implicitly.

The linguistically relevant structure of phoneme inventories is provided by the natural classes appearing in them. Following an idea of Fudge (1967), I will define the *degree of naturalness* of a subset S

[6] Feature cooccurrence restrictions, or FCRs, as proposed by Stanley (1967) will have to be stated for every category (such as degree of voicing) for which the number of contrasts in a given language is not a power of 2. For instance the low/mid/high opposition in vowel height will require the exclusion of the combination *⟨+high, +low⟩. FCRs, however, are not necessarily the by-products of the consistent application of the binarity principle: it is simply an empirical fact that certain combinations of otherwise well-motivated features (corresponding to, say, uvularity and clicks, respectively) do not appear simultaneously in natural languages.

in a phoneme inventory P as the number of rules in which S occurs as context. This makes it necessary to enlarge the surface inventory so as to include underlying phonemes as well. But this can be done without increasing the complexity of the model — the set P will simply have to be replaced by some other set P'. Thus, the class {*b,u*} of Hungarian phonemes will have 0 degree of naturalness, while the class {*p,b,m*} has positive degree of naturalness because the phonemes in it trigger $n \rightarrow m$ assimilation (see ch. 2.5 of Vágó 1980). Needless to say, a subset S' of P' can be natural also by virtue of undergoing, rather than triggering, some process. From the formal point of view, the criteria we employ in distinguishing natural classes from 'unnatural' ones (cf. e.g. Rubach 1982 ch. 3.2) are unimportant: all that matters is that natural classes (and their degree of naturalness) are given externally. In other words, natural classes are present in the system (much like the phonemes) before we attempt to develop a feature analysis.

This is particularly clear for major classes, such as stops or vowels (listed in the first line of (1A) and in (1B) above): the system of major class features (to be discussed in 2.2) has to be constructed in accordance to these classes, rather than the other way round. Those classes that can be expressed by fewer features than their individual members will be called N-classes (cf. Halle 1964:328). Geometrically, N-classes are the intersection of Q with (hyper-) planes parallel to the coordinate axes, and, as such, will depend on the feature analysis (the mapping C) chosen. This definition makes it possible to reformulate the 'compactness' condition discussed above as

(4C) The N-classes given by C should coincide with the natural classes of P. (Compactness)

It is worth noting that the 'classical' feature analysis model, as outlined above, makes rather specific predictions concerning natural classes. First, there cannot be many. If there are p (surface) phonemes, we will need at least $f = \log_2 p$ features to satisfy condition (4A). Each N-class is given by assigning some specific value (+ or –) to some features, while leaving the rest unspecified. This gives us 3 options for each feature, so with the aid of f features, 3^f N-classes can be defined. Therefore, the number of natural classes is expected to be $3^{log_2 p} = p^{1.58}$. This is a very small fraction of the possible (natural or unnatural) subsets, of which there are 2^p. To get an idea of the order of magnitudes involved, suppose $p = 30$. The number of natural classes is $30^{1.58} = 220$ (or $3^5 = 243$ if we use 5 features, rather than the theoretical 4.91); while the number of classes is $2^{30} = 1,074,000,000$.

Second, the set of natural classes must be closed under intersection. More precisely, two natural classes, both containing some phoneme q will intersect either in q, or in a full natural class.[7] This is easy to prove for N-classes (no matter what function C we choose): therefore, the (meta)theory predicts that it must be true for natural classes as well. In practice, natural classes are defined by a few features, i.e. with the majority of the features left unspecified. Since the intersection of such classes will be specified for every feature for which at least one of the original classes were specified, we will eventually get classes in which only a few features are still unspecified — these are more often called *archiphonemes.*

For example, the set of high (diffuse) vowels contains *i,í,u,ú,ü,* and *ű* — this is clearly a natural

[7]It is a matter of definition whether we call one-member classes natural. If we take lexical items to be identity rules, as proposed by Kiparsky (1982), each phoneme will appear in a huge number of rules and will, therefore, have a very high degree of naturalness.

class. So is, presumably, the set of rounded (or the set of flat) vowels: the intersection gives us {*u,ú,ü,ű*}. This is often taken to be an archiphoneme U, the backness (gravity) of which is determined by rules of vowel harmony.

Let me conclude the discussion of the classical theory of features with an historical remark. The first extant treatment of natural classes is given in Pāṇini 1.1.71 (see e.g. Böhtlingk 1887). Pāṇini's method is to arrange the phonemes in a linear sequence (the *śivasūtras*) with some indicatory letters *(anubandha)* interspersed. (For a discussion of the considerations governing the placement of *anubandha* see Staal 1962). Natural classes *(pratyāhāra)* are defined as those subintervals of the *śivasūtras* which end in some *anubandha*.[8] Surprisingly enough, this gives the same predictions. The number of *pratyāhāra* on p phonemes with k equidistant *anubandha* is $\approx p(k+1)/2 \leq p(p+1)/2$, again a small[9] power (at most the square) of p. Furthermore, the intersection of two *pratyāhāra*, if not empty, can also be expressed as a *pratyāhāra*, and is, therefore, 'natural'.

Let us now turn to autosegmental theory. The autosegmental representations of phonemes will be treated as graphs: the edges correspond to association lines (in the usual sense of autosegmental phonology, cf. van der Hulst and Smith 1982), and are therefore undirected. Vertices will usually correspond to autosegments — but unlabelled vertices might also be present. In general, I will make no distinction between vertices and their labels; and in order to maintain this convenient identification, sometimes spurious labels will be added to those nodes (such as the root node and class nodes of Clements 1985) that have no labelling of their own.

The trend in autosegmental phonology is to put more and more features on separate tiers: first tone (as in Leben 1973), then vowel features (such as ⟨ATR⟩ and ⟨back⟩) participating in vowel harmony (Clements 1976), laryngeal features (Thráinsson 1978), nasality (Hyman 1982), syllabicity (Clements and Keyser 1983:3.8), etc. The logical endpoint of such a trend is a representation resembling a paddlewheel (Archangeli 1985:337), in which every feature is on a separate tier, and they are all linked to a central (root) tier. This proposal, also known as the Independent Linking Hypothesis (ILH) gives us a graph commonly known as a *star.* Since it is precisely this configuration that I want to propose for Hungarian vowels, let us examine this first. (More sophisticated 'geometrical' configurations will be discussed in 2.2, as we turn to the representation of consonants.)

Given a central node (here labelled 'V'), the peripheral nodes will have to be labelled either by a *feature value* (such as '–back'), or by the *name of the feature* (i.e. 'back'). In the first case, Stanley's (1967) objections against a ternary use of binary features apply with even greater force in the framework of autosegmental phonology, because in a context like

[8] In addition to using *pratyāhāra,* Pāṇini employs a variety of other devices, most notably, the concept of 'homogeneity' *(sāvarṇya)* as a means of cross-classification. For a modern exposition, see Cardona 1965.

[9] In the *Aṣṭādhyāyī* (Böhtlingk 1887) Pāṇini uses k=14 *anubandha* for p=43 phonemes, which yields 291 potential natural classes (most of which are not actually used in the grammar) – this has to be compared to the eight trillion possibilities given by 2^{43}.

(5)	timing tier	C	V	C
		|		|
	F-tier	+F	____	+F

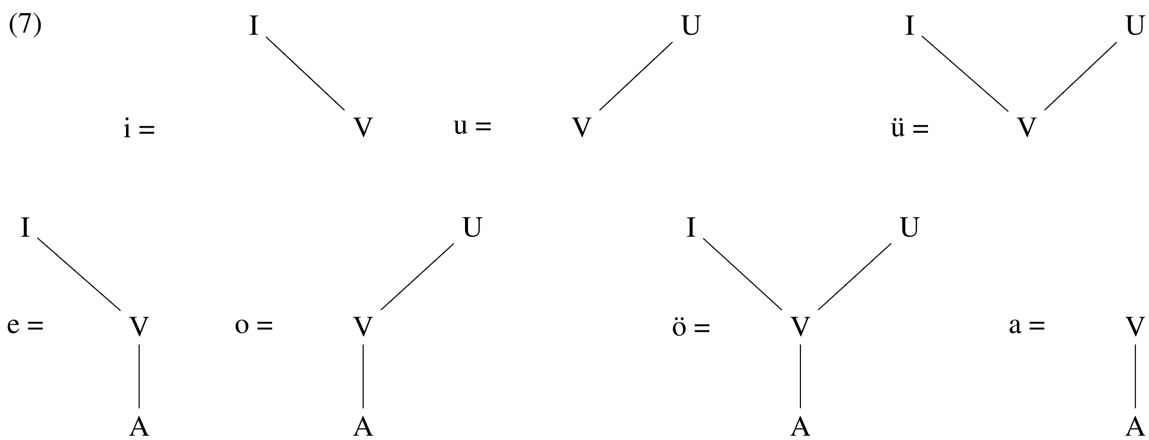
the vowel in the middle can take place in a five-way opposition as given in (6):

(6)	(a)	(b)	(c)	(d)	(e)
CV-tier	V	V	V	V	V
	|	|			
F-tier	+F	-F		+F	-F

In order to escape Stanley's objections, I will use the *name* of the feature as label. This gives us only two choices: a feature is either present on its tier (as in (6a)), or absent (as in (6c)) — conceptually, this is a straightforward restatement of the notion 'privative opposition' as defined by Trubetzkoy (1958: ch. III 2.B). Floating features of type (6e) are still possible, but only as diacritics marking exceptionality. I will return to this matter in 2.3.

Let us suppose, then, that we have n tiers t_1, t_2, ..., t_n arranged in a paddlewheel fashion around a central (timing) tier. In order to handle geminates, affricates, etc. in the usual autosegmental fashion (cf. Clements and Keyser 1983:3.7), it will be expedient to introduce the notion of temporal *span.* The span of an autosegment is defined simply as the set of timing units to which it is linked (either directly or via class nodes). In Hungarian, long vowels will be associated to two (adjacent) timing units. Keeping this in mind, we can disregard the feature ⟨long⟩ (⟨tense⟩) in the following discussion. For Hungarian vowels, the case of *contours*, i.e two or more features (on the same tier) with overlapping spans, can also be disregarded: the features associated to a single timing unit form a 'temporal slice' of the paddlewheel. For (short) Hungarian vowels, we have the following configurations:

(7)



This proposal will be defended in 2.3, here I will simply contrast it to earlier (linear) solutions (for similar proposals, see Rennison 1984, Kaye et al 1985). It is at this point that the formal models developed earlier become useful, for if we set up a one to one correspondence between the tiers t_i and the features f_i (i = 1,2, ...,n), the classical model and the ILH can be shown to be isomorphic. This is particularly clear for the representations depicted in (7): the U-tier corresponds to the feature ⟨flat⟩, the A-tier to the feature ⟨compact⟩ (⟨–diffuse⟩), and the I-tier to the feature ⟨acute⟩ (⟨–grave⟩). If we recast (7) in a feature matrix similar to those given under (2) and (3), it is easy to see that *a* stands for the feature A, *i* stands for I, and *u* stands for U.

But the ILH, be it just a notational variant of the classical model, suggests a geometrical picture far removed from that of n-dimensional linear spaces. Each star lies in a (2-dimensional) *plane,* and the features can serve as basepoints in a barycentric coordinate system. Since *e* is specified as +A, +I, –U and the appropriate vowels stand for these features, the algebraic equation $e = a+i$ that holds among the coordinate vectors can be interpreted as saying that e lies in between a and i. Similarly, *o* lies in between a and u, *ö* in between a, i, and u, etc. The picture emerging from these considerations resembles, to a large extent, the traditional phonemic diagrams for vowels (Delattre et al. 1951, reproduced as Fig. 1.1.2 in Neary 1978).

Let us see what kind of a construction corresponds to natural classes in this model. Suppose, for instance, that features f_1 and f_2 have the value +, features f_{n-2}, f_{n-1}, and f_n have the value –, and every other feature is unspecified. If we take these unspecified values to be +, we get a star T with 'rays' for f_1, f_2, ..., f_{n-3}. If we take them to be –, we get a star B with only two rays (for f_1 and f_2). Every other star corresponding to some element of the N-class must be a proper subset of T and a proper superset of B, and vice versa, every such star corresponds to some element of the N-class. Since labelled graphs form a (distributive) lattice for the usual (set theoretic) operations of union and intersection, we might say that the elements of the N-class correspond to a (closed) *interval* (in the lattice-theoretic sense) between B and T. Every interval of this sort will be called an *M-class.* For the case of the ILH, N-classes and M-classes coincide, making the parallelism between the classical model and the ILH complete. M-classes, however, can be defined for more complicated representations as well: they are simply classes of graphs that contain some graph B and are contained in some larger graph T.

For the case of privative features at least, this provides a definition of the notion 'natural class' that is applicable in any model of feature geometry, and reverts to the standard definition for the ILH. If equipollent features are also permitted, further restrictions must be imposed before we can call the resulting classes truly 'natural'. I will return to this issue in 3.1, where the geometry of morphosyntactic features is investigated in greater detail, and propose the restriction that in a 'geometrical' arrangement of features, only the marked '+' value of an equipollent feature can dominate other features (a condition obviously met by any tree structure composed of privative features).

2.2 Consonants

So far, we have made little use of the possibilities of autosegmental representation. The feature composition of consonants, however, will not be given in the simple 'star' geometry employed for vowels. The reasons for this move of adopting a more complicated, tree-like organization of features are discussed in Clements 1985; here it is sufficient to say that *laryngeal* features will be separated from *supralaryngeal* features, and, in the latter category, *manner* features will be separated from *place* features. These "superfeatures" will be represented by separate nodes (called class nodes) as in (8) below:

```
(8)                 Root node
                  /            \
         Laryngeal node     Supralaryngeal node
                               /              \
                        Manner node       Place node
```

In the model of Clements 1985, the laryngeal node dominates features such as ⟨voice⟩, ⟨spread glottis⟩, and ⟨constricted glottis⟩; the manner node dominates ⟨consonantal⟩, ⟨nasal⟩, ⟨lateral⟩, ⟨continuant⟩, ⟨strident⟩, and ⟨sonorant⟩; and the place node dominates ⟨coronal⟩, ⟨labial⟩, ⟨anterior⟩, ⟨distributed⟩, ⟨high⟩, ⟨back⟩, etc.

In the present model, the laryngeal and place features are left largely in this form (though the features ⟨high⟩, ⟨back⟩, ⟨round⟩, and ⟨low⟩ are replaced by I, U, and A), but the manner features are reorganized. The reasons for this move are twofold: first, the 'barycentric' interpretation of features outlined in 2.1 makes it desirable to employ features that 'stand for' complex articulatory targets (fully specified phonemes), and, second, this way certain redundancies in the system of major class features can be eliminated (see 2.4). Therefore, certain 'partial' features will be replaced by 'phoneme-like' features (e.g. ⟨nasal⟩ by N, just as ⟨spread glottis⟩ will be replaced by H under the laryngeal node), and the manner node, rather than being the center of a star, will have its own 'geometry'.

```
(9)             I/U  (=Hi)
                 |
                V/C  (=X)
              /      \
             A      Re/Ob (=Son)
                     /   \
                  N/L  St/Fr (=Cont)
```

The structural unity of a phoneme is established by the *root node* (not depicted above), which links directly to the timing tier. In the case of geminates, the root node spreads over two adjacent timing units, and in the case of affricates (in Hungarian, *c, cs, dzs*), two root nodes share a single timing unit. The nature of the timing units is somewhat controversial — I will suppose that they are C and V units in the sense of Clements and Keyser 1983. X will denote a 'pure' timing unit that has neither C nor V specification (see Levin 1985, Guerssel 1986). Here X is used simply as a *variable* that ranges over the two-member set {C,V}; its function is essentially that of a cover feature. Hi is a similar cover feature: it can take the value I or the value U.

(9) is specific to Hungarian inasmuch as the Hi features I and U can be linked to vowels and consonants alike (directly under the root node for vowels and glides, and indirectly under the place node, for palatalized consonants). Hungarian makes no phonemic distinction between labialized and non-labialized consonants, so there is no need to link U to any node in the representation of consonants, but in other languages (e.g. Kabardian, cf. Kuipers 1976), one might link even the feature A to consonants.

The features C and V (corresponding to the SPE feature ⟨syllabic⟩) are in equipollent opposition (in the sense of Trubetzkoy 1958:III), therefore they appear on the same autosegmental tier.[10] On the next lower level, we find the features Re(sonant) and Ob(sturent), corresponding roughly to the feature ⟨sonorant⟩. In Hungarian, both of these features can be linked only to Cs. To the Re feature we might link the feature N(asal) or L(ateral), or we might leave it unlinked. In Hungarian this last possibility corresponds to the phoneme *r*. Obstruents, on the other hand, can be further specified as St(op) or Fr(icative). Ignoring redundantly assigned feature values, this distinction corresponds to the SPE feature ⟨continuant⟩.

After this preparation, the autosegmental representation of Hungarian consonants becomes a trivial matter. (10) below depicts the representations in a planar format. The arrows in the representation of affricates serve to indicate progression along the third (temporal) dimension[11] The orthographic representation of the consonants is given at the root node solely for ease of identification — it is *never* part of the representation itself. Voiced/voiceless pairs are given by a single representation containing the voicing feature (Z) in parentheses.

Since the feature matrices corresponding to these graphs have the usual values (cf. Halle and Clements 1983) they require no special justification. Nevertheless, as (10) departs from the feature system of Vágó in certain respects, I feel obliged to comment on these discrepancies.[12]

[10] Gradual oppositions in the sense of Trubetzkoy (loc cit) involve more than one tier. Various degrees of voicing (see e.g. Ladefoged 1971:ch. 2) or register tone are typical examples; the usual treatment is that the tier where the features determining the finer subdivisions appear is anchored in the tier where features for gross distinctions appear. In the graph-theoretical formalism outlined in 2.1 this means that the relevant class of graphs is that of *chains*. For example, if X is a voiceless stop, X–Z can be a murmured, and X–Z–Z a fully voiced stop.

[11] Thus, they should not be confused with the 'rewrite' arrows of phonological rules.

[12] One possible objection to the representations given in (10) is that the feature ⟨ant⟩ is redundant for labials, such as *m, p/b,* or *f/v*. While this is certainly true, removing the redundant ⟨ant⟩ from the representation would make it necessary to take a stand on the issue of the exact level of representation at which redundant features are introduced. Readers familiar with the recent underspecification debate (Archangeli and Pulleyblank 1986, Steriade 1987, Itô and Mester 1988) are no doubt aware that this is an extremely complex issue in its full generality. Therefore, it seemed less advisable to discuss underspecification in the context of labials where the competing proposals can all be made to work, than in the domain of vowel harmony, where there is a more complex body of data to justify the analysis. For a discussion of Hungarian vowel harmony in terms of underspecified (SPE) features see 2.6.

(10)

```
         h         j             r             l
        / \       / \           / \           / \
       H   |     Z   \         Z  /\         Z  /\
           Ob        /\          Re \          Re \
                  cor  I            /\         /   /\
                                 cor ant      L cor ant

           n                ny                 m
          / \              / \               / \
         Z  /\             Z  /\               Z  /\
           Re \             Re \               Re \
          /   /\           /   /\             /    /\
        N  cor  ant       N  cor I          N  lab ant

          t/d             ty/gy               p/b
          / \              / \                / \
        (Z) /\           (Z) /\             (Z) /\
           Ob \             Ob \               Ob \
          /   /\           /   /\              /   /\
         St cor ant       St cor I          St lab ant

         sz/z              s/zs               f/v
          / \              / \                  \
        (Z) /\           (Z) /\                 /\
           Ob \             Ob \           ____Ob \
          /   /\           /    |          /   /   /\
         Fr cor ant       Fr   cor      (Z)  Fr lab ant

          k/g             c                cs/dzs
          / \             |                 / \
        (Z)  |           / \              (Z)  /\
             Ob         Ob  \___               Ob \___ant-->0
             |        / \   \  \          / \  \
             St     St-->Fr cor ant      St-->Fr cor
```

First, Vágó takes *ty, gy, ny,* and *j* to be ⟨–cor⟩, probably in order to simplify the statement of some of his assimilation rules. Since in the autosegmental formalism no significant gain in simplicity can be achieved by taking these elements to be ⟨–cor⟩, I will employ the (phonetically more natural) ⟨+cor⟩.

Second, Vágó takes *v* to be a “sonorant consonant” (which is certainly a possibility) and *h* to be an “obstruent glide”. Here *h* is represented by the feature H on the laryngeal tier. Since aspiration gives rise to no phonemic contrast in Hungarian, we may put H on the same tier as Z (the feature for voicing). This will make the formulation of certain allophonic rules (e.g. the automatic voicing of intervocalic *h*) a bit awkward, but captures the *phonologically* complementary distribution of aspiration and voicing. Since H stands for *h* (in the sense of 2.1), we may assign it the feature Ob on the sonorant tier (and thereby capture the essential insight of Vágó), without being forced to the position that *h* is a glide.

As for *v*, the main reason for ascribing any special status to it comes from the fact that it does not trigger regressive voicing assimilation, while every other voiced obstruent does. Since voicing assimilation is simply the (backward) spreading of the feature Z, *v* will not give rise to it if it is specified for Z in the phonemic core (cf. Halle and Vergnaud 1982) or, as proposed in (10) above, under the manner node. Obstruent devoicing, on the other hand, cannot be described as a rule of spreading (in the framework adopted here, there is no feature -Z to spread) but has to be a rule of delinking. Since the environment triggering this delinking is specified negatively, i.e. as a *non*resonant consonant *not* linked to Z (notice that this includes *h* without any further stipulation), it is possible that these processes cannot be subsumed under a single rule of voicing assimilation.

Perhaps the interaction of the well-known degemination rule (cf. Vágó 1980:2.11) and the voicing assimilation rule(s) can provide some evidence in favor of the solution adopted here. Given the phenomenon of geminate blocking (Schein and Steriade 1986), we would expect (de)voicing to be inoperative for geminates. Therefore, the correct surface form [töpše:g] can be derived from underlying *több* + *še:g* ‘more + suff.nom.; majority’ only if degemination (which simplifies a geminate consonant cluster before a consonant, see (43)), feeds devoicing. But if degemination feeds voicing assimilation as well, we would derive the surface form [skivbö:l] from underlying *skiff* + *bö:l* ‘skiff + IAL’. The more regular [skifbö:l] can be derived only if degemination follows voicing assimilation. However, the intuitions of native speakers are not a reliable source of data here, and in the absence of experimental results, the case for the separation of the voicing rule from the devoicing rule is not very strong.

In general, there are quite a few conflicts between Vágó’s (1980:ch. 2) treatment of consonant sandhi and my chief descriptive source (Elekfi 1973) — clearly, a great deal of empirical work is necessary before we can go any further. Since the issues involved are tangential to morphology proper, I will not develop a full-fledged rule system here and will only outline the most prominent rules in 2.5.[13] Instead, I will tap another source of evidence, namely interconsonantal *consonant epenthesis* in dialects.

This kind of epenthesis is a most puzzling phenomenon from the point of view of linear phonology: it does not simplify the cluster at all (in fact, it makes it more complex), and there seems to be no natural phonetic process in the background. From the autosegmental point of view, however, it becomes

[13]The reader who wishes to familiarize himself with the feature analysis proposed above might attempt to recapitulate the essence of Vágó’s ‘natural assimilation rules’ in an autosegmental format. The solution proposed in 2.5 makes a number of predictions that set it apart from Vágó’s system: for instance *a:č+t*^{s}*eruza* ‘carpenter’s pencil’ or *el+ja:r* ‘proceed’ will not become *[a:t^{s}t^{s}eruza] or *[ejja:r] but will remain [a:čt^{s}eruza] and [elja:r] on the surface.

clear that it is precisely the breakdown of the most fundamental tenet of linear phonology, namely the 'absolute slicing hypothesis' (see Goldsmith 1976) that gives rise to this sort of consonant epenthesis. Autosegmental tiers can be interpreted as lines in an 'orchestral score' that governs the position of the articulators. For instance, the cluster *pr* is represented as (11A) in linear (SPE) phonology:

```
(11)
(A)                (B)                   (C)
   p      r              p      r             p      t      r

                    ______________        _________________
[-syll] [-syll]     |    -syll    |       |     -syll      |
                    --------------        -----------------
[+cons] [+cons]     |    +cons    |       |      +cons     |
                    --------------        -----------------
[-son ] [+son ]     | -son | +son |       |   -son   |+son |
                    --------------        -----------------
[-cont] [+cont]     | -cont| +cont|       |   -cont  |+cont|
                    --------------        -----------------
[-cor ] [+cor ]     | -cor | +cor |       |-cor|   +cor    |
                    --------------        -----------------
[+lab ] [-lab ]     | +lab | -lab |       |+lab|   -lab    |
                    --------------        -----------------
[+ant ] [+ant ]     |    +ant     |       |      +ant      |
                    --------------        -----------------
```

In autosegmental phonology, this is depicted as (11B) (for the notation, see Anderson 1976). Now, if the place and the manner 'parts' of the score are not perfectly synchronized, so that the place node of *r* chimes in a bit earlier, we get a score like (11C). If we slice this up into parts homogeneous in every feature, the result will be the sequence *ptr*.

In fact, the epenthetic *t* is well attested in this context: dialectal forms such as *ptrücsök* (ECH *prücsök*) 'cricket', *ptrüsszent* (ECH *prüsszent*) 'sneeze', etc. are not uncommon (see B. Lőrinczy 1979:ch. 133). (ECH stands for Educated Colloquial Hungarian, and SLH for Standard Literary Hungarian — see Nádasdy 1985)

This way, autosegmental phonology can describe the mechanism of interconsonantal consonant epenthesis in two steps: first, desynchronization of tiers, and, second, attributing phonemic status (i.e. assigning a full C slot) to the resulting intermediate element. (That the second step need not be complete can be seen from a phonetic description of the facts. B. Lőrinczy (1979:144) cites Laziczius (1944:87), who takes the epenthetic consonant to be a 'bilabial tremulant' showing, in effect, that the 'tremulant' manner node of *r* chimes in before the 'bilabial' place node of *p* could end.)

Moreover, the autosegmental treatment can go further than simply describing the facts: it can make predictions concerning the nature of the epenthetic consonant. For instance, the framework leaves open the possibility that the place node, rather than advancing over the manner node, lags behind it. This possibility is manifested in dialectal *tüsténtkedik* (ECH *tüsténkedik*) 'bustle', *mintket* (ECH *minket*) 'we-ACC', etc. (see ch. 134 of B. Lőrinczy 1979). Also, tiers dominated by a lower node (e.g. the ⟨anterior⟩ node, dominated by ⟨place⟩) can also be desynchronized: cf. dialectal *istmét* (ECH *ismét*) 'again'.

However, it must be admitted that at present the theory (as presented e.g. in Wetzels 1985) makes no predictions concerning the direction of the desynchronization or about the tiers involved. Nevertheless, the limitations imposed on the class of epenthetic consonants are borne out by the data. Of the twenty-odd cases discussed by B. Lőrinczy (chs. 133-136, 138-139, 141-143, 154-155), a few are likely to involve folk etymology: *nyársfa* (ECH *nyárfa*) 'poplar' p. 159, *pironkság* (ECH *pironság*) 'shame' p. 157. Aside from these, there is but one class not amenable to the proposed treatment. This class (discussed in ch. 138 of B. Lőrinczy 1979) involves *f* or *v* in the rightside environment of the epenthetic consonant — examples are:

partfűm	(ECH parfűm)	'perfume'
készakartva	(ECH készakarva)	'willfully'
epertyfa	(ECH eperfa)	'mulberry tree'

This suggests that we should treat *f* and *v* as ⟨–continuant⟩, i.e. with Ob dominating the feature St (rather than Fr). But the fact that some minor 'cleanup' is needed here should not obscure the fundamental superiority of the autosegmental approach. Rather then resorting to a *functionalist* (teleological) explanation like onomatopoeia or even euphony, the autosegmental treatment provides a *causal* explanation by connecting the nature of the epenthetic consonant to the phonetic process of desynchronization.

Finally, we can treat the epenthetic *d* of *péndz* (SLH *pénz*) 'money', *bendzin* (SLH *benzin*) 'gasoline', etc. (see E. Abaffy 1975) in a similar manner: here the ⟨continuant⟩ tier lags behind the ⟨sonorant⟩ tier. This takes care of a potential counterargument to the proposal of É. Kiss and Papp (1984) that we should treat *dz* as a sequence of two phonemes. As the perceptive reader will have noticed, only *dzs* is granted phonemic status here. Although this move decreases the symmetry of the consonant system (*c* will have no voiced counterpart on the systematic phonemic level), it explains the otherwise mysterious distributional limitations of *dz*, which cannot appear word-initially or after a consonant,[14] and must take two timing units intervocally and word-finally, if preceded by a vowel.

[14]The single exception to these rule appears to be the dialectal *brindza* 'coach' in which the *d* is arguably epenthetic just as in the examples *péndz* etc discussed above.

2.3 Vowel harmony

The basic characteristics of Hungarian vowel harmony are well known. In fact, they are a little too well known, inasmuch as the same set of data has been subjected to an astonishing number of analyses. I do not assume the reader to be familiar with this literature (for an overview, see van der Hulst 1985 or Siptár 1984) and I will exemplify all the facts that appear to be relevant. As a first step, let me deduce the 'Jakobsonian' feature system that was assumed without argumentation in 2.1.

Vowel length is distinctive in Hungarian: it is easy to find minimal pairs like *kor* 'age' vs. *kór* 'sickness'; *vidd* 'carry!' vs. *vídd* 'fight!'; *üröm* 'artemisia' vs. *űröm* 'my space'; *öröm* 'joy' vs. *őröm* 'my guard'; *buja* 'lush' vs. *búja* 'his sorrow' that differ only in the length of the vowel in question. However, as we have seen in 2.1, *é* and *á* differ from *e* and *a* not only in quantity, but in quality as well, and a key question of the analysis is to understand why we should follow Becker-Makkai 1970 and adopt the same features for short and long vowels nevertheless.

In order to show that *a/á* and *e/é* belong to the same short/long series as the other vowels, I will consider two independent processes. First, there is a a rule of Low Vowel Lengthening (cf. Vágó 1980:1.1.2) whereby stem-final *a* becomes *á* and *e* becomes *é* before certain suffixes such as the accusative *at/et/ot/öt*, the dative *nak/nek,* and the superessive *on/en/ön*.

Since the lengthening of stem-final *a* is triggered by exactly the same set of suffixes that trigger the lengthening of stem-final *e*, we have to treat Low Vowel Lengthening as a unified process, i.e. we must suppose that the change from *a* to *á* happens along the same features as the change from *e* to *é*. Thus, if *a* and *e* have identical values for some feature, *á* and *é* must also have identical values for that feature, and vice versa.

The lengthening rule shows that *á* and *é* differ from *a* to *e* at least in length, but it leaves open the possibility of other differences such as raising or unrounding. Recall, however, that the height differences work in the opposite direction: *a* is *lowered* to *á* but *e* is *raised* to *é* by the same rule. In order to derive the correct values, Vágó (1980:19) is forced to posit two 'adjustment rules' which can be collapsed only by using Greek letter variables referring to the opposite values of ⟨low⟩ and ⟨long⟩. From this I conclude that the rule of Low Vowel Lengthening should be formulated so as not to affect height.

Another way of showing that length is the *only* feature that distinguishes *a* from *á* and *e* from *é* is by studying a process of shortening which affects a different set of stem vowels. In a lexically restricted set of consonant-final stems, certain suffixes, such as the the accusative *at/et/ot/öt*, the dative *nak/nek*, but *not* the superessive *on/en/ön,* will trigger the shortening of the vowel in the last syllable of the stem. Examples:

(12)	*tél/telet*	'winter/ACC'	*nyár/nyarat*	'summer/ACC'
	tűz/tüzet	'fire/ACC'	*víz/vizet*	'water/ACC'
			nyúl/nyulat	'rabbit/ACC'

Again, the change is triggered by exactly the same set of suffixes for each stem, and this establishes the unity of the process. But here the non-low vowels show that the only feature the rule affects is length.

Thus, it follows that this is the only change for low vowels as well. In particular, if we would permit the rule to round *á* to *a*, we would mistakenly create a round vowel for *e* and *i* as well. Since this is completely unmotivated, we must derive the roundness of *a* by a late rule.

Before turning to the larger issue of discrepancies between the underlying and surface representations, let me summarize the implications of the preceding discussion. In order to maintain the unity of Low Vowel Lengthening, we must suppose that it operates on representations which are fairly distant from what one would propose on phonetic grounds. The shortening rule leads to the same conclusion, namely that the only difference between corresponding short and long vowels is in the number of timing units they take, at least at the level where these rules operate. Therefore, the difference in the surface quality of *a* and *á*, or *e* and *é,* cannot be exploited to explain the purported asymmetry between the ⟨low⟩ values of *e* and *é*, as it was done e.g. in Steriade 1987. Rather, we predict the harmonic behavior of the corresponding long and short vowels to be identical. As we shall see later, this conclusion holds not only for the 'regular' vowels, but for most of the 'abstract' ones as well.

So far we have established that for the purpose of writing phonological rules that preserve the unity of vowel shortening/lengthening processes in Hungarian, we need a feature system which is fairly removed from the phonetic characteristics of these vowels. This situation is typical: rules that capture linguistically significant generalizations can be far removed from surface 'naturalness' (Anderson 1981). This being the case, we can not make use of what I called at the beginning the 'observational' aspect of the features to argue for a particular feature analysis. Rather, we are forced to make use of their 'descriptive', classificatory function. Let us first see how one can deduce the remaining features from the patterns of harmonic alternations in Hungarian. The basic 'binary' alternation involves, among others, the following suffixes:

(13)

nak/nek	'DAT'	dative
nál/nél	'ADE'	adessive
tól/től	'ABL'	ablative
nok/nök	'professional characterized by'	
	e.g. *szó* 'word', *szónok* 'orator'	
unk/ünk	'1PLPOSSG'	1st plural possessive singular
ú/ű[15]	'having'	e.g. *nagy láb* 'big foot'
		nagy lábú 'having big feet'

This list in (13) is representative: there are no binary alternations involving other pairs of vowels.[16] Disregarding exceptional stems for the time being, we can say that stems in which the last vowel is *a, o,*

[15] In ECH, or at least in the author's own dialect, this suffix is actually short. In a list of more than 3,000 suffixes and suffix-combinations (Veenker 1968) I have found only one more candidate: *tyú/tyű* as in *pattantyú, billentyű* etc. (The back alternant, as well as the suffix *ú/ű* as in *szomorú, keserű,* were suggested to me by Péter Siptár, who follows the SLH standard and pronounces these with long vowels.)

[16] The alternation in the 3rd sg. present definite suffix *ja/i* involves consonantal changes as well — for a discussion see 4.1.

or *u,* (short or long) take the first alternant of the suffixes listed above, and that all other stems take the second alternant. (Except for stems in which an *a, o,* or *u* is followed by one or more *i* or *e* — these will be discussed later.) Examples:

(14)

stem	2-ary *a/e*	2-ary *o/ö*	2-ary *u/ü*	gloss
bab	babnak	babtól	babunk	'bean'
báb	bábnak	bábtól	bábunk	'puppet'
rum	rumnak	rumtól	rumunk	'id.'
húr	húrnak	húrtól	húrunk	'chord'
bot	botnak	bottól	botunk	'stick'
drót	drótnak	dróttól	drótunk	'wire'
hit	hitnek	hittől	hitünk	'belief'
víz	víznek	víztől	vizünk	'water'
fej	fejnek	fejtől	fejünk	'head'
érv	érvnek	érvtől	érvünk	'argument'
tök	töknek	töktől	tökünk	'pumpkin'
bőr	bőrnek	bőrtől	bőrünk	'skin'
füst	füstnek	füsttől	füstünk	'smoke'
bűn	bűnnek	bűntől	bűnünk	'sin'

Thus, there should be a feature that distinguishes *a, o, u,* from *e, i, ö, ü.* This, of course, is the feature 'I' (⟨–grave⟩, or ⟨–back⟩). Next, we should note the existence of a four-way[17] alternation as shown, among others, by the following suffixes:

(15)

at/et/ot/öt	'ACC'
ak/ek/ok/ök	'PL'
am/em/om/öm	'1SG POS'
as/es/os/ös	'having'

Quaternary suffixes, though ignored in standard treatments, are in fact anything but marginal: the most frequently encountered suffixes, such as the accusative, the 1st and 2nd sg possessive, and the plural are all of this form. Nor are they restricted to inflection: the last example, *as/es/os/ös* 'having, having to do with' is a high-frequency derivational suffix that forms adjectives from nouns. Note that *a, e, o, ö* form a natural class because they are the only vowels that appear in four-way alternations. Thus, we will need a feature that distinguishes these vowels from *u, i, ü.* This is the feature 'A' (⟨–diffuse⟩, or ⟨–high⟩).

In order to keep the seven vowels distinct, we need a third feature, which I called 'U' in 2.1. The values of this feature are established on the basis of two kinds of vowel alternations, namely ternary harmonic alternations (which are automatic and meaningless), and proximal/distal vowel symbolism

[17] In vowel-initial suffixes, the suffix vowel is dropped regularly after stems ending in vowels, and occasionally after consonant-final stems — for the investigation of vowel harmony, this phenomenon can safely be disregarded.

(which is meaningful). Ternary suffixes include the allative *hoz/hez/höz* and the superessive *on/en/ön*. There are no vowels other than *o, e, ö* that take part in a three-way alternation. Again ignoring exceptional stem vowels, the basic pattern can be stated as follows: stems in *a, o, u* take the *o* alternant, stems in *e, i* take the *e* alternant, and stems in *ö, ü* take the *ö* alternant. Examples:

(16)	stem	SUE	ALL	gloss
	bab	babon	babhoz	'bean'
	báb	bábon	bábhoz	'puppet'
	rum	rumon	rumhoz	'id.'
	húr	húron	húrhoz	'chord'
	bot	boton	bothoz	'stick'
	drót	dróton	dróthoz	'wire'
	hit	hiten	hithez	'belief'
	víz	vizen	vízhez	'water'
	fej	fejen	fejhez	'head'
	érv	érven	érvhez	'argument'
	tök	tökön	tökhöz	'pumpkin'
	bőr	bőrön	bőrhöz	'skin'
	füst	füstön	füsthöz	'smoke'
	bűn	bűnön	bűnhöz	'sin'

The pattern exemplified in (16) divides the vowels that took the I (⟨–grave⟩, or ⟨–back⟩) alternant in the binary case into two groups, *i, e* and *ö ü*. These groups must be separated by the new feature U. The question is where the other vowels belong in this classification. In order to answer this question, I will invoke a 'functionalist' principle of maximal contrast which says that phoneme pairs that carry meaning distinctions should be maximally contrasted in form.

Hungarian has a pervasive system of proximal/distal distinctions in pro-forms. Examples are *itt/ott* 'here/there', *ez/az* 'this/that', *így/úgy* 'this way/that way', *ekkor/akkor* 'this time (now)/that time (then)'. In order to keep the representation of *i* and *u* maximally distinct, we must classify *u* with *ü, ö*, and similarly, if we want to keep the representation of *i* and *o* maximally distinct, we have to classify *o* with *ü, ö*. By the same token, *a* would fall together with *ü, ö*, but if we want to keep the feature matrices of *o* and *a* distinct, we must put *a* together with *i, e*. This observation completes the deduction of the feature system on the basis of vowel alternations.

Note that I provided justification for three features, I, A, and U solely on the basis of vowel alternations. The probability of these features coinciding with the phonetically motivated ⟨back⟩ (⟨grave⟩), ⟨high⟩ (⟨diffuse⟩), and ⟨round⟩ (⟨flat⟩) is 1/840, even if we take it into account that the arguments used for these distinctive features provide no evidence as to which value should be taken as marked (+), or unmarked. Although the argument based on the principle of maximal contrast is less than fully compelling, the chances that the partially determined feature matrix developed on the basis of harmonic behavior alone should be compatible with (2) and (3) are still less than one percent. It is particularly noteworthy that the only case where the functional argument must be overridden by considerations of distinctness is the case of *a*, which is actually ⟨+round⟩ on the surface.

Such an improbable 'coincidence' would call for an explanation even if it was restricted to Hungarian — but in fact the phonological patterning of languages is nearly always less abstract than what one would expect on the the basis of the apparent arbitrariness of the methods by which the feature analysis was established. In this note I can offer only the beginnings of an explanation: I would like to maintain that the process for discerning the phonological patterning is not really arbitrary. The idea is that in general we can deduce a feature analysis in two steps. First, the rules can be established using purely phonemic notation. Second, the sets of phonemes triggering or undergoing some rule can be taken as natural classes. This means that natural classes are independently given to us (provided that the rules are given), and the feature analysis is just a way of explicating natural classes in terms of N-classes (cf. 2.1 above).

The view that the 'objective' (phonetic) relations between speech sounds can obscure the true 'psychological' (phonological) relations goes back at least to Sapir (1925). From this perspective, the high degree of similarity between the abstract natural classes containing phonemes which behave similarly in rules, and the concrete natural classes containing phonemes which have similar articulatory properties, is not really surprising: a rule with a simple articulatory (or perceptual) statement is easier to learn (or recognize) than one requiring a more complex statement. Thus, we expect the 'abstract' features emerging from the pattern given by the rules to be very similar (though not necessary identical) to the concrete phonetic features which are universally given.

As the examples in (16) show, the basic pattern of harmonic alternations is the same for short and long vowels. Since this is also true of the exceptional patterns presented below, I will not mention the long vowels separately. For the sake of simplicity, *a, o, u* will be called *back; i, e* will be called *neutral;* and *ü, ö* will be called *front* vowels. Neutral vowels will *not* be called 'front', their featural classification and phonetic properties notwithstanding.

Exceptional vowels, such as the *í* of *híd* 'bridge' are well known in the literature: the fact that they take back suffixes (*hídnak, *hídnek*), rather than the regular front suffixes, was taken as evidence for absolute neutralization rules (Ringen 1977). Indeed, they provide a clear counterexample to the Alternation Condition (Kiparsky 1968), and the problem posed by 'abstract vowels' could not be solved until the emergence of autosegmental phonology.

Here I will show that the set of abstract vowels in Hungarian is much more complex than previously suspected: in particular, not only neutral vowels have abstract counterparts. To see this, I will group stems together according to their harmonic behavior, i.e. according to the suffix-alternants they select. I will develop a taxonomy which is applicable not only to stems, but also to fully formed words that can undergo further suffixation. The classification presented below offers a theory-neutral descriptive framework encompassing all the non-vacillating Hungarian data.

In the binary case, all stems can be divided into two classes, 'BACK' and 'FRONT', according to the quality of the alternant they select. Binary alternants can be arranged in parallel series: if a stem takes

-nak in the dative, it will take *-nál*, rather than *-nél,* in the adessive, *-tól,* rather than *-től,* in the ablative, etc. In a corpus of more than 35,000 nouns,[18] I was unable to find non-vacillating counterexamples that would take, say, *-nak* in the dative but *-től* in the ablative. Thus, the generalization that binary suffixes can be arranged in parallel series is extremely robust, with no systematic counterexamples.

With the introduction of quaternary suffixes, a four-way partitioning results, depending on the quality of the vowel in the plural suffix *ak/ek/ok/ök*. Those stems that take *-ak* are grouped together in Class I, those that take *-ok* are in Class II, those that take *-ek* are in Class III or IV, and those that take *-ök* are in Class V. The difference between Class III and Class IV will be established shortly — until then, we will consider Class III and Class IV together.

It should be emphasized here that the classes are set up on the basis of the suffix vowels, rather than on the basis of the stem vowel. Since Hungarian has root-controlled harmony, there is a correspondence between the two. Class I contains the exceptional back vowel stems, and Class II contains the regular back vowel stems. Class III contains the regular neutral vowel stems, Class IV contains the exceptional front vowel stems, and Class V contains the regular front vowel stems. The distribution of exceptional vowels will be discussed in greater detail later on.

The partitioning into classes is justified by the fact that other quaternary suffixes will take the alternant with the same vowel. Indeed, no stem can subcategorize for *-ak* in the plural but *-ot* in the accusative, and in general the distributions of the quaternary alternants, as far as it can be established,[19] are completely parallel. In the following table (17), quaternary suffixes are represented by the plural.

The list in (17) is exhaustive in the sense that each surface vowel is exemplified in every class where it appears: for instance, there are no *i*-stems in Class IV or Class V. Moreover, each non-vacillating Hungarian word falls into one of these classes, irrespective of morphemic composition or lexical category. For the sake of clarity, all the possibilities are exemplified by monosyllabic (and monomorphemic) noun stems.

[18] Most of the data discussed here was gleaned from a dictionary database (Kornai 1986).

[19] The alternating vowel is lost after certain stems for some of the quaternary suffixes, e.g. we have *borok* 'wine-PL' but *bort* 'wine-ACC'. This phenomenon will be discussed in 4.2.

(17)

	stem	4-ary	3-ary	2-ary *a/e*	2-ary *o/ö*	2-ary *u/ü*	gloss
I							
	had	hadak	hadhoz	hadnak	hadtól	hadunk	‘army’
	ház	házak	házhoz	háznak	háztól	házunk	‘house’
	lyuk	lyukak	lyukhoz	lyuknak	lyuktól	lyukunk	‘hole’
	kút	kutak	kúthoz	kútnak	kúttól	kútunk	‘well’
	hold	holdak	holdhoz	holdnak	holdtól	holdunk	‘moon’
	ló	lovak	lóhoz	lónak	lótól	lóunk	‘horse’
	híd	hidak	hídhoz	hídnak	hídtól	hídunk	‘bridge’
	héj	héjak	héjhoz	héjnak	héjtól	héjunk	‘crust’
II							
	bab	babok	babhoz	babnak	babtól	babunk	‘bean’
	báb	bábok	bábhoz	bábnak	bábtól	bábunk	‘puppet’
	rum	rumok	rumhoz	rumnak	rumtól	rumunk	‘id.’
	húr	húrok	húrhoz	húrnak	húrtól	húrunk	‘chord’
	bot	botok	bothoz	botnak	bottól	botunk	‘stick’
	drót	drótok	dróthoz	drótnak	dróttól	drótunk	‘wire’
	zsír	zsírok	zsírhoz	zsírnak	zsírtól	zsírunk	‘fat’
	cél	célok	célhoz	célnak	céltól	célunk	‘goal’
III							
	hit	hitek	hithez	hitnek	hittől	hitünk	‘belief’
	víz	vizek	vízhez	víznek	víztől	vizünk	‘water’
	fej	fejek	fejhez	fejnek	fejtől	fejünk	‘head’
	érv	érvek	érvhez	érvnek	érvtől	érvünk	‘argument’
IV							
	hölgy	hölgyek	hölgyhöz	hölgynek	hölgytől	hölgyünk	‘lady’
	tőgy	tőgyek	tőgyhöz	tőgynek	tőgytől	tőgyünk	‘udder’
	sült	sültek	sülthöz	sültnek	sülttől	sültünk	‘roast’
	tűz	tüzek	tűzhöz	tűznek	tűztől	tűzünk	‘fire’
V							
	tök	tökök	tökhöz	töknek	töktől	tökünk	‘pumpkin’
	bőr	bőrök	bőrhöz	bőrnek	bőrtől	bőrünk	‘skin’
	füst	füstök	füsthöz	füstnek	füsttől	füstünk	‘smoke’
	bűn	bűnök	bűnhöz	bűnnek	bűntől	bűnünk	‘sin’

As can be seen from (17), the selection of the suffix vowel in a quaternary suffix determines the selection of binary suffixes, i.e. whether a stem belongs to the FRONT or to the BACK class: all stems that take *-ak* (Class I) or *-ok* (Class II) in the plural take back suffixes and the rest take front suffixes. In other words, the BACK class is made up from Class I and Class II, and the FRONT class is made up from the rest. However, the choice of ternary alternant cannot be fully determined from the fact that a stem takes the *e*-alternant of quaternary suffixes: such stems generally take the *e*-alternant (as expected), but not always. Those stems that take the *e*-alternant both in ternary and quaternary suffixes are grouped together in Class III, which contains the overwhelming majority of neutral vowel stems. But there are a number of stems in *ö* and *ü* that also take *-ek* in the plural, these are collected in Class IV.

It should be emphasized that the behavior of Class IV stems is qualitatively different from that of vacillation: forms like **hölgyhez, *tőgyhez, *sülthez*, and **tüzhez* are unacceptable in every idiolect. In ECH or SLH, the forms **hölgyöt, %tőgyöt*[20] **sültöt*, and **tüzöt* are clearly unacceptable. Thus, the stems in Class IV require new kinds of abstract vowels.

I have mentioned above that the quality of the stem vowel determines the harmonic behavior of the stem to a large extent. If the stem vowel is front (*ö* or *ü*), the stem must be in Class IV or Class V, and if it is back (*a, o,* or *u*), the stem will belong in Class I or Class II. As long as the last vowel of a polysyllabic stem (or word) is not neutral (*i* or *e*), the quality of the last vowel will decide the harmonic behavior of the whole stem (if the last vowel is neutral, the situation is more complex, see 2.6).

The choice between Class IV and Class V is lexically determined, as is the choice between Class I and Class II. However, only the selection of Class I or Class IV has to be marked in the lexicon — the default case is Class II for stems in *a, u,* and *o*, and Class V for stems in *ü* and *ö*. This is particularly clear for Class IV, which contains roughly 20 monomorphemic stems, as opposed to the 150 monomorphemic stems in Class V. That Class IV is the exceptionally marked class can also be seen by the fact that all recent loans in *ü* or *ö* are in Class V. Although Class I is much larger (it contains more than five hundred monomorphemic members), it is still considerably smaller than Class II (which has several thousand). It is also closed: nonce-words and recent loans in *a, u,* and *o* always belong in Class II. For the same reasons, the default is Class III for monosyllabic stems in neutral vowels.

Thus, the stems of Class I and Class IV must be marked by some diacritic. As we shall see shortly, there are reasons to suppose that this diacritic is the same for Class I and Class IV — let us call it +ML. For the time being, it is sufficient to note that each 'regular' vowel has a +ML counterpart, which is in Class IV if the vowel is front, and in Class I if it is neutral or back. The overwhelming majority of exceptional vowels are +ML. However, we find a few neutral vowel stems in Class II as well. For example, the triple *híd, víz, zsír* shows that we must have two kinds of exceptional *í*-s. Unlike +ML stems, which can contain every kind of surface vowel, –ML exceptional vowels are restricted to *í* and *é*. They appear only in a handful of monomorphemic stems and will require some other form of diacritical marking.

The reader familiar with earlier treatments of Hungarian vowel harmony might wonder how a phenomenon involving thousands of stems and the most common suffixes like the accusative or the plural could have been ignored. I believe that this is because most of the literature on this topic is based on

[20] This form has a decidedly *vidéki* 'rural' flavor for Budapest speakers, but is, perhaps, acceptable in some dialects. Note that different dialects can have different vowel inventories.

Vágó 1975, who addressed the problem posed by exceptional stems outside the framework of the investigation of vowel harmony. I will argue that Vágó's treatment was mistaken, and in fact the exceptional stems are fully integrated with the rest of the harmonic system. (In the following critical discussion, all page number references are to Vágó 1980, henceforth SPH.)

Although SPH starts with a discussion of vowel harmony (pp. 1-30), we do not learn of the existence of Class I stems until ch. 4.3 (p. 110), where "lowering stems" are discussed together with other (harmonically regular) classes of exceptional stems. On p. 111 we learn that lowering stems are "historically old, unproductive, and constitute a closed set — a set that is large, to be sure". I will argue that this characterization of Class I and Class IV is only partially correct: although they are closed as far as non-derived stems go (as can be seen from the fact that all recent loans are in Class II, III or V), they are open in the sense that *every* non-derived, non-neutral (and non-vacillating) Hungarian stem has derived forms that belong in Class I or Class IV.

To see this, let me recapitulate the defining property of lowering stems: a back stem is lowering (Class I) if and only if it selects the *a*-alternant of quaternary suffixes, and a front stem is lowering (Class IV) if and only if it selects the *e*-alternant of quaternary suffixes and the *ö*-alternant of ternary suffixes. Given an arbitrary stem of Class II or V, such as *rum* 'rum' or *'tök'* 'pumpkin', both the possessive and the accusative forms will show the regular quaternary *o* or *ö*. But the possessive forms themselves will belong to Class I (or Class IV), because they take the *a* (or *e*) alternant of a following quaternary suffix:

(18)

stem	1SG POS	ACC	1SG POS+ACC	
rum	rumom	rumot	rumomat	*rumomot
tök	tököm	tököt	tökömet	*tökömöt, *tökömhez

The accusative forms cannot be tested because the accusative suffix is absolute word-final, but almost all other suffixes have the same effect of turning regular stems into lowering stems. In the verbal paradigm, the past, imperative, and conditional markers all show this effect (as we shall see in 4.1) and, in the nominal paradigm the 1st and 2nd singular possessive suffixes also create lowering forms. Suffixes of this sort are not restricted to inflectional morphology either. For instance, the names of the villages *Rum* and *Tök* are declined just as the corresponding common nouns but, being place names, they permit the addition of the denominal adjective-forming suffix *-i* 'characterized by the location', which derives *rumi, töki* 'inhabitant of Rum, Tök', which in the plural becomes *rumiak, tökiek* 'inhabitants of Rum,Tök' rather than **rumiok, *tökiök*.

The operation of such suffixes shows that Class I stands in the same relationship to Class II as Class IV to Class V. Let us see how SPH captures this generalization. Vágó derives the *a*-alternant of quaternary suffixes by a minor lowering rule (MIN-LOW, p. 111), which lowers and unrounds the quaternary vowel: this rule is triggered by the diacritic +ML, which marks all lowering stems. In the description of MIN-LOW, Vágó mentions only the fact that it takes *o* to *a*, but it is clear from the formulation of the rule and from the discussion of the Class IV stem *föld* 'earth' on p. 112 that MIN-

LOW also takes *ö* to *e*.

By positing a single rule for front and back lowering stems, Vágó correctly captures the fact that the irregularity of Class IV stems has the same source as the irregularity of Class I stems, namely arbitrary lexical marking. This approach predicts that the +ML class of stems can contain front and back vowels alike and thereby leaves only the *zsír*-type (surface) neutral vowels as truly exceptional. In fact, the distribution of abstract vowels strongly supports this conclusion. As discussed above, (surface) neutral vowels can show three kinds of behavior (Class I, II, or III), while all other vowels show only two kinds of behavior (Class I or Class II for back vowels, Class IV or Class V for front vowels).

However, the technical details of Vágó's system do not quite work. Ternary vowels are treated as underlying *o* (p. 18), which is turned into *ö* by the rule(s) of binary harmony, and gets subsequently unrounded by the rule of Rounding Harmony (p. 19) if the preceding vowel is unrounded. Quaternary vowels are introduced by a rule of *o*-Epenthesis (p. 63, p. 109), which feeds MIN-LOW. This correctly derives the quaternary forms. The problem is that MIN-LOW necessarily overapplies to ternary suffixes after lowering stems: because the underlying representation of the ternary and the quaternary vowel is the same, we necessarily derive the incorrect **hölgyhez* along with the correct *hölgyet*.

Another problem is that Vágó needs two distinct epenthesis rules, one for the suffix adjacent to the stem (*o*-Epenthesis), and and one for subsequent suffixes (*a*-Epenthesis) (p. 110) to describe the fact that the vowel of quaternary suffixes shows up as binary *a/e* after certain suffixes, as exemplified in (11) above. These epenthesis rules are reproduced below:

$\underline{\text{a}}$-epenthesis

$$\emptyset \rightarrow \text{a}/C \begin{bmatrix} + \\ -\text{stem} \end{bmatrix} \langle\overline{\text{-VERB}}\rangle_{\text{a}}\text{C}\langle\#\rangle_{\text{b}}$$

Condition: if $\underline{\text{a}}$, then $\underline{\text{b}}$.

$\underline{\text{o}}$-epenthesis

$$\emptyset \rightarrow \text{o}/C + \langle\overline{\text{+VERB}}\rangle_{\text{a}}\text{C} \begin{Bmatrix} + \\ \text{C} \end{Bmatrix} \langle\#\rangle_{\text{b}}$$

Condition: if $\underline{\text{a}}$, then $\underline{\text{b}}$.

As can be seen, these are very complex rules: they involve angled brackets (with Boolean conditions), reference to lexical category (a feature ⟨±VERB⟩), and reference to the fact that a certain +-boundary does *not* follow a stem. Moreover, the two rules cannot be collapsed even with the powerful abbreviatory devices of SPE, because of the dissimilar environments and the ordering relation that obtains between them.

It is highly questionable whether angled brackets and Boolean conditions are permitted by phonological theory at all. But even if they are, the rules fail to capture the generalization that the relationship between the the quaternary vowel of a suffix, when following the stem, and the binary vowel of the same suffix when following another suffix can be subsumed under the rule of MIN-LOW, simply by marking the first suffix as 'lowering' (i.e. bearing ML in the lexicon).

The *a*-Epenthesis rule makes the incorrect prediction that a quaternary vowel must show up as binary whenever it is separated from the stem by another suffix. The following counterexample, which demonstrates the problem with this approach, comes from the numeral system. The suffix *-ad/ed/od/öd* 'th' forms fractions when attached to a numeral base (or any form such as variable names that can be interpreted as having numerical value). As the second column in (19) below shows, *-ad/ed/od/öd* is a genuine quaternary suffix, which takes the same alternant as the accusative (third column) or any other quaternary suffix[21] for every stem. However, the accusative form of the stems formed by suffixing *-ad/ed/od/öd* never shows up with *-at* (or with *-et* and *höz*):

(19)	stem	FRAC	ACC	FRAC+ACC	gloss
	öt	ötöd	ötöt	ötödöt	five
	hat	hatod	hatot	hatodot	six
	hét	heted	hetet	hetedet	seven
	nyolc	nyolcad	nyolcat	nyolcadot	eight
	iksz	ikszed	ikszet	ikszedet	x
	ipszilon	ipszilonod	ipszilonok	ipszilonodot	y
	sok	sokad	sokat	sokadot	much

It is easy to explain this paradigm as I do by not marking the fraction-forming *-ad/ed/od/öd* with ML. The SPH system, however, requires an extra rule which turns the epenthetic *a* into *o*. I would like to emphasize that the issue is *not* whether we have an epenthesis or an elision analysis of quaternary vowels, but rather the manner in which the quality of the vowel (epenthetic or otherwise) is determined by vowel harmony. To make this point even more clear, consider the second person singular possessive *-ad/ed/od/öd*, which gives us minimal pairs like *nyolcadat* 'eight.POS2SG.ACC' vs. *nyolcadot* 'eight.FRAC.ACC'; or *ötödet* 'five.POS2SG.ACC' vs. *ötödöt* 'five.FRAC.ACC'. The accusative suffix shows binary alternation after POS2SG, and ternary alternation after FRAC: this follows without further stipulations if we mark POS2SG by ML, but leave FRAC unmarked.

These minimal pairs show that the choice between *o*-Epenthesis and *a*-Epenthesis depends on the identity of the preceding morpheme and not on the position (stem-adjacency) of a suffix. Noun/adjective pairs (Tompa 1957) provide another, perhaps even more striking, source of minimal pairs: *akadémikusok* 'academician.PL' vs. *akadémikusak* 'academic.PL'. In the light of these and similar pairs, even if we take epenthesis for granted, we should eliminate the rule of *a*-Epenthesis in favor of a single rule that inserts the quaternary vowel, and derive the observed binary or ternary alternation by making this vowel subject to MIN-LOW.

Thus, even if we are interested only in front/back alternations such as the *a/e* alternation of the vowel preceding the *t* of the accusative (when it appears adjacent to the stem or in other positions), in order to capture the full pattern of *a/e* alternations we must come to grips with the fact that binary, ternary, and

[21] Since in the accusative of *ipszilon* the vowel is elided, the plural is given instead.

quaternary alternations are deeply intertwined in Hungarian. In addition to the effects of MIN-LOW, this can also be seen from the fact that in the same structural position, such as the noun-final position for case markers, we can find unary (*-ként*), binary (*-nak/nek*), ternary (*-hoz/hez/höz*), and quaternary (*-at/et/ot/öt*) suffixes. Moreover, suffixes can also change their arity as a result of morphological processes, as will be shown through an analysis of the possessive paradigm in 4.2.

To sum up, Vágó (1975, 1976) fails to treat harmonic alternations in a unified manner. The subsequent literature, including Clements 1976, Phelps 1978, Jensen 1978, Vágó 1980, Ringen 1980 van der Hulst 1985, Goldsmith 1985, and Farkas and Beddor 1987 has continued to concentrate on binary harmony, and in particular on the treatment of neutral vowels and vacillating stems. This is somewhat unfortunate, inasmuch as the generalizations these systems intend to capture are not equally clear.[22]

The most important generalization established so far is the unity of 'minor lowering' processes. Since these are lexically determined, we need a diacritic like +ML to mark those stems and suffixes that trigger such processes. It is a major achievement of autosegmental phonology that the lexical marking in such cases need not involve *ad hoc* rule exception features, but can be chosen from a restricted set of diacritics, namely that of floating features. For minor lowering, the changes involved are *o* to *a* and *ö* to *e*. In the feature system developed here both of these consists of the deletion of a feature U. Therefore, the theory forces upon us the feature –U as the diacritic marking 'minor lowering' stems.

The most straightforward analysis of binary and ternary harmony, which will have to be supplemented by other rules later, is to spread the feature values for I and U onto the suffixes. For quaternary harmony, we have to adjust the underlying representation of lowering stems so that the spreading of U is blocked for them. This is achieved by marking the stems in Class I and Class IV by a floating –U. Thus, the cornerstone of the analysis is the following pair of spreading rules:

(20A)

$$\begin{array}{lll} \mathrm{I} & & \\ | & \ddots & \\ \mathrm{V} & \mathrm{C}_0 & \mathrm{V} \end{array}$$

(20B)

$$\begin{array}{lll} \mathrm{U} & & \\ | & \ddots & \\ \mathrm{V} & \mathrm{C}_0 & \mathrm{V} \end{array}$$

Supposing that the vowel of quaternary suffixes is specified only for A, the rules in (20) will derive e.g. *tökök* from *tök* simply by spreading the I and U features of the stem vowel. This solution can readily be extended to ternary suffixes by supposing that their underlying representation contains a floating U which can dock only if the I feature of the stem did not spread.

[22] As a representative example let me discuss here the following generalization (due to Esztergár (1971), and quoted in Vágó 1980:111): lowering stems always have an epenthetic vowel before the *-t* of the accusative. SPH (fn. 18 on p. 134) gives a clear counterexample: *férfi* 'man' accusative *férfit (*férfiat)*. As can be seen from the plural *férfiak*, this is clearly a lowering stem. Papp 1968 lists more than fifty lowering stems as taking *-t* in the accusative, including *hegyoldal* 'hillside', *magtár* 'granary', *szökőár* 'bore(tide)'. In addition, there is a systematic class of exceptions: denominal adjectives, formed from back vowel stems with the suffix *-i* discussed above, are all lowering (because *i* is lowering). Some of these, e.g. *szarházi,* are epithets generally used as nouns. In addition, there are hundreds of lowering stems in which the quaternary vowel of the accusative is optional.

The most important problem with this simple solution is that it does not capture the exceptional status of lowering stems. Without additional 'cleanup' rules, we would have to mark the elements of Class II, rather than those of Class I, in the lexicon, in spite of the fact that Class II is productive, and Class I is closed.[23]

In order to deal with this problem, we must posit an independent source for the feature U that appears in the suffix of Class II stems. The field of computer science provides many examples of recently borrowed Class II stems such as *fájl/fájlok* 'file/PL', *ram/ramot* 'random access memory/ACC', which make clear that there must be a U in the representation of the quaternary suffixes themselves.

Adopting a proposal of Halle and Vergnaud (1982), I will take this U to be specified in the phonemic core.[24] For those features like A in Hungarian that do not harmonize, there is no good reason to establish a separate tier. I suppose that the unmarked place for a segmental feature is on the segmental tier, and we autosegmentalize a feature only if there is some evidence for this. However, the segmental core remains a possible location even if the feature has been lifted to a separate tier — the present analysis will make use of this option only in the case of exceptional elements.

In sum, the underlying representation of quaternary suffixes contains the same features, namely A and U, as that of the ternary suffixes, the difference being that, for the quaternary case, the U is in the core, while, in the ternary case, it is floating. This 'geometrical' difference will surface only after exceptional stems and after some suffixes. Before turning to these, let me show first how the non-exceptional forms are derived.

In Class II, *zsír* and *cél* are exceptional (there are less then 10 monomorphemic *i* and *e* stems there) — the Class II pattern is regular only for stems in *a, o,* and *u*. Since these do not contain the feature I, (20A) is inoperative. Whether (20B) actually spreads the feature U in the case of stems in *u* and *o* cannot be decided on the basis of these forms, since both ternary and quaternary suffixes have an underlying U (albeit on different tiers).

I-spread gives the right result in Class III, provided that the U floating over *hoz/hez/höz* is stopped from linking up. This is achieved by a rule of floating U deletion:

```
(21)  I-tier              /   I
                         /    |
    CV-tier             /     VCoV
                       /
      U-tier   U-->0  /         ____
```

[23]The internal logic of the proposed system suggests that at an earlier stage Class I must have been the regular one. In 4.2 I will argue that synchronically the quaternary vowel is part of the suffix. But at the historical stage when Class I was regular, the quaternary vowel might very well have been epenthetic. If this was indeed the case, the phenomenon of Low Vowel Lengthening (discussed from a synchronic perspective in 4.2) can be adduced as evidence in favor of the treatment of long *a* and *e* as geminates, because under such a treatment Low Vowel Lengthening would arise naturally as a by-product of generalizing the epenthesis rule from consonant-final stems to any stem.

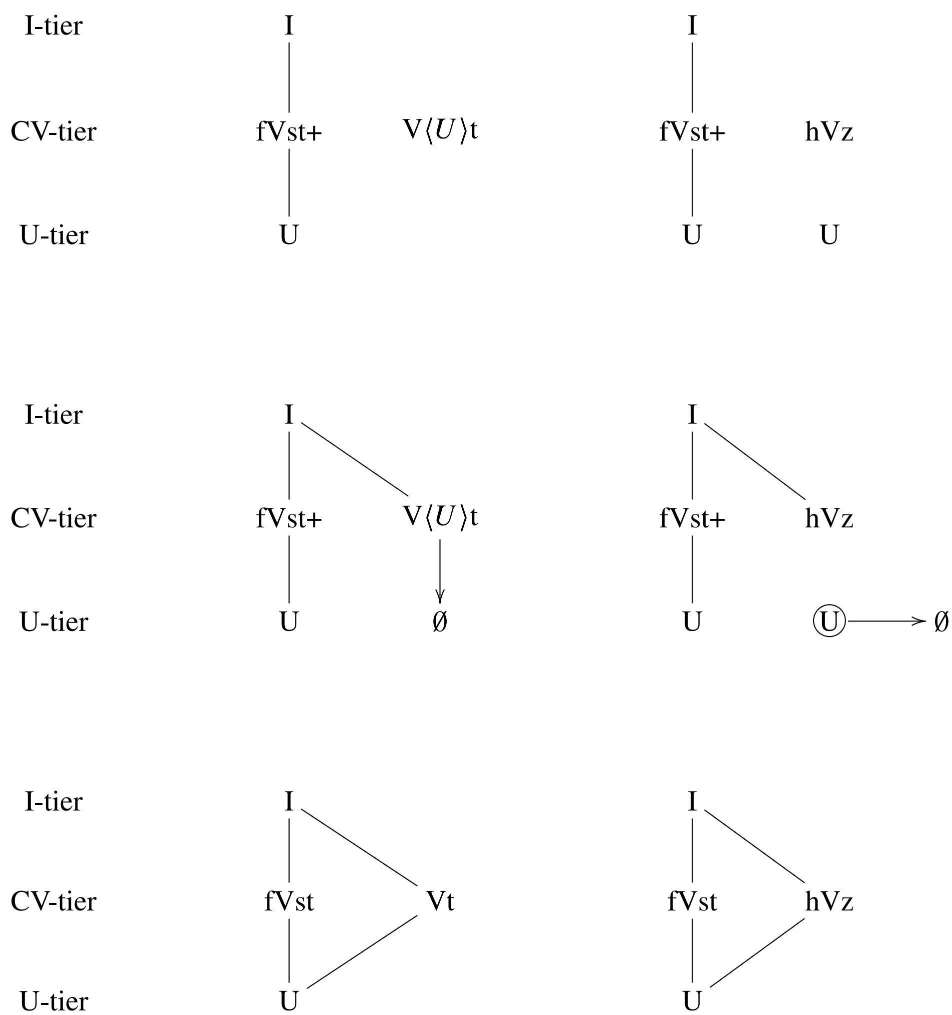
[24]The root tier is omitted from the display, and the segmental core is identified with the CV tier. Nothing hinges on these simplifying assumptions.

In the same environment, the core U of *ak/ek/ok/ök* must also be deleted:

```
(22) I-tier               /  I
                         /   |
    CV-tier   V<U>-->V /    VCo___
```

Finally, in Class V, I-spread (20A), U-deletion (21-22), and U-spread (20B), in this order, give the right result.[25] This is illustrated in (23) by the derivation of the forms *füstöt, füsthöz*. Notice that the derivation is essentially the same for forms like *tököt, tökhöz*, as the presence or absence of A-specification plays no role in any of the rules. Although the suffixes in question are specified for A, the high-mid parallelism makes it possible to omit the As from the display altogether.

(23)



The lack of negative specification in the rules means that we can interpret U, I, and A as *simplex* features representing privative oppositions.[26] With the rules (21-22), the representations of non-exceptional

[25] Roughly speaking, (21) and (22) have the same effect as the 'Rounding Harmony' rule of SPH (ch. 1.7).

[26] Except of course for exception features which has to be mentioned by the rules explicitly. I will return to this question in 2.6, where the objections of Farkas and Beddor 1987 to treating I as privative in Hungarian are discussed.

items (*a á o ó u ú* in Class II, *i í e é* in Class III, and *ö ő ü ű* in class V) were kept simple. The exceptional items are treated as follows. In Class IV, we have to stop the core U of quaternary suffixes from taking effect, but we do *not* want to derive the *e*-alternant for ternary suffixes in the manner of SPH. To do this, let us suppose that the diacritic –U triggers the deletion of core Us:

```
(24)   CV-tier      V<U>-->V   / VCo__
                              /
        U-tier               /   -U
```

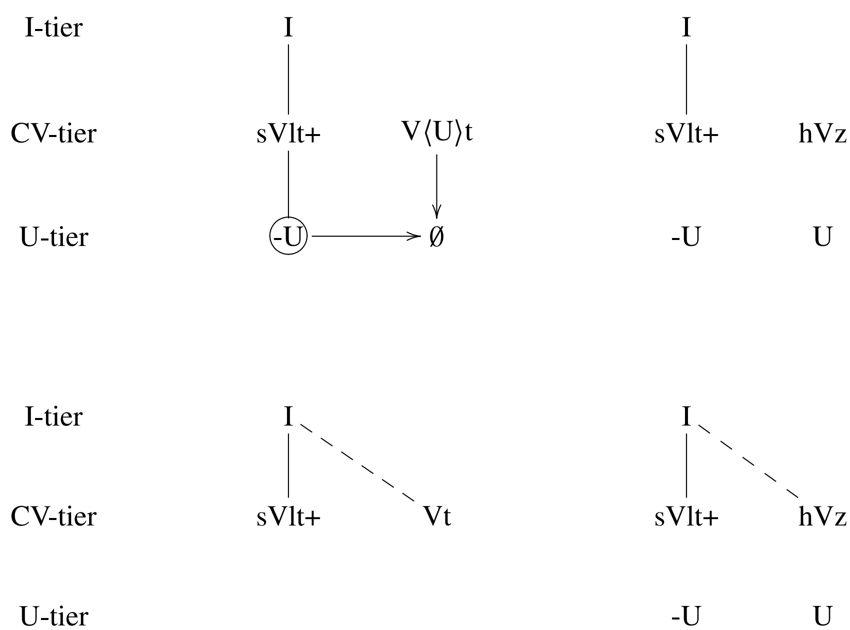
In other words, for a negative feature, 'spreading' to a positively specified core amounts to deleting the offending feature from the core. I will assume that the negative feature also disappears. This can be formulated as a general rule subsuming (24):

```
(25) Core/Neg  Annihilation

    CV-tier       V<+X>-->V   / VCo__
                             /
     X-tier                 /   -X
```

This rule, ordered before the others, takes care of Class IV. Let us show this on the stem *sült*.

(26)



At the top of (26) we find the underlying representations to which Core-Neg Annihilation (25) applies to yield the intermediate representations found in the middle section. At this stage, only I-spread (20A) will apply[27] – again, the derivation proceeds independent of the contents of the A-tier.

[27] In particular, there is no U-deletion as the -U of *sült* is absent from the structural description of (21) — cf. the previous footnote.

The remaining exceptional elements are treated as follows. The *i* of the *híd*-type words is specified for I in the core (therefore it will not spread); in addition, it will have a floating –U, which will derive the correct *hidat (*hidot)*. The *í* of the *zsír*-type words is also specified in the core *(*zsírnek)*, but will have no other exceptional property; therefore, we derive the correct *zsírok (*zsírak)*. The exceptional *é* of *héj* has I in the core, and floating –U, giving us the correct *héjak, héjhoz*. The only stem in *e* or *é* paralleling the behavior of *zsír* is *cél* 'goal' which have I in the core but no other exceptional marking.

Every other exceptional element will be marked by a floating –U : the U specification (where present, e.g in *lyuk, hold, hölgy*) must be relegated to the core (CV tier). This simple and unified treatment of back stems in Class I is a highly desirable result, given the fact that Class I contains more than a thousand monomorphemic noun stems (and several thousand compounds) of this kind, while the remaining exceptional types are only sporadically represented.

Moreover, the use of the feature –U unifies the treatment of the exceptional classes. The only vowels we do not find in Class I are *ü* and *ö* – when these are marked by –U, they belong in Class IV. In addition, the tridirectional feature system captures the parallelism between the set {*u ü i*} of high vowels and the set {*o ö e*} of mid vowels. Since no rule makes reference to the feature A, we expect to find both +A and –A elements in every class. As can be seen from the data presented in (17), this expectation is fulfilled not only by the regular vowels, but by the irregular vowels as well.

The analysis can be extended to capture the behavior of suffix-combinations with no epenthesis rule(s), simply by marking most quaternary suffixes (but not the fraction-forming *ad/ed/od/öd*) with a floating –U. By doing this, Core-Neg Annihilation will do not only the work of MIN-LOW, but also that of *a*-Epenthesis. Forms such as those in (19) can be derived without further complications.

The adjective-forming denominal suffix *as/es/os/ös* 'having (to do with)', which must also be marked this way, offers a particularly good way of testing the proposed mechanism. In isolation, a form containing this suffix can exhibit behavior characteristic of Class II forms:

(27) *ház* 'house' *házas/házasok* 'married/PL'
**házasak*

However, if the same form appears in a non-lexicalized meaning, as in

(28) A kertes könyveket jobbra tedd, a házasakat pedig balra!
'Put those books (about) gardens to the right side,
and those (about) houses to the left'

its plural will be *házasak*.[28] This behavior can be explained only if we suppose that the exceptional marking introduced by *-as* is lost in the lexicalized form *házas*, but can be present if the form is derived anew, as required in (28). The derivations are given in (29) and (30)

[28] In ECH, *%házasokat* is tolerated, but hardly ever produced, in (28) and similar contexts.

(29A)

CV-tier			hVVz+	V⟨*U*⟩s
			|	↓
U-tier	Ø ←	-U	-U	Ø

(29B)

CV-tier	hVVzas
U-tier	-U

(29C)

CV-tier	hVVzas	V⟨U⟩k
U-tier	-U	-U

The only rule that applies in the derivation is Core/Neg Annihilation, which derives (29B) as the representation of the form *házas* from the underlying (29A). If the exceptional marking contributed by the suffix *as/es/os/ös* is retained, we derive (29C) on the next cycle. If the exceptional marking is lost, so that we start with (30A), we derive (30B) as the result of plural suffixation. Finally, since the plural form is not in the lexicon, the exceptional marking contributed by the plural suffix cannot be lost, so in the last cycle we derive (30C) in both cases.

(30A)

CV-tier	hVVzas
U-tier	

(30B)

CV-tier	hVVzasV⟨U⟩k
U-tier	-U

(30C)

CV-tier	hVVzasV⟨U⟩k	V⟨U⟩t
U-tier	-U	-U

This solution readily extends to cases like *aranyak* 'gold pieces' vs. *Aranyok* 'books of Arany', or *pirosak* 'red(adj)PL' vs. *pirosok* 'red(n)PL', where we have reason to suppose that a derived form re-entered the lexicon. In the case of *arany*, the common noun 'gold' is obviously the primary form. The proper name *Arany* retains the exceptional marking of this primary form (*Aranyat/*Aranyot nem szerették a kortársai* 'Arany was not liked by his contemporaries'), but when the common noun 'book of Arany' is formed from the proper name, the exceptional marking is lost. Similarly, the primary color term is obviously *pír* (which is Class II and thus does not have the –U exceptional marking), and the secondary color term *piros* is derived by the suffix *as/es/os/ös* which, as discussed above, adds the floating –U. The noun *piros* 'the color red' is formed from the adjective, and enters the lexicon with a loss of exceptional marking.

2.4 Syllable structure

The classical "mirror rule" of traditional grammars subsumes three, logically independent observations:

(31) If PQ is a possible syllable onset (P, Q arbitrary consonants), then QP is not.

(32) If PQ is a possible onset, then QP is a possible coda, and, conversely, if RS is a possible coda, then SR is a possible onset.

(33) If PQ is a possible coda, then QP is not.

Of course, if (32) holds, (31) and (33) are equivalent – but there might well be languages where (32) turns out to be false, but the other two statements are true. In fact, every language where consonant clusters are disallowed as codas but permitted as onsets is a counterexample to (32), and the same holds for those languages that allow complex codas but do not allow complex onsets. Before turning to the investigation of the mirror rule in Hungarian, let me add a further clause, (cf. Clements and Keyser 1983:47-48) which I will call Hjelmslev's Law:

(34) If PQR is a possible onset, then so are PQ and QR, and similarly for codas.

(35) If PQ and QR are possible onsets, then so is PQR, and similarly for codas.

This last requirement (the converse of Hjelmslev's Law) and (34) have the effect of extending (31) and (33) to arbitrarily long consonant clusters: in Hungarian, the longest cluster that we will encounter contains three consonants. If the notion "Sonority Hierarchy" (in the sense of Jespersen 1897-99) has any validity, then the statements (31-35) will follow automatically. Suppose that phonemes are partitioned into sonority classes, and the classes are ordered linearly among themselves in such a manner that vowels are at the high end of this ordering (called the sonority hierarchy). Now if we define a well-formed syllable as one with sonority (strictly) increasing from the onset to the nucleus, and (strictly) decreasing thereafter, (31-35) will necessarily hold.

The traditional view, which was based on non-borrowed (pre-16c) material, is that Hungarian has no syllable-initial consonant clusters. (The idea that this could be retained in synchronic descriptions has been criticized at length by Siptár 1980b.) The following table gives an overview of the two-member initial and final clusters attested.

(36)

	s	sz	f	p	k	c	t	cs	ty	h	v	b	g	d	gy	zs	z	ny	m	l	n	j	r	
s	2	4		3	3		5	5			5	4	4	4	4	4	4		1	3	3	2	3	s
sz	4	2	3	3	3	5	5		3		5	4	1	4	4	4	4	2	3	3	3	2	2	sz
f		2	2	2			1				4	4	4	4	4	4	4		2	3		3	3	f
p		1	1	2			1			1	4	4	4	4	4	4	4		2	3	5	2	3	p
k		3			2	2	2	2	2		5	4	4	4	4	4	4			3	3	2	3	k
c		4	2	2		2					5									2	2	2	2	c
t	2	6	2	2	2		2		4	2	5	4	4	4	4	4	4	6	2	2	2	2	3	t
cs	2				2			2											2	2	2	2	2	cs
ty							4		2			4	4	4	4	4	4	2			4		2	ty
h										2								2	1			3	1	h
v	4	4	4	4	4	4	4	4	4	4				2				2	2	3	2	2	2	v
b	4	4	4	4	4	4	4	4	4	4		2				2			2	1	4		3	b
g	4	4	4	4	4	4	4	4	4	4	1		2							3	3	2	3	g
d	4	4	4	4	4	4	4	2	4	4	2	2		2	6	6	3	6	2	2	2	2	3	d
gy	4	4	4	4	4	4	4	4	4	4				4	2	2	2	2		2	4		2	gy
zs	4	4	4	4	4	4	4	4	4	4				4			4		2			2	2	zs
z	4	4	4	4	4	4	4	4	4	4				4		4	2		2	1	2	2	3	z
ny							4											2			4		2	ny
m																			2	2	5	2	2	m
l																				2	2	2	2	l
n									4				1		4						2	2	2	n
j			2	2	2						2	2	2						2			2	2	j
r																							2	r
	s	sz	f	p	k	c	t	cs	ty	h	v	b	g	d	gy	zs	z	ny	m	l	n	j	r	

1 at the intersection of row x and column y means that xy is an attested onset; 2 means that yx is an attested coda; 3 = 1 + 2 means that xy is an attested onset and yx is an attested coda; 4 means that whenever a combination like xy and/or yx arises in compounding, it is subject to consonant sandhi; 5 = 1 + 4 means that yx is subject to sandhi but xy is attested as onset; and 6 = 2 + 4 means that xy is subject to sandhi but yx is attested as coda.

If (32) were a universal law, there could be no ‘1’ or ‘2’ entries in (36) at all. However, the high number of exceptional entries is counterbalanced by the ‘low quality’ of the exceptional words. First, a few interjections, namely *pfuj* ‘phooey’, *phű* ‘phew’, *hm* ‘id’, and *hja* ’well’ are included in (36), because they were mentioned in the dictionaries (Bakos 1974, Juhász et al 1982, Országh 1977, Papp 1968) or papers (Abondolo 1988, Kassai 1981, Siptár 1979, 1980a) I have included in my corpus. But in a larger corpus it would be possible to find other interjections, such as *grmbh, hmpf,* or *brrr*. The onomatopoeic nature of these expressions makes it hard to exclude *any* combination categorically. Therefore I am inclined to discard these elements. With the exception of *bolyh* ‘tomentum’, which will have no onset counterpart, this move will increase the symmetry of the system.

Second, certain entries such as *nganaszán* ‘name of Uralic tribe’, *ptózis* ‘ptosys’, *mnemotechnika* ‘mnemonics’, *ftálsav* ‘ftalic acid’, *szgrafitto* ‘wall painting’, and perhaps a few others, can hardly (if at all) be pronounced by ECH speakers in the manner suggested by the orthography. This is not to say that all ‘foreign’ words should be eliminated from the corpus,[29] but surely the line must be drawn *somewhere.* The items deemed exceptional above are likely to manifest exceptional behavior in simple reading and repetition tasks, while most ‘learned’ or foreign words such as *gnóm* ‘gnome’ or *pszichológia* ‘psychology’ are likely to pattern with native words. In fact, certain foreign words, such as *szféra* ‘sphere’ are likely to fare better than certain native words (especially proper names as *Szakcs*).

Third, the entries in the diagonal reflect the fact that geminate codas can be found with almost every consonant, but geminate onsets are absent. Examples are: *juss* ‘share’, *hossz* ‘length’, *muff* ‘id’, *épp* ‘just’, *sikk* ‘fashion’, *vicc* ‘joke’, *ott* ‘there’, *priccs* ‘bunk’, *pötty* ‘dot’, *pech* [hh] ‘bad luck’, *alább* ‘below’, *agg* ‘old’, *haddelhadd* ‘rumpus’, *meggy* ‘sour.cherry’, *nézz* ‘look 2nd.sg.imp.indef’, *könny* ‘tear’, *bumm* ‘boom’, *toll* ‘pen’, *kinn* ‘out’, *falj* [jj] ‘devour 2nd.sg.imp.indef’, *orr* ‘nose’. The examples given are always the ‘best’ in the sense that proper names are avoided if common nouns can be found, monomorphemic or at least uninflected words are preferred to overtly inflected forms, surface forms having the same underlying representation are preferred to those coming from different URs, and finally native words are preferred to ‘foreign’ or ‘learned’ words. This means that in any case (e.g. the coda *dd*) the reader can infer from the given example (the compound *haddelhadd*) that no word can be found in its class which has lesser morphemic complexity but is not uninflected (as e.g. *add* ‘give 2nd.sg.imp.def’) or is not a proper name (as e.g. *Fadd*).

Conditions on well-formedness are frequently ‘enforced’ by a conspiracy of rules which modify the offending combinations that arise in the course of the derivations (cf. e.g. Kisseberth 1970). The lack of geminate onsets in Hungarian appears to be a purely ‘static’ well-formedness condition in the sense that no rules are necessary to enforce it: as there are no single-consonant prefixes in Hungarian, the disallowed combinations simply do not arise. This observation indicates that even static conditions can have priority over the generalization expressed in (32).

Fourth, dynamic well-formedness conditions or, equivalently, the phonological rules enforcing these, can also distort the picture. For instance, the onsets *sv, szv, kv, cv*, and *tv* (as in *svéd* ‘Swedish’, *szvit* ‘suite’, *kvarc* ‘quartz’, *cvikker* ‘pince-nez’, and *tviszt* ‘twist’) have no coda counterparts, since devoicing (see (39) below) would turn these into *fs, fsz, fk, fc*, and *ft*. Similarly, the onsets *szt, szc, scs, st*, and *pn* (as in *sztár* ‘movie star’, *szcenárió* ‘screenplay’, *scsi* ‘kind of soup’, *steril* ‘sterile’, and *pneumatikus* ‘pneumatic’) correspond to codas that trigger rules of affrication and assimilation.

Taking all these factors into consideration, there remain only six ‘1’ entries in (36), and half of these are caused by onsets appearing in a single word (*zlotyi* ‘Polish currency’, *hradzsin* ‘castle in

[29] On the contrary, foreign words tend to fill in the accidental gaps in the system. For instance, the epenthetic *t* of *ptrücsök* etc. discussed in 2.2 gives rise to an onset *pt* by Hjelmslev’s Law, and this gap is filled only by foreign words such as *ptózis* or *Ptolemaiosz* ‘Ptolemy’.

Prague', *gvárdián* 'Father Superior'). The rest might be attributed to accidental gaps in the coda system (*blúz* 'shirt', *smaragd* 'emerald'), with the exception of **szp* codas (cf. *pszichológia* 'psychology'), the absence of which appears to be systematic in the light of (33), given the coda *psz* (e.g. in *gipsz,* 'gypsum').

Word initial *szp* is also possible (*szpiker* 'announcer'), and similarly with *ksz* we have *kszilofón* 'marimba', *maszk* 'mask', *szkiff* 'skiff' and *keksz* 'biscuit'. Aside from a couple of proper names (*Szakcs* and *Recsk*), and some inflected forms, the pairs listed above, and the type *liszt* 'flour' vs. *sztár* 'movie star' constitute the only counterexamples to (31) and (33) in Hungarian. Selkirk (1984) deals with the same problem in English by treating [sp,st,sk] clusters as affricates, but Clements (pc) notes that

(i) these clusters do not pattern with 'true' affricates (**črV, ǰrV*, as opposed to *sprV*)

(ii) unlike true affricates, these clusters are easily broken up by speech errors

(iii) Selkirk's theory leaves the lack of syllable-final

```
VC              VCC
 /\      vs.     ||
s  p             sp
```

contrasts unexplained. Since the counterarguments based on distribution and speech errors are equally valid in the case of Hungarian, I will not adopt Selkirk's solution here. But as these cases constitute the only serious counterexample to (31) and (33) in a number of languages (e.g. Danish *fisk* 'fish' vs. *fiks* 'fix'), Selkirk is obviously right in trying to explain them by some special principle that leaves the larger generalization intact.

To sum up what we have so far, the mirror image of the onset structure of Hungarian syllables is, by and large, properly included in the set of attested coda combinations that make up the coda structure. As Algeo (1978) notes, it is possible to make the system look a great deal more elegant by being more selective with the data we include. For instance, the exclusion of inflected forms would go a long way in eliminating the 'irregular' entries that appear under the diagonal. The 2nd.sg.imp.indef suffix *j* is the only source of codas such as *'döfj'* 'pierce!', *lopj* 'steal!', *bukj* 'fall!', *szivj* 'suck!', *dobj* 'throw!', or *vágj* 'cut!', and the 2nd.sg.imp.def suffix *d* is the only source of codas such as *dobd* 'throw!', *szivd* 'suck!', or *tanitsd* [čd] 'teach!'.

Of the remaining 10 entries under the diagonal, two (*NOSZF* and *MAFC*) could be excluded on the basis that they are acronyms (these were the only acronyms included in the corpus, because the 'spelling pronunciations' **enóeszef* and **emaefcé* are never heard), and two (*Apc* and *Szakcs*) because they are proper names. One (*borscs* 'kind of soup') is the single example of its type, and another one (*szomj* 'thirst') can be argued to end in [h] rather than [j] on the surface. But no matter how hard one tries, there seems to be no way to exclude *füst* 'smoke', *szaft* 'gravy', *recept* 'prescription' or *akt* 'nude'.

As the 'hard' counterexamples all end in *t*, one might try to reshuffle the matrix in (36) so that only 'easy' items appear under the diagonal. To see what is involved here, let us take a look at the high end of the hierarchy. The ordering $m\langle l\langle n\langle j\langle r$ appears to be extremely well motivated: the codas in *film* 'id', *slejm* 'phlegm', *farm* 'id', *ajánl* 'recommend', *fájl* 'file', *görl* 'girl', *kombájn* 'combine-harvester', *modern* 'id', and *férj* 'husband' are all decreasing in sonority. Moreover, every pair of decreasing sonority corresponds to an attested coda, with the exception of $n\rangle m$, but even this is attested (in the reversed order) as onset. Yet it is possible to rearrange the sequence (e.g. to $m\langle n\langle l\langle r\langle j$) by the data-manipulation techniques discussed above, and the reader is invited to try it. For those who prefer to work with a larger set of data, I list here the 'best' example of each onset and coda type not mentioned so far:

srác 'kid', *sors* 'fate', *mersz* 'daring', *francia* 'French', *turf* 'id', *próba* 'trial', *szörp* 'juice', *krém* 'cream', *park* 'id', *bérc* 'peak', *trágya* 'manure', *szirt* 'cliff', *tekercs* 'scroll', *korty* 'gulp', *érv* 'argument', *bróm* 'bromide', *szerb* 'Serbian', *gróf* 'peer', *burg* 'castle in Vienna', *drót* 'wire', *kard* 'sword', *tárgy* 'object', *törzs* 'tribe', *zri* 'rumpus', *borz* 'badger', *szárny* 'wing', *Majs* 'placename', *fájsz* 'hurt 2nd.sg.pres. indef', *fjord* 'id', *dölyf* 'haughtiness', *selyp* 'lisp', *sztrájk* 'strike', *Svájc* 'Switzerland', *sejt* 'cell', *ejts* [ejč] 'drop 2nd.sg.imp.indef', *ölyv* 'buzzard', *cajg* 'calico', *majd* 'then', *pajzs* 'shield', *rajz* 'drawing', *snájdig* 'neat', *pikáns* 'piquant', *sznob* 'snob', *pasziánsz* 'solitaire', *knock-out* [knokaut] 'id', *fánk* 'doughnut', *tánc* 'dance', *pont* 'dot', *kilincs* 'doorknob', *rokonszenv* 'sympathy', *gnóm* 'gnome', *rang* 'rank', *gond* 'worry', *pénz* 'money', *sláger* 'hit', *fals* 'out of tune', *szláv* 'Slavic', *félsz* 'fright', *flaska* 'bottle', *golf* 'id', *plakát* 'poster', *talp* 'sole', *klassz* 'groovy', *halk* 'silent', *polc* 'shelf', *bolt* 'shop', *kulcs* 'key', *vlach* 'id', *nyelv* 'tongue', *blúz* 'shirt', *glória* 'halo', *rivalg* 'whoop', *föld* 'earth', *völgy* 'valley', *smaragd* 'emerald', *szmötyi* 'sediment', *tömsz* 'stuff 2nd.sg.pres.indef', *tromf* 'retort', *kolomp* 'bell', *teremt* 'create', *teremts* [mč] 'create 2nd.sg.imp.indef', *hamv* 'ash', *domb* 'hill', *nyomd* 'push 2nd.sg.imp.def', *tömzs* 'lode', *nemz* 'beget', *dzeta* 'id', *gerezd* 'clove', *küzdj* [zj] 'fight 2nd.sg.imp.indef', *idösb* [žb] 'elder', *Pünkösd* [žd] 'whitsun', *esdj* [žj] 'beg 2nd.sg.imp.indef', *kedv* 'mood', *yacht* 'id', *sztyepp* 'prairie', *hagysz* [cs] 'let 2nd.sg.pres.indef', *Batyk* 'placename', *steril* 'sterile', *Detk* 'placename', *barack* 'peach', *skála* 'scale', *voks* 'vote' *spicli* 'informer', *taps* 'clap', *szpáhi* 'spahi', *gipsz* 'gypsum', *copf* 'pigtail', *Apc* 'placename', *szféra* 'sphere', *szivsz* [fs] 'suck 2nd.sg.pres. indef'.

The point of the exercise is that only a few gross statements about the sonority hierarchy appear to be incontrovertible. Voiceless consonants will precede the voiced ones, and obstruents will precede the resonants in every reasonable rearrangement of (36), but besides these (rather trivial) observations, little can be said with certainty.

With that, the question becomes the following: what can we possibly gain by employing a theoretical construct (the sonority hierarchy) if, on the one hand, it is next to impossible to model the facts (i.e. to arrange the consonants on a scale) by it in an unambiguous manner, and if, on the other hand, the predictions (i.e. 31-33) made by the theory do not really fit the data? My answer is based on the well-known facts that syllables are psychologically real units of speech production (cf. e.g. Kim 1971) and of speech perception (Savin and Bever 1970).

The sonority hierarchy makes it possible to factor out a large part of the linear precedence (LP, see Gazdar and Pullum 1982) information that must be encoded with every syllable node immediately dominating a number of timing units. In fact, no LP information has to be stored with C*V and VC* syllables conforming to the hierarchy. In syllables containing Cs on both sides of the V it is sufficient to store only the fact that a given consonant precedes or follows the vowel (so that *pit* will not be confused with *tip)* – of course, this will have to be stored with V-initial and V-final syllables as well, so as to know which is which. The consonants can be arranged among each other on the basis of sonority: the more sonorant a consonant, the closer it comes to the vowel.

This proposal can be implemented without recourse to an abstract scale if we take it into account that sonority can be expressed in terms of *features.* To quote Basboll (1973:132): "In fact, the claim is that the features of the "hierarchy" are distributed around the peak of the syllable, so that each feature may spread continuously over several segments in the way indicated in the hierarchy. This could be formulated so that 'one instance of' e.g. the feature ⟨+sonorant⟩ 'belongs to' several segments at the same time." In autosegmental terms this means that the timing units must be arranged around the vowel in such a manner that the features linked to them can undergo contour simplification maximally. For instance, in the monosyllable *brancs* 'gang', we have to store only the facts that *b* and *r* precede, and *n* and *cs* follow the vowel. The alternative ordering **bracsn* is excluded because the ⟨+son⟩ features of *n* and the vowel are not adjacent, and thus cannot be simplified. Similarly, the order **rbancs* can be excluded because the ⟨+son⟩ of *r* cannot be collapsed with that of the vowel, and the order **rbacsn* is excluded even more strongly, as it would require 3 instances of ⟨+son⟩ instead of the optimal 1.

From this perspective, the existence of isolated counterexamples is not really worrisome; with those, we will simply have to store more LP information. The mechanism outlined above acts as a default; extra information concerning the position of the features can override it. This means that it matters but little whether we have proper names, foreign, or learned words: it is quite conceivable that such items require a larger amount of storage in a system that works best with native words. Inflected forms, however, belong in a different class, at least if we suppose that these are not stored in the lexicon but are created 'on the fly'. In generating a form like *lopj* 'steal 2nd.sg.imp.indef' we know independently that the suffix *-j* will follow the stem *lop*, so the default mechanism need not be engaged at all. In other words, in the investigation of language-particular sonority facts or 'core syllables' the consonant clusters arising from affixation can safely be disregarded, since for these the LP information is derived from the linear precedence of the affixes, rather than from the sonority hierarchy.

2.5 Postlexical rules

While internal sandhi rules are in the lexicon, external sandhi must take place in the postlexical component of the grammar. Since cliticization rules can enlarge the lexical words (the resulting phonological words will still have a single accent), it is possible to have a third class of sandhi rules having word-internal effects but belonging to the postlexical component. Apparently, internal and external sandhi rules are highly similar in Hungarian. This, and the lack of detailed descriptions, makes it impossible to investigate this third class of rules here. The aim of this section is just to prepare the ground for such an investigation by outlining the most prominent sandhi rules, and by describing the effects of cliticization.

Cliticization phenomena are traditionally ignored in grammars of Hungarian. Vágó 1980:ix is typical in this respect: "...the initial syllable of a word is predictably stressed. This rule is merely assumed." Nevertheless, certain connectives like *ha* 'if' (or *is* 'too') are always proclitic (resp. enclitic). If we suppose that the usual word-initial stress is supplied by a lexical redundancy rule, and that proclitics (enclitics) are marked as such in the lexicon, then the phonological words can be created by readjustment rules (in the sense of Chomsky and Halle 1968:9ff,371ff).

Personal pronouns (case-marked or otherwise) belong in the class of *simple clitics* (Anderson 2005) because they are "elements of some basic word class and appear in a position relative to the rest of the structure in which the normal rules of syntax would put them". They are proclitic if they precede the main verb of the clause, but enclitic otherwise. This makes it hard to classify them according to the typology proposed by Klavans 1985. The relevant cliticization rule can be subdivided into two extremely simple rules:

(A) attach the pronoun to the following tensed verb, elsewhere

(B) attach it to the preceding word.

Another such rule will delete the (lexical) stress from verbs preceded by some VM (Verbal Modifier, see Ackerman and Komlósy 1983). Ackerman 1984 demonstrates that VM+V constructions can be lexical. In such cases the required stress pattern comes simply from the lexical redundancy rule mentioned above. If redundancy rules are part of the grammar, the relatively open nature of the V^1 class can readily be explained: the Derivational Simplicity Criterion of Kiparsky (1982) forces us to reanalyze VM+V constructions as V^1s. In a 'percolation' theory of word-formation (such as Lieber 1981), no reanalysis is called for, since the trees resulting from compounding, affixation, and cliticization are identical.

These rules, and in general rules governing the placement of simple clitics require only local information (namely the category of the adjacent words) for their operation. The rules governing the placement of *special clitics* (in the sense of Anderson 1986a), however, will require global information inasmuch as they are sensitive to the labelling of constituents above the word level. If special clitics are simply phrasal affixes, as alluded to in several recent papers dealing with clitics (including Klavans 1985, Zwicky 1985), and explicitly stated by Anderson (1986a), then it should be possible to find clitics that go with phrases of arbitrary bar-level.

The definite article *a/az* in Hungarian cannot be attached to full NPs (as shown in Kornai 1985a), just to bare nouns (N^0) or to nouns modified by an adjective and/or numeral (N^1 and N^2 constructions).

Since *a/az* is clearly proclitic, we are forced to the conclusion that definiteness is an inflectional category in Hungarian. This is supported by the existence of the "definite object" and "indefinite object" conjugations to be discussed in 4.1.

Certain words appear in the syntactic representation with a diacritic **f** (read foykes). Words in Focus position (see É. Kiss, 1981) or having particular communicative functions (contrastive topic, see Szabolcsi, 1981; contrafocus, see Varga 1982; etc.) are marked by this diacritic. The rules generating the intonation patterns of corrective sentences (phrases) are also triggered by **f**. (For the syntax and semantics of corrective sentences see Kálmán 1985, and for sentences with more than one foykes, see Bánréti 1983.)

The role of **f** in the formation of the phonological representation is twofold. First, elements marked by it cannot cliticize. Second, if a word is marked by **f** in the (postlexical) underlying representation, a rule of Eradication will delete the accent and word boundaries of the subsequent words up to the next **f** or phrase boundary (#). (S_0 denotes zero or more syllables.)

$$[_{\mathbf{f}}S_0 \overset{*}{S} S_0][S_0 \overset{*}{S} S_0]_1 \rightarrow [_{\mathbf{f}}S_0 \overset{*}{S} S_0(S_0 S S_0)_1]/\_ \left\{ \begin{array}{l} [_{\mathbf{f}} \\ \# \end{array} \right.$$

Therefore, the segments from one **f** to the next one (or to the end of the phrase) will form a single phonological word having only one accented syllable.[30] The (intonation) rules operating on phonological words are discussed in Kornai and Kálmán 1989; here it is sufficient to say that 'derived' phonological words behave exactly as the 'non-derived' ones that come from the lexicon.

The point here is that certain pieces of information provided by the syntax (e.g. bar-levels and the diacritic **f**) play a crucial role in the formation of phonological words. Therefore, the grammar must be organized in such a manner that these rules will *follow* lexical insertion. The model adopted here is quite traditional:

(37) Syntax → Lexical Insertion ← Lexical Pronology and Morphology

↓

(Surface) Syntactic Representation

↓

Cliticization Rules, Postlexical Internal Sandhi

↓

Underlying Tone Level

↓

External Sandhi, Intonation Rules

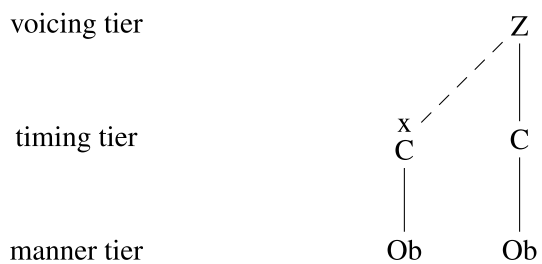
↓

Surface Phonetic Form

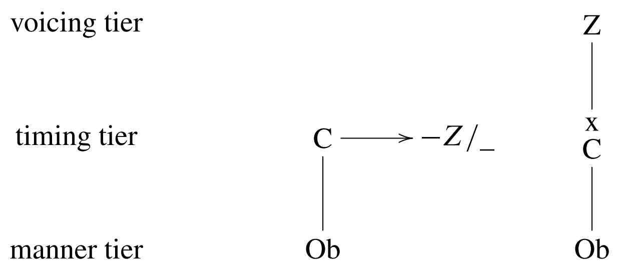
[30] 'Eradicating' stress (see Kálmán et al. 1984) is generally believed to be stronger than the normal stress. Minimal pair tests, however, do not show a linguistically significant contrast: apparently, the quantitative difference between normal and eradicating stress is a secondary phenomenon due to the fact that eradicating stress generally appears on substantially longer phonological words (see also Kálmán et al. 1984).

The sandhi rules that ought to be mentioned are the following:

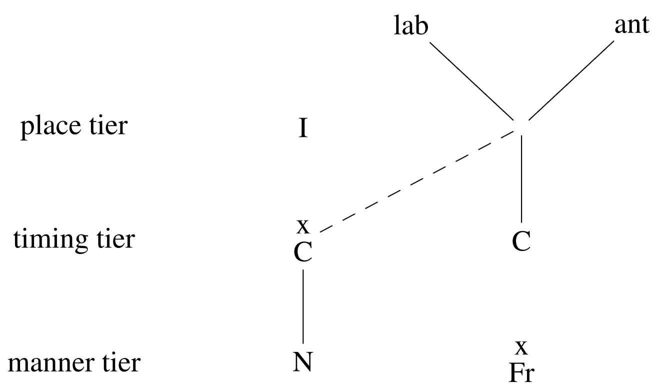
(38) Voicing Assimilation



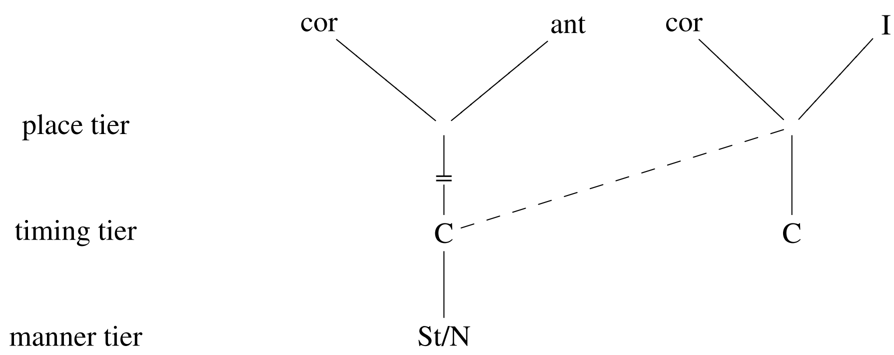
(39) Obstruent Devoicing



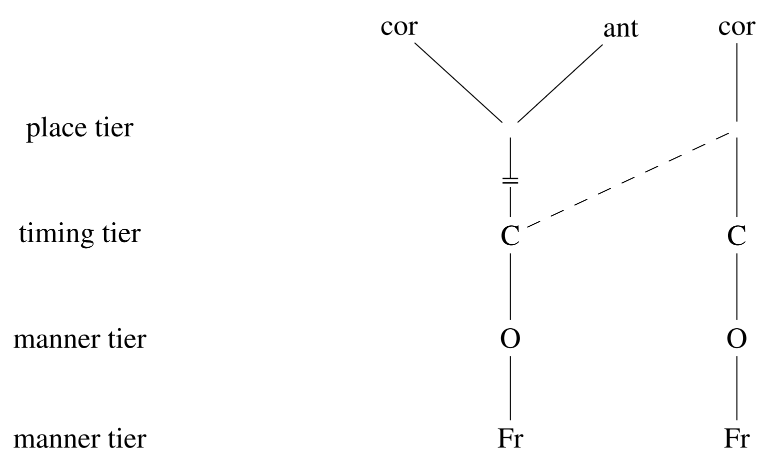
(40) Nasal Assimilation



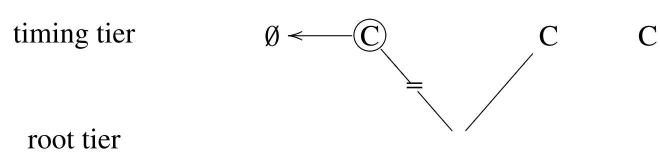
(41) Palatal Assimilation



(42) Fricative Assimilation (Optional)



(43) Degemination



2.6 Appendix

As we have seen in 2.3, the key generalization concerning lowering stems is that they have the same (lowering and unrounding) effect both in Class I and Class IV. By marking lowering stems with a single diacritic feature, we also explain the distribution of lowering stems, i.e. the fact that every (surface) vowel can be lowering. In a four feature system, there are two features which can possibly trigger lowering and unrounding, namely ⟨+low⟩ and ⟨–round⟩. For reasons to be discussed later, we will take ⟨+low⟩ as the diacritic playing the role of ML.

The major advantage of having a fourth feature at our disposal is that we can resolve conflicts between underlying and surface feature values by means of underspecification. In a system containing only three binary features, we can describe at most eight phonemes as long as we interpret the features as privative; leaving one or more features out of the representation leads to other phonemes, rather than archiphonemes.[31] But if we can distinguish some phonemes by the fourth feature, the way is open for leaving the values of some other features underspecified. Let me illustrate this on the harmonic pair *a/e*, occurring e.g. in the dative suffix *nak/nek*.

In the tridirectional system, *a* is defined as [+A –I –U], and *e* is defined as [+A +I –U]. Thus, the archiphoneme *a/e* should be defined as [+A 0I –U]. However, the absence of I in a privative feature system is equivalent to a negative specification, so that in fact we end up with [+A –I –U], which is *a*. Thus, a system of three privative features simply has no room for archiphonemes. The only possible interpretation is that the suffixes alternating between *a* and *e* have underlying *a*, and the operation of harmony (I-spread) changes the I-value. In such a system, the conflict between the surface ⟨+round⟩ of *a* and its underlying –U has to be resolved by late 'realization' rules.

In the SPE system we can define both *a* and *e* as underlying [0round], and collect these vowels into an archiphoneme which is also [0back]. This archiphoneme can be kept distinct from other vowels and archiphonemes by defining it as [+low]; as we shall see, the specification for ⟨low⟩ will play a pivotal role in the definition of the ternary and quaternary archiphonemes as well. Before turning to these, let us define the remaining binary archiphonemes: *o/ö* = [+round –high –low 0back]; *u/ü* = [+round +high (–low) 0back].

There are two ways to handle binary harmony in such a system: one is to suppose that both ⟨+back⟩ and ⟨–back⟩ spreads (so that the archiphoneme always receives its specification for backness from the stem), and the other is to suppose that only one of them spreads, and the other value is inserted by a later rule. In this respect, privative systems are more restrictive since a negative value is encoded by the absence of a feature, so only the positive value can spread.[32]

Thus, the basic pattern of binary harmony will be captured by a pair of spreading rules which supply the backness value for the binary archiphoneme on the basis of the backness specification of the stem vowel:

[31] Leaving all features out would lead to the empty vowel, which has no phonetic realization in Hungarian.

[32] Since the present system has the resources to spread both values, it is immune to the criticism levelled against privative backness by Farkas and Beddor (1987). I will return to this issue below.

(44A)

$$\begin{array}{lll} \mathrm{F} & & \\ | & \ddots & \\ \mathrm{V} & \mathrm{C}_0 & \mathrm{V} \end{array}$$

(44B)

$$\begin{array}{lll} \mathrm{B} & & \\ | & \ddots & \\ \mathrm{V} & \mathrm{C}_0 & \mathrm{V} \end{array}$$

In case the stem vowel is neutral, we have a number of options. First of all, we might take the surface frontness of neutral vowels as indicative of underlying frontness, and define neutral vowels as underlyingly linked to the autosegment F. Second, we might take the transparency of neutral vowels as diagnostic, and define them as underspecified for frontness (and suppose that they receive their final specification by a late rule). Third, we might suppose that the autosegment F floats over neutral vowels, and links up only at the end of the derivation.

Given the discrepancies between underlying and surface representations (cf. the discussion in 2.3), we can not *a priori* exclude the remaining two possibilities, namely that neutral vowels are underlyingly linked to a B autosegment, or that B is floating in the underlying representation of neutral vowels. Although it is possible to invoke principles such as *structure-preservation* to rule out these combinations, it should be kept in mind that neutral vowels can show three kinds of harmonic behavior (two of which involve back suffixes), so that three out of the five logical possibilities are actually needed.

Let us now turn to the case of ternary harmony. Since ternary alternation is between *o,e* and *ö*, we will suppose that the ternary archiphoneme is simply an otherwise underspecified mid vowel, given by [0round –high –low 0back]. If a back stem precedes, this archiphoneme receives a B by (44A) and thereby becomes *o*; the redundant ⟨+round⟩ is supplied by a later rule. If a front round vowel precedes, both F and R spread. This is the rule of Ternary Harmony, given as (45) below.

(45)

$$\begin{array}{llll} \text{backness tier} & \mathrm{F} & & \\ & | & \ddots & \\ \text{CV tier} & \mathrm{V} & \mathrm{C}_0 & \mathrm{V}< -H > \\ & | & \cdot^{\cdot^{\cdot}} & \\ \text{rounding tier} & \mathrm{R} & & \end{array}$$

In case the stem vowel is neutral, the *e* alternant is derived by binary harmony, because there is no R to spread, and (45) remains inoperative. This leaves only quaternary harmony to be accounted for; as we shall see, this will follow from the rules we already have.

Since quaternary alternation is between *a,e,o* and*ö*, we will suppose that the quaternary archiphoneme is simply an otherwise underspecified non-high vowel, given by [0round –high 0low 0back]. Since this archiphoneme differs from binary *a/e* only in that the latter is ⟨+low⟩, we are in a position to recapitulate the correct part of Vágó's a-Epenthesis analysis simply by taking the feature ML to be ⟨+low⟩. The effect of this ⟨+low⟩ is to turn the quaternary archiphoneme into binary *a/e* which will be subject to the rules of binary harmony. Before discussing in greater detail how this system works, let me summarize the underlying representations of the vowels and archiphonemes encountered so far.[33]

(46)	a	e	i	o	u	ü	ö	a/e	o/ö	u/ü	o/e/ö	a/e/o/ö
back	+	-	-	+	+	-	-	0	0	0	0	0
high	-	-	+	-	+	+	-	-	-	+	-	-
low	+	+	-	-	-	-	-	+	-	-	-	0
round	0	0	-	+	+	+	+	0	+	+	0	0

Let us now see how the four-feature solution accounts for the harmonic behavior of the stems in each of the five classes.

Class I stems are marked with a floating ⟨+low⟩. In addition, they are associated with a B autosegment. In the binary case, this B spreads — the floating ⟨+low⟩ cannot dock because all binary archiphonemes are specified for ⟨low⟩. In the ternary case, ⟨+low⟩ cannot dock because the ternary archiphoneme is specified as ⟨–low⟩, and Ternary Harmony is inoperative because there is no F to spread. Thus, the spreading of B gives us a back mid vowel, and a late rule, given as (47) below, will supply the redundant ⟨+round⟩ to derive *o*. In the quaternary case, however, the ⟨+low⟩ can dock on the quaternary vowel, and thereby turns it into binary *a/e*, and the spreading of B will derive the correct *a*.

(47) –high αback $\rightarrow$ αround

Thus in the binary and quaternary cases we make use of the fact that rounding on short low vowels is redundant (determined by backness). In the ternary case we use (47) to determine the rounding of a mid vowel. Given that *ö* is ⟨–back⟩ but ⟨+round⟩, we must make use of the fact that (47) applies in a feature-adding rather than a feature-changing fashion. One way to assure this manner of application would be to add ⟨0round⟩ to the structural description of (47) — but this would amount to adopting a three-valued feature system.

Class II stems are not marked with a floating ⟨+low⟩, but they are associated with a B autosegment. To be more precise, the B is associated with the vowels that appear in Class II regularly, namely *a,u* and *o*, and appears as a floating feature with neutral vowel stems. In Class I and Class II, neutral vowels are associated with F in the lexicon, and the unassociated B follows this F on the B/F tier. In the binary case, the B of the stem spreads (or if it was floating, just links up) to supply the backness of the following archiphoneme. In the case of *a/e*, (47) also applies. In the ternary case the derivation is same as for

[33]Since long vowels are treated as geminates, only the short vowels are given. The phonetic differences between short and long vowels are treated by late realization rules for *á* and *é* in the manner of Pāṇini 8.4.68 **a a**. Given the complexity of the interpretative metatheory behind the Aṣhṭādhyāyī of Pāṇini, the following comments on 8.4.168 by Dr. Ballantyne (quoted in Vasu 1891) might be helpful:

"In actual use the organ in the enunciation of the short *a* is *contracted*; but it is considered to be *open* only, as in the case of other vowels, when the vowel *a* is in the state of taking part in some operation of Grammar. The reason for this is that if the short *a* were held to differ from the long *ā* in this respect, the *homogeneousness* mentioned in I.1.9 would not be found to exist between them and the operation of the rules depending on that homogeneousness would be debarred. In order to restore the short *a* to its natural rights, thus infringed throughout the Aṣhṭādhyāyī, Pāṇini with oracular brevity in his closing aphorism gives the injunction *a a* ; which is interpreted to signify — Let short *a* be held to have its organ of utterance contracted, now that we have reached the end of the work in which it was necessary to regard it as being otherwise".

Class I, but in the quaternary case the situation is different, because now there is no ⟨+low⟩ to create a binary suffix, and the spreading of B leaves the quaternary vowel underspecified for ⟨low⟩, even after (47) has supplied the value ⟨+round⟩.

There are two ways of supplying the missing ⟨–low⟩: it can be inserted by *rule*, or it can be part of the *representation* of quaternary vowels as a floating feature. Since most quaternary suffixes lower the following suffix, we must suppose that they carry a floating ⟨+low⟩: this means that if we take the missing ⟨–low⟩ to be part of the representation, we would have to presume a floating ⟨–low⟩ ⟨+low⟩ melody. Since the alignment of such melodies with the vowels appearing in suffix-combinations could not be achieved by straightforward Left-to-Right One-to-One mapping,[34] I will suppose that the ⟨–low⟩ is in fact supplied by a rule, which is given in (48):

(48) –high → –low

This rule also applies in a feature-adding fashion: it operates *after* the ⟨+low⟩ of the stem had a chance to link up, but *before* the ⟨+low⟩ of the suffix could link up (in the case of those suffixes which are themselves marked with a ⟨+low⟩).

Class III stems are not marked with a floating ⟨+low⟩, nor are they associated with a B autosegment. There is no reason to suppose that the neutral vowels are transparent in such stems, so I will simply assume that they are associated with an F autosegment. In the binary case, this F spreads on the suffix vowel, and for *a/e* rounding is again supplied by (47). In the ternary case, there is no ⟨round⟩ for Ternary Harmony to spread, so we derive the unrounded front mid vowel *e* by (44A) and (47). In the quaternary case, we get the same result by applying (48) as well.

These derivations are fairly straightforward, except for the fact that the result is a mid, rather than a low *e* in the ternary and quaternary cases. Thus, we will need some cleanup rules for neutralizing the ⟨+low⟩ *e*-s that we derive from binary *a/e* alternation and the ⟨–low⟩ *e*-s resulting from ternary harmony. The need for such neutralization rules (which are arguably synchronic remnants of a diachronic neutralization process in Hungarian) is not unproblematic. I will return to this question below.

Class IV stems are marked with a floating ⟨+low⟩, and they are associated with an F autosegment. In the binary case, this F spreads on the suffix vowel, and for *a/e*, we also use (47). In the ternary case, the conditions of Ternary Harmony are met, and we derive the *ö*-alternant. Notice that (47) cannot supply the ⟨–round⟩ value, since (45) has spread the ⟨+round⟩ from the stem, and (47) is feature-adding, rather than feature-changing. In the quaternary case, the ⟨+low⟩ of the stem docks on the quaternary vowel, and thereby turns it into the archiphoneme *a/e*. Binary harmony then supplies the feature F, and ⟨–round⟩ is given by (47). Notice that the quaternary *e* derived this way is ⟨+low⟩, while the quaternary *e* after Class III stems was ⟨–low⟩.

Class V stems are not marked with a floating ⟨+low⟩, and they are associated with an F autosegment. In the binary case, the derivations are the same as in Class III and Class IV, and in the ternary case the

[34] In general, harmonic systems do not have melodies (floating or otherwise), but the highly exceptional behavior of Class I and II neutral vowels makes it necessary to assume an F-B melody for these. Notice, however, that this melody is still docked by the F, and thus can be made subject to the standard rules of autosegmental mapping.

rule of Ternary Harmony applies the same way as in Class IV. In the quaternary case, Ternary Harmony applies, and the ⟨–low⟩ feature is supplied by (48).

Finally, let me contrast and evaluate the solutions presented in 2.3 and in this Appendix. First I will discuss whether both F and B, or just one of them, spreads. Next, I will present what I take to be the relevant data concerning neutral vowel stems, and issue some cautioning notes concerning widely accepted 'generalizations' about neutral vowels and vacillation. I will argue that the vacillation is not understood well enough to serve as a testing ground for descriptions of Hungarian vowel harmony, and that we need other data to decide between the the solutions offered above.

One of the major differences between the solutions presented in 2.3 and 2.7 is in the basic rule of binary harmony: in the three-feature solution I have supposed that only I (=⟨front⟩) spreads, while in the four-feature solution the assumption was that both ⟨front⟩ and ⟨back⟩ spread. This difference corresponds to the traditional distinction between *privative* and *equipollent* features (Trubetzkoy 1958).

My starting point will be Farkas and Beddor 1987, which objects to privative backness in Hungarian on the following grounds. First, privative feature analyses cannot predict the pattern of abstract vowels, and, second, non-harmonizing non-neutral suffixes such as the diminutive *-kó* are problematic for feature-changing accounts. Let us take these objections in turn.

An important advantage of the tridirectional system is that it predicts what kinds of binary alternations are possible in Hungarian. If an alternating pair is defined by the presence vs. absence of the feature I, *a* will be paired with *e*, *ó* will be paired with *ő* and *u* will be paired with *ü*. These are precisely the pairs attested in Hungarian. Moreover, since *i* is paired with the empty vowel, which is phonetically uninterpreted in Hungarian, the system predicts that *i* cannot take part in alternations. As the definition of binary archiphonemes in (46) makes clear, the four-feature system does not make any comparable predictions.

However, neither the tridirectional nor the SPE system predicts the existence or the behavior of ternary and quaternary alternations, and both systems make wrong predictions concerning the distribution of abstract vowels. For instance, the tridirectional system would permit a fourth kind of *i* which has a floating –U but no other exceptional property,[35] and the SPE system would permit a floating ⟨round⟩. Thus it is true that both systems have the resources to describe harmonic behavior that cannot be found in Hungarian — but this is no argument against privative backness in and of itself.

Let us turn to the problem of non-harmonic non-neutral suffixes. If the feature I spreads from the stem to the right in a feature-changing manner, how can we protect an invariant suffix such as *kó* from becoming *kő*? The privative approach makes it necessary to posit an internal word-boundary for such suffixes (cf. SPH:ch. 1.6), as Farkas and Beddor correctly point out. However, the seriousness of this problem is proportional to the number of invariant back suffixes, and in Hungarian only a few invariant back suffixes can be found.

[35] In the present system, all neutral vowels with a floating –U are specified for I in the core, rather than on the I-tier.

First, disyllabic suffixes in which the first vowel is neutral (e.g. *-izmus, -ista*) require no word boundary. They can be treated as containing a non-transparent neutral vowel, i.e. an I on the I-tier, which blocks the spreading of the stem I. (The blocker itself will not spread because the environment is not derived.) Second, some of the remaining suffixes are arguably harmonic, although SPH lists them as invariant. Most notably, the diminutive *-us* has an alternant *üs* (cf. *Tünde, Tündüs*), and the diminutive form of *Gergely*, *Gergő* makes it plausible that the diminutive suffix *kó* also alternates (cf. also *Anikó, Enikő*).

This leaves us with only one indisputably invariant back suffix, namely *-kor* '(temporal) at', and here a word boundary can be justified by the fact that *kor* also appears as a free form (in the meaning 'age'). Thus I conclude that in Hungarian no argument against privative features can be based on invariant back suffixes.[36]

The behavior of neutral vowels in Hungarian does not lend itself to clear-cut generalizations. The basic generalization concerning neutral vowels in Hungarian is that they are *transparent*, i.e. that they let the F/B feature of the preceding vowel spread onto the following suffix. For suffix vowels, this has been illustrated above on the denominal adjective-forming suffix *-i* 'characterized by the location', which derives e.g. *rumi, töki* 'inhabitant of Rum, Tök' from the place names *Rum, Tök*. As we recall, in the plural these become *rumiak, tökiek* 'inhabitants of Rum, Tök' rather than **rumiek* or **tökiak*, and similarly with binary suffixes (*ruminak, tökinek*). Some further examples are given below in (49):

(49)	stem	ACC	LOC	LOC+PL	
	ház	házat	házi	háziak	'house'
	kert	kertet	kerti	kertiek	'garden'
	nyár	nyarat	nyári	nyáriak	'summer'
	tél	telet	téli	téliek	'winter'

This is somewhat problematic for the tridirectional feature system, since the *feature* I spreads, but the *vowel i*, which contains only this feature, leaves this spreading unaffected. This problem is solved by relegating the I specification of neutral vowel suffixes to the core, so that the I-tier remains available for spreading. In those cases where the harmonic spreading is actually affected (as in the invariant latinate suffixes discussed above), the I appears on the I tier and *blocks* the spreading.

In the four-feature solution, we can use underspecification instead of core-specification. This is also problematic, since the frontness of the neutral vowels is redundant only among non-low unrounded vowels, and so far we have treated rounding as predictable from backness. However, if we take the transparent suffix vowels to be lexically specified as ⟨–round⟩, the ⟨+front⟩ specification can be supplied by a late rule.

In stems, the behavior of neutral vowels is quite complex. I will try to make sense of this complexity in terms of a probabilistic description below. Roughly speaking, a single neutral vowel in the last syllable of the stem is transparent: if the preceding vowel is back, the stem takes back suffixes, otherwise it takes

[36] Both solutions predict the existence of invariant front suffixes. Apparently, no such suffixes exist in Hungarian.

front suffixes. If the vowel preceding the final neutral vowel is also neutral, we encounter *vacillation*.[37] There are a number of vacillating stems that contain only one neutral vowel (e.g. *Ágnes*), and there are lots of non-vacillating stems with one or more neutral vowels. We can find a number of generalizations in the literature concerning vacillating stems: for instance, van der Hulst 1985:276 lists the following:

- Stem-final *i, í, é* are typically transparent
- A stem containing a back vowel followed by *e* typically vacillates
- A stem containing a back vowel followed by two neutral vowels typically vacillates

These 'generalizations' are listed in decreasing order of plausibility. It appears to me that stems containing a back vowel followed by two neutral vowels typically take front suffixes. The present treatment is based on native speaker judgments[38] rather than on exhaustive testing like in Ringen and Kontra 1989. Nevertheless, I think that the pretheoretical, probabilistic model presented below gives a fair picture of the situation.

Let us model the decision process by which Hungarians decide on the quality of the suffix vowel by a three state finite automaton that scans the stem right to left. In the initial state the front (default) alternant is selected. As the automaton moves backwards, it will encounter a front, a neutral, or a back vowel first and accordingly move into state F, N, or B. The F and B states correspond to making a decision in favor of the front or the back alternant respectively — these are final states. The N state corresponds to the situation where the last vowel of the stem was neutral, so that harmony is determined by the preceding vowel: if it is front, the automaton moves into state F, and if it is back, the automaton moves into state B. If the preceding vowel is neutral, the automaton stays in state N with some probability (p $\approx$ .2) *or* goes to F (with probability 1-p $\approx$ .8): it is this choice that gives the vacillating behavior. Since this state is non-final, the automaton goes on and investigates the preceding vowel by repeating the steps described above until it falls into a final state or runs out of vowels (in which case the front alternant is selected). Clearly, the more neutral vowels the automaton encounters, the more likely it is to fall into the F state.

In this model, vacillation corresponds to a probabilistic choice, and there is no attempt to analyze the mechanism that leads to such choices. Nevertheless, it is remarkable that in this probabilistic setting the operation of the automaton is greatly simplified by scanning right to left.[39] The reason for this is that spreading is left to right in Hungarian, so the decision in favor of taking the front or the back alternant of a suffix has to be delayed until the rightmost vowel of the stem is inspected. This is particularly clear

[37] The literature in general makes no distinction between the cases where vacillation means idiolectal variation (i.e that each individual speaker uses either the front or the back alternants quite consistently), and the cases where we can speak of true vacillation in the sense that one and the same speaker uses both back and front alternants.

[38] In particular, the judgments encoded in the 'Debrecen Thesaurus', which is the common source of Papp 1968 and the dictionary database (Kornai 1986) that was my main source of data.

[39] Lauri Karttunen (pc) notes that the direction of scanning has no effect on the simplicity of the automaton itself, only on the average number of steps it takes to make a decision. This difference is completely immaterial in a Two-level implementation (see 4.3) where everything must be completely scanned anyway.

in the case of disharmonic stems. Stems like *föderativ* 'federal' always take back suffixes, and stems like *zsonglőr* 'juggler' always take front suffixes. Thus, the decisive factor is the last non-neutral vowel, although its effect might be obscured if two or more neutral vowels follow. The automaton given above will work with disharmonic stems without any modifications, and will scan the whole stem only in case every stem vowel (except perhaps for the first one) is neutral. In order to handle disharmonic stems, the comparable automaton that scans left to right must always scan the whole stem.[40]

In the three-feature solution, the treatment of neutral vowels in polysyllabic stems is based on the assumption that *i* and *e* have the feature I in the core unless they are stressed (= appear in the first syllable). In the four-feature solution the same vowels are treated as underlyingly ⟨–round⟩, with no specification for frontness. However, we do not have independent evidence that stress and harmony interact in Hungarian, so the decision to take the *first* neutral vowel in a neutral stem as the source of the spreading is quite arbitrary. As Farkas and Beddor 1987 points out, we could use the last (or the next to last) position just as well.

Let us see how the generalizations listed above fare in the light of the probabilistic model. If a stem ends in *i, í,* or *é,* the automaton moves to N, and, in the next step, it will choose the B (or F) state if the preceding vowel is back (or front). This means that *i, í, and é* are indeed transparent. In contrast to the second generalization, the model predicts *e* to be also transparent.[41] Trisyllabic stems in which the first vowel is back and the other two are neutral are predicted to be vacillating by the third generalization listed above. My impression is that such stems show front harmonic behavior with a very high probability.

While previous treatments have concentrated on vacillating stems, in the present study I have shifted the emphasis onto exceptional but non-vacillating stems where I believe it more properly belongs. The reasons for this shift are methodological. For every speaker of standard Hungarian, the exceptional (Class I and Class IV) stems show the same unambiguous lowering behavior irrespective of sentence stress, syntactic environment, register, etc. In the phonological study of vacillation it would be necessary to control for all of these factors, and perhaps for others as well — in contrast, the information on exceptionality is readily available from standard dictionaries.

The probabilistic model presented above is a temporary solution until more data become available. Until then, the following facts should be noted. First, there are suffixes which often form vacillating stems (perhaps the best example is *-né* 'Mrs'). Second, vacillation is not a transient phenomenon due to some recent sound change. The first systematic grammar of Hungarian, Szenczi 1610, already mentions a few vacillating stems such as *Józsué* 'Joshua'. Nearly four hundred years later, these stems are still vacillating.[42] Third, as demonstrated by Kontra et al. 1987, vacillation is highly influenced by the

[40] Another argument in favor of right to left scanning, due to Péter Siptár, runs as follows. As long as we do not derive stem-internal harmony by rule we can suppose that there is no spreading in non-derived environments. In this case the only possible location for a harmonic trigger is at the end of the stem.

[41] The continuing empirical investigations of Kontra and Ringen (1986, 1987) are expected to shed more light on this matter.

[42] László Elekfi (pc) pointed out the fact that the time span involved here is probably much longer, since vacillating stems such as *fráter* 'monk' were already vacillating in the 15c mss. where they are first attested.

harmonic properties of the preceding words. Fourth, neutral vowels do not seem to form a homogeneous group with respect to vacillation (see Ringen and Kontra 1989).

Given that vacillation is clearly influenced by factors such as the harmonic properties of the preceding words which are not taken into account by any of the existing models, there is no reason why the evaluation of these models should be based on vacillating data. If there is any conclusion that can be drawn on the basis of such data, it is the fact that a (probabilistic) description involves right to left scanning of the stem. However, even this conclusion can be established on the basis of non-vacillating data as we will see in 4.2.

The original aim of underspecification (leaving feature values blank) was to capture *archiphonemes*, i.e. underdifferentiated entities such as a 'placeless nasal' that will be fully specified only in the course of the derivation. In the case of vowel harmony this means that archiphonemes are similar to vowels in their featural composition, only they lack any specification for the harmonic feature, which will be supplied by spreading. Contemporary theories of *restricted underspecification* (Steriade 1987, Itô and Mester 1988) are based on the assumption that only redundant features can be underspecified. But in the case of harmony, the missing features are not redundant (cf. minimal pairs like *kor* 'age' vs *kör* 'circle'), so archiphonemes must be exceptional elements.

The exceptionality of archiphonemes is encoded as a condition on rule application in the four-feature system, where feature-filling rules are blocked from applying by the presence of underlying feature-specifications. In the three-feature system, a mechanism for filling in blanks is incompatible with the 'simplex' feature interpretation in which the lack of specification is equated with negative specification. Here the exceptionality of archiphonemes is encoded in the representations by core specification.

For instance, if we take the dative suffix as underlying *-nak*, the alternant *-nek* will be derived by a feature-changing rule of I-spread (20A). However, we have to block U-spread (20B) from applying. In other words the *a/e* archiphoneme has to be –U throughout the derivation. This is achieved by specifying *a/e* as –U in the core. Core specifications do not spread, which is compatible with the view that negatively specified autosegments, when floating, will trigger the deletion of the following positively specified feature.

We will see in 4.2 that the three-feature solution yields a somewhat simpler analysis of the possessive paradigm. However, its advantage over the four-feature solution is slight, especially in the light of the fact that it must make recourse to two kinds of exceptional devices, namely floating negative features and core-specification.[43] Whether these devices are just technical tricks that will be eliminated from the theory of phonology as soon as a better notation presents itself or whether they capture essential properties of the grammar, only time will tell.

The four-feature solution is also problematic, especially in the light of the general principle of *structure-preservation* (Kiparsky 1982). The source of the problem is the height specification of *e*. In binary alternation, as well as in quaternary alternation after lowering stems, *e* receives the value ⟨+low⟩. In ternary alternation and in the regular quaternary case (after non-lowering neutral stems), however, *e*

[43] This is the main reason why the U-deletion rules (21-22) and (24) could not be collapsed.

surfaces with ⟨–low⟩ specification. Since phonetically *e* is mid-low, the choice between low and mid is arbitrary. But no matter how we choose to represent *e*, one of the intermediate representations will violate structure-preservation. Again, only time will tell whether structure preservation represents an essential property of grammars. If it does, however, as I believe to be the case, the four-feature solution will have to be reworked completely.

In sum, the exceptional items (irregular vowels and archiphonemes) are formally closer to the non-exceptional items (regular vowels) in the four-feature solution than in the three-feature solution. Floating features are necessary in both solutions, but in the three-feature solution core-specification for the exceptional cases is also necessary, while in the four-feature solution blank entries are common to exceptional items and regular (non-alternating, non-lowering) vowels such as *a*. But this homogeneity is bought at the price of an extra feature, namely ⟨low⟩.

Following Householder 1952, we can distinguish two polar opposites in the way we interpret phonological features. Under the 'God's truth' view, features, together with their phonetic interpretation, are part of Universal Grammar and, as such, they are likely to be innate. Under the 'hocus-pocus' view, features are part of the descriptive machinery we call Universal Grammar, but they are not necessarily related to anything outside this grammar. In particular, they need not have a universal phonetic interpretation.

In 2.3 I have performed the hocus-pocus necessary to derive three features on the basis of vowel alternations (both harmonic and length alternations) in Hungarian, and concluded that that these three features correspond to the standard SPE features only to a limited extent. In particular, the derivation made it necessary to import the abstract ⟨–round⟩ analysis of *a* into the system from the beginning, and to treat *e* and *é* as having the same height.

From this perspective, the feature ⟨low⟩ is completely unmotivated, and the four-feature analysis, which makes essential reference to ⟨low⟩ in order to keep the ternary and quaternary archiphonemes apart, must have its justification outside the domain of vowel alternations. Under the standard view, this justification comes simply from the phonetic quality of the vowels; after all, *a* is low, and *i* is high, which is exactly what we need for the analysis to work.

However, this simple justification does not stand up in the light of the facts that *a* acts as ⟨–round⟩ at every stage of the derivation and that the height differences between *a* and *e* on the one hand, and *á* and *é* on the other hand, are not reflected in the underlying representations of these vowels. In fact, there is no reason to suppose that the use of four features involves any less hocus-pocus than the use of three features.

3 Words and Paradigms

In this chapter I will address two fundamental questions of morphology: What is a word? What are lexical categories? After a brief overview of the defining criteria for words, in 3.1 I propose a definition of lexical categories based on (inflectional) morphological criteria. On the basis of the Hungarian nominal paradigm I argue that the morphosyntactic features describing the inflectional possibilities of a stem should be arranged in a tree structure in such a way that only marked features dominate other features. Stems with defective paradigms are shown to fit into this scheme. In 3.2 the lexical categories of Hungarian are described with the aid of a feature system that is argued to be cross-linguistically relevant.

3.1 Some definitions

The problem of defining the term 'word' is a notorious one. First, the term is ambiguous between 'lexical entry', 'stem', 'citation form' and 'fully formed word'. In the following discussion only the last meaning of 'word' will be important, as this is the one that should correspond to observables. Lexical entries and stems are theoretical constructs, and their precise form will greatly depend on the degree of abstractness allowed in phonology/morphology. Citation forms have no theoretical status at all; they are, at best, artifacts of a lexicographic convention that insists on representing lexical entries by surface forms (I will return to this question in 4.1).

The classical definition of words as '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 will be taken as maximal associated sequences of timing units. More precisely, a phonological word contains all the (feature) material associated to its timing units. In this sense, at least, phonological words are indeed made up from phonemes. The phonological words defined this way will generally coincide with

(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 (see 2.5), and bracket retention (tmesis), this variety can be reduced even further. The remaining types are frequently called 'morphological word', 'lexeme', etc. This fact makes it possible to organize the grammar in the manner of (37) in 2.5.

In this work it is assumed without further argumentation that morphology supplies fully formed words (see e.g. Jensen and Stong-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, a fully formed word like *boy* cannot be inserted under a node V, or under a node N⟨+PL⟩.

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.[44]

Other categories, though highly relevant for the grammar, are not susceptible to 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 or 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.

[44]For a description of this 'concept formation' technique, see Ohala 1984:III D, Jaeger 1986.

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 (for a discussion, see Muysken and van Riemsdijk 1985). This creates a need for a theory which makes the development and the comparison of such feature analyses a feasible task.

The central claim of this section 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 Lexical Integrity Hypothesis (see e.g. Scalise 1985) 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 into 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 cannot, it is possible to classify morphemes as stems or affixes solely on the basis of their word-internal distribution. (The actual 'discovery procedure' will be somewhat more 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.)

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 traditional insight summed up by Anderson (1982) in the following criterion:

> "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, then 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 called irreducible, otherwise it is reducible to its superset paradigm. (A few intuitively reducible paradigms turn out to be irreducible under this definition, and this makes it necessary to introduce the concept of *suppletion.)*

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.

Since this definition 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 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.

As 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. 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). Anderson (1982:4.1) proposes that morphosyntactic features be arranged into layers, and Zwicky (1986) tags each feature with the relevant paradigmatic dimension. A similar representation is used in GPSG (see e.g. ch. 2 of Gazdar et al. 1985). These are all special cases of the solution adopted here, in which morphosyntactic features are arranged into trees. The node labels in morphosyntactic trees will be the features themselves, with the additional constraint that

(3) 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, which are presented here together with the rest of the nominal paradigm using the stem *sógor* 'brother in law' (cf. Antal 1961).

(4)

				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 affixes listed in the second column of (4) 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 arbitrary 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 1SG, ...,3PL, or there can be no possessor at all. It is this latter case which is truly unmarked; the remaining cases, therefore, can be subordinated to a feature ⟨POS⟩. Condition (3) makes it impossible 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 presence 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 singular, 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 (4) describes a different possessive suffix, one which behaves anaphorically (see e.g. Lotz 1967b). The presence of this suffix *-é* will be encoded by the feature ⟨ANP⟩. If the anaphoric referent is plural, we have the suffix *-éi,* which can be captured by subordinating the feature ⟨PL⟩ to ⟨+ANP⟩. The case system will be discussed in 4.2 – here a separate feature is used for every surface case ending.[45]

The unmarked member of the case system is the nominative: every other feature in the last column of (4) will be subordinated to an abstract feature ⟨CAS⟩. This gives us the nominal paradigm in the form of the following tree:

```
(5)
                          |
       ________________________________________
       |                  |       |         |
     +POS                +PL    +ANP      +CAS
       |                  |       |         |
  _______________         |       |         |
  |      |       |       FAM     PL       XYZ
 ME     YOU     PL
```

The rules that spell these trees out will be given in 4.2.

The paradigms of numerals and adjectives differ only minimally from (4): numerals have no plural forms, while adjectives can also have (comparative *-bb* and superlative *leg-*) degree affixes in addition to the suffixes listed in (4). This means that if we apply the criterion in (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: affixes 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 the comparative degree this means that word forms appearing as ⟨+COMP⟩ on the surface behave as ⟨–COMP⟩ from the morphological point

[45] This makes it necessary to introduce a feature cooccurrence restriction (much the same way as in GPSG, see e.g. Gazdar et al 1985 ch 2.3) that permits only one case feature to have positive value in any morphosyntactic tree.

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* are all encoded by the same feature (on the second branch of (5)), and we do not have the blocking effect for ±PLs on different branches as part of the grammatical metatheory. This means that if such a blocking effect is encountered, we need a specific rule to capture it (see the discussion of 'dialect S' in 4.2).

In what follows, the actual vs. potential interpretations of a feature will be denoted by double (resp. single) lines in the trees formed by the morphosyntactic features. This 'coloring' 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 viewpoint 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 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 conceived of as the end-product of certain marking conventions (redundancy rules) in accordance with the Principle of Brevity. Alternatively, morphosyntactic trees with unspecified features might be taken to represent archi-elements. This alternative is extremely relevant in the context of 'defective' stems that lack certain paradigmatic forms.

Each defective paradigm can be thought of as a smaller, but full paradigm. Viewed from this perspective, the system of morphosyntactic features should yield these smaller paradigms as natural classes of paradigmatic forms. In 2.1 I argued that in a 'geometrical' arrangement of features natural classes should be defined as consisting of those trees that contain one tree and are contained in another one. But this definition yields the desired result only for trees composed of privative features — in order to extend it to trees containing equipollent features, we must limit the class of possible 'geometries' by condition (4). In addition, we must ensure that archi-elements (required for the treatment of defective paradigms) do not get fully specified by the redundancy rules (required by the Principle of Brevity).

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 and Halle 1968:9.2), because the (only) redundancy rule employed here is not sensitive to context, and always inserts the unmarked (–) value. This rule 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, *ház*⟨+POS⟩ is the abbreviated form of *ház*⟨+POS⟨–ME –YOU –PL⟩ –PL –ANP –CAS⟩ = *háza* 'his house', while *ház*+POS stands for the class {*házam, házad, háza,* *házunk, házatok, házuk*} which is the same as *ház*⟨+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, házuk*} 'our house, their house'.

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 collected 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, which also included 'saepe' words) show some defect with respect to the possessive paradigm. One class (*fia* 'his son', *mása* 'his image',... 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', *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-fiesta', *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), as these are always lexicalized with plural morphology in Hungarian, and there are no nouns of the 'cattle, police' type. There is only one word lexicalized with the familiar plural, namely *katonáék* 'the army'. In addition, 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). But not every combination is attested: for instance, there are no nouns with a 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. Thus, labelnode trees provide a relatively straightforward and effective theory of morphosyntactic representation. As Anderson (1988) has noted, such a theory has to be supplemented by "an account of the way morphosyntactic representations are constructed and manipulated by the syntax".

The syntactic theory hovering in the background of this work is GPSG (though not the current version) — this makes it relatively easy to answer the questions posed by Anderson. In particular, there is no need to say "what INFL is, where it is, how it gets there, and how it affects binding relations". The "assignment of configurational properties" (such as case on NPs and agreement) can be handled in an identical manner, namely by context free rule schemata with greek letter variables of the kind employed by Chomsky and Halle (1968:4.3). (For a fragment of Hungarian syntax described along these lines, see Kornai 1985a.)

This mechanism is probably not restrictive enough. Notice, however, that (3) above makes it possible to eliminate the ± signs altogether. If we use the *absence* of a feature as its "–" value, much as in 2.1-2.3 above, the abbreviation conventions outlined here must be modified. Fortunately, the notion of M-classes is neutral enough to be applicable in the morphosyntactic case as well, so the generalizations concerning defective paradigms can be stated in substantially the same manner.

This leaves us with an extremely homogeneous theory in which phonological features and in general phonological representations differ from morphological features and representations only in that the former have no meaning. A natural consequence of this homogeneity is the idea to treat agreement, case-assignment, and the like as feature spreading. This is certainly an alluring possibility, but given the limitations of this work, it cannot be pursued here.[46] Anderson 1986 also asks for "a theory of the phonological realization of inflectional properties" — this fits into the morphological frame much better, and will be discussed in detail in ch. 4.

[46]In 1989 I couldn't resist the temptation to call the reader's attention to the seminal work of Sadock (1985, 1989), and also Yip et al. 1987, Lapointe 1989 in a footnote. *By 2007* it is clear that autolexical syntax (as presented e.g. in Sadock 1991) has not taken the field by storm, but it remains a valuable theory whose insights have still not been entirely assimilated in mainstream syntax.

3.2 The lexical categories of Hungarian

The person/number features that appear 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), 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

> (6) the morphosyntactic features employed in the description of paradigmatic forms can be used as X-bar features as well.

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 arbitrary features like [+Subj, +Obj] (as in Jackendoff 1977:3.2) 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, following Nida (1949), 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 these are clearly derivational.

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 cooccurrence 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'

In the category 'direction' we have a 3-way opposition to/at/from with 'at' as the unmarked member: this could be described by the feature ±OUT subordinated to the + value of the feature LATIVE. In the system of abbreviations used here, the last letter of the three-letter code for cases listed in (4) above is mnemonic for this **e**ssive/**l**ative opposition.

(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,

(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’

For further details see 4.3.[47] 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. definiteness can be an inherent property of the stem (cf. Anderson 1982: 1.3, Cooper 1983). As I have already mentioned in 2.5, ±D is not silent in the verbal paradigm, since verbs must agree with their objects in definiteness.

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. Though a few forms like *engedtessék* ‘let ... be’, *adatott* ‘was given’, etc. are preserved in idioms, these are at best derivational now. 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 (4). 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, e.g. *legeslegeslegeslegnagyobb* ‘the very very very greatest’. These features give rise to the following feature analysis for major categories in Hungarian:

[47] The 3-way directional opposition is not as transparent as the examples in (7) suggest. The gap in (7A), *felett*, is an alternant of *fölött*, and certain other forms such as *körött* ‘around’ are at best marginal in present-day Hungarian. The situation is further complicated by the fact that the ‘from’ forms *fölűl, felől* (and sometimes also *elől*) are pronounced with short *ü* and *ö* in ECH, which makes them homophonous with the members of another essive series that contains *alul* ‘down’, *hátul* ‘back’ *körül* ‘around’ etc.

(8)

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

In addition to these, there are 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 to', *kifolyólag* 'because of',...) govern oblique cases.

An important aspect of this method of category definition is that not every item that can be thought of as belonging to a broad pre-theoretical category like 'noun' will actually end up in that category according to the theoretical definition presented here. For instance, the existence of a limited set of predicate nouns like *testvér* 'brother (of)' or adjectives like *büszke* 'proud (of)' should not be taken as implying that nominals in general can govern complements. Rather, the method applied in (8) is to define nouns (or adjectives) as *never* governing other items, i.e. to reserve the term 'noun' for the non-exceptional members of the pre-theoretical category.

Predicate nominals therefore will have their own lexical categories, and the fact that they are 'nominals' will be captured by the fact that they share every feature but GOVT with ordinary nominals. But aside from this distinction, the morphological classification presented here agrees remarkably well with the traditional classification that one finds in dictionaries of Hungarian. More than 99% of the 35,171 lexemes classified as nouns in the Debrecen Thesaurus will be classified as nouns in the scheme presented here.[48] The remaining discrepancy is due to words such as *farkaskasza* 'wolf.scythe' and *jotta* 'iota' that are attested only in idioms. Since these idioms will have to be listed in the lexicon anyway, this is no great loss.

There are several open classes not listed in (8): numerals, for instance, take all the possessive suffixes, but have no number or definiteness. (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* 'the army' *elei* 'ancestors' in 3.1), but there are one-member classes that can hardly be called defective. For instance, the reciprocal pronoun *egymás* 'each other' is potentially +CAS +ANP, and actually +D –POS, and there are no other words with exactly these inflectional possibilities. Some pronouns, such as the interrogative *ki(csoda)* 'who' and *mi(csoda)* 'what' will be categorized as full-fledged nouns, but in general pronouns are defective. For instance, the personal pronouns *én, te, ő* 'I, you(sg), he/she/it' are –POS, and the possessive pronouns, which are traditionally listed under a separate minor category,

[48]This figure includes the defective nouns as well.

emerge as paradigmatic (namely +ANP) forms of the personal pronouns (see Kornai 1989).

In Hungarian, as in most (and perhaps all) languages, 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 accommodated only at the price of some arbitrary assignment of feature values. The system proposed here avoids this pitfall: all 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.

4 Inflectional morphology

In this final chapter the tools developed so far are put to use in the description of verbal (4.1) and nominal (4.2) inflection in Hungarian. Following the program of chapter 3 to its logical conclusion would require the description of the adjectival, numeral, etc. paradigms in a similar fashion — the first steps in this direction are taken in 4.3, where Hungarian inflectional morphotactics is described in the framework of a two-level implementation. In 4.4 I offer my conclusions.

4.1 Conjugation

Let me begin by simply tabulating the paradigmatic forms and a few other suffixed forms of the stems *vár* 'wait', *kér* 'ask', and *tűr* 'suffer'. In (1-3), the items in the first, second, and third columns are first, second, and third person forms, respectively. With each stem, the first two rows give the singular, and the last two rows give the plural forms: the indefinite conjugation is in the odd rows, and the definite conjugation is in the even rows.

	1st	*2nd*	*3rd*	number	conjugation
(1A)	várok	vársz	vár	*sg*	*indef*
	várom	várod	várja	*sg*	*def*
	várunk	vártok	várnak	*pl*	*indef*
	várjuk	várjátok	várják	*pl*	*def*
(1B)	vártam	vártál	várt	*sg*	*indef*
	vártam	vártad	várta	*sg*	*def*
	vártunk	vártatok	vártak	*pl*	*indef*
	vártuk	vártátok	várták	*pl*	*def*
(1C)	várjak	várj(ál)	várjon	*sg*	*indef*
	várjam	vár(ja)d	várja	*sg*	*def*
	várjunk	várjatok	várjanak	*pl*	*indef*
	várjuk	várjátok	várják	*pl*	*def*
(1D)	várnék	várnál	várna	*sg*	*indef*
	várnám	várnád	várná	*sg*	*def*
	várnánk	várnátok	várnának	*pl*	*indef*
	várnánk	várnátok	várnák	*pl*	*def*

	1st	*2nd*	*3rd*	number	conjugation
(2A)	kérek	kérsz	kér	*sg*	*indef*
	kérem	kéred	kéri	*sg*	*def*
	kérünk	kértek	kérnek	*pl*	*indef*
	kérjük	kéritek	kérik	*pl*	*def*
(2B)	kértem	kértél	kért	*sg*	*indef*
	kértem	kérted	kérte	*sg*	*def*
	kértünk	kértetek	kértek	*pl*	*indef*
	kértük	kértétek	kérték	*pl*	*def*
(2C)	kérjek	kérj(él)	kérjen	*sg*	*indef*
	kérjem	kér(je)d	kérje	*sg*	*def*
	kérjünk	kérjetek	kérjenek	*pl*	*indef*
	kérjük	kérjétek	kérjék	*pl*	*def*
(2D)	kérnék	kérnél	kérne	*sg*	*indef*
	kérném	kérnéd	kérné	*sg*	*def*
	kérnénk	kérnétek	kérnének	*pl*	*indef*
	kérnénk	kérnétek	kérnék	*pl*	*def*
(3A)	tűrök	tűrsz	tűr	*sg*	*indef*
	tűröm	tűröd	tűri	*sg*	*def*
	tűrünk	tűrtök	tűrnek	*pl*	*indef*
	tűrjük	tűritek	tűrik	*pl*	*def*
(3B)	tűrtem	tűrtél	tűrt	*sg*	*indef*
	tűrtem	tűrted	tűrte	*sg*	*def*
	tűrtünk	tűrtetek	tűrtek	*pl*	*indef*
	tűrtük	tűrtétek	tűrték	*pl*	*def*
(3C)	tűrjek	tűrj(él)	tűrjön	*sg*	*indef*
	tűrjem	tűr(je)d	tűrje	*sg*	*def*
	tűrjünk	tűrjetek	tűrjenek	*pl*	*indef*
	tűrjük	tűrjétek	tűrjék	*pl*	*def*
(3D)	tűrnék	tűrnél	tűrne	*sg*	*indef*
	tűrném	tűrnéd	tűrné	*sg*	*def*
	tűrnénk	tűrnétek	tűrnének	*pl*	*indef*
	tűrnénk	tűrnétek	tűrnék	*pl*	*def*

The rows of (1-3) are organized into blocks lettered A,B,C,D. In the first three columns, each form is present (indicative) in block A, past (indicative) in block B, (present) imperative in block C, and (present) conditional in block D.[49] The tables (1-3) are best thought of as vertical layers of a three-dimensional array: in what follows I will refer to the morphemes spanning the 'depth' of this array by their position in the tables. For instance, B13 is the 1PL past indefinite *tunk/tünk*.

The following tables (1I-3I) exemplify the 'infinitival conjugation', which is the form taken by a verb if it appears in construction with certain auxiliaries such as *kell* 'must' or *kár* 'pointless'.[50] The numbers and letters indicate a similar three-dimensional arrangement as above — for instance, I32 refers to the 3PL morpheme *niuk/niük* of the infinitival conjugation.

	1st	*2nd*	*3rd*	number
(1I)	várnom	várnod	várnia	*sg*
	várnunk	várnotok	várniuk	*pl*
(2I)	kérnem	kérned	kérnie	*sg*
	kérnünk	kérnetek	kérniük	*pl*
(3I)	tűrnöm	tűrnöd	tűrnie	*sg*
	tűrnünk	tűrnötök	tűrniük	*pl*

The following table (4) exemplifies the 'implicative' of verbs used when the subject is 1SG and the object is 2SG or 2PL. The numbers and letters follow the scheme used above: for instance, 4C gives the implicative of the imperative.

(4)	1	2	3
(A)	várlak	kérlek	tűrlek
(B)	vártalak	kértelek	tűrtelek
(C)	várjalak	kérjelek	tűrjelek
(D)	várnálak	kérnélek	tűrnélek

The following table (M) is a mixed collection of the most productive deverbal suffixes that are arguably derivational (since they are category-changing). In the manner of the previous tables, morphemes can be referred to by their position as given by column and row: for example: MPU is the infinitival suffix *-ni*.

[49] For a succinct description of the tense/mood system reflected in the block structure see Abondolo 1988:2.0.2.

[50] The infinitival conjugation is optional in ECH, but obligatory in SLH. For a detailed discussion of infinitival constructions see Kálmán et al 1989, Kornai 1988

(M1)	P	Q	R
(U)	várni	várva	várván
(V)	várás	váró	várandó
(M2)	P	Q	R
(U)	kérni	kérve	kérvén
(V)	kérés	kérő	kérendő
(M3)	P	Q	R
(U)	tűrni	tűrve	tűrvén
(V)	tűrés	tűrő	tűrendő

It is important to note that in (M) the arrangement of the elements into rows and columns is largely arbitrary. MQU and MRU are adverbial forms: the former is fully productive, but the latter is not. MPV is produced by a (fully productive) nominalizing suffix, MQV and MRV are participial forms, but only MQV (the *active participle)* is fully productive.

So far, we have 64 forms for every stem, and in a purely 'inflecting' language we could do little more than list the alternants appearing with the stems in (1-3) and note that the alternations are governed by the same rules of vowel harmony that were given for nouns in 2.3. But Hungarian is claimed to be an 'agglutinating' language (e.g. by Hall 1944:11), so one would expect a significant reduction in the number of endings as a result of segmenting the endings into component morphemes. For instance, were we able to isolate 2 person markers (plus zero for third person), a plural marker (plus zero for singular), 3 tense/mood markers (plus zero for present indicative), and a definiteness marker (plus zero for indefiniteness), the inventory of 48 portmanteau endings could be reduced to an inventory of 2+1+3+1=7 'pure' morphemes, plus a few rules of their combination.

In fact, the theoretical limits to the enterprise are even smaller. Given only *six* suffixal morphemes, each opposed to zero in some paradigmatic dimension, the most transparent 'agglutinative' morphology gives rise to exactly $2^6 = 64$ suffix combinations. Although the relevant paradigmatic oppositions are not so neatly binary here, it looks as if we have a good chance with this program: for instance, every plural form ends in *k*, and we already know that the plural marker on nouns is also *k*. Since the stem is always followed by *t* in past tense forms, by *j* in imperative forms, and by *n* in conditional forms, there are obvious candidates for the tense/mood markers as well. Even the 0 vs. *ja/i* opposition of third person singular indefinite vs. definite forms seems to recur in second person plural forms, which gives a (this time, less obvious) candidate for a suffix of definiteness.

In the case of adverbial suffixes, this approach would lead us to positing an underlying 'active' adverbial suffix *vA* (that would surface as *va* or *ve* according to the rules of vowel harmony) and an underlying suffix *An*. Affixation of this latter suffix would turn the active adverbial into the 'continuous' form *ván/vén* without any further stipulation.

Yet it would be hard to explain the curious effect *An* has on the meaning of *vA*. If the information stored with the former (namely, that the suffixed form is continuous) simply overrides the information stored in the latter (namely, that the suffixed form is active), then it also has to override the information that the form is an adverbial. If this is so, the result of the second affixation could be of any category, and the fact that *An* is *not* a category-changing affix has to be built into its representation.

Given the preponderance of category-changing affixes in the languages of the world, it is very unlikely that category-preservation is the unmarked case that can be left unmentioned. Therefore, we gain practically nothing by segmenting *vÁn* into *vA* + *An*. The overall complexity of the lexical representation belonging to the hypothetical *An* is on a par with that of the unsegmented *vÁn*.[51]

The point here is that partial phonetic similarity (e.g. of *vA* and *vÁn*) is not a sufficient basis for segmentation, even if it is coupled with partial similarity of meaning (in the example chosen, both suffixes form adverbials). But then what *does* constitute sufficient basis for segmentation? The answer to this depends on certain general assumptions which lie, in a sense, deeper than the axioms and/or conventions of the particular grammatical framework one adopts in describing a language. The discussion of such matters is usually couched in terms of a 'simplicity measure', and I will follow this practice here. Throughout the discussion, the term 'evaluation measure' or 'simplicity measure' will be used in the technical sense imparted to it by Chomsky (1965: 1.7), which is quite different from the ordinary sense of 'simplicity' based on the number and complexity of constructs ('principles', as opposed to 'rules') a theory employs.

My main proposal will be surprisingly simple: I suggest that instead of counting *features* (as proposed by Chomsky and Halle 1968:ch. 8), we should count the *labels* appearing in representations. As the reader already knows from 3.1, morphosyntactic representations are conceived of as labelnode trees, so the proposal here is tantamount to counting the nodes in these trees. In this particular case, labels are (morphosyntactic) features, so we have feature-counting of sorts. Although in the case of phonological representations, the labels will be (phonological) features, autosegmental representations are not necessarily trees (because they can have circular subgraphs wherever the absolute slicing hypothesis fails). But if counting of labels reduces to counting of features, what have we gained? The answer lies in part with the different (Prague-style) conventions of markedness employed here and in part with the thoroughness in applying this principle.

The complexity of the morphology of any language is due in large part to the exceptions. While the basic facts are usually describable with the aid of a few dozen rules, we will generally need hundreds or even thousands of minor rules and exceptionally marked lexical entries in order to describe any sizable fragment of the language. To know a language is to know its grammar *and* to know its lexicon: the main claim embodied in the label-counting measure is that these two are essentially inseparable. In particular, autosegmental rules look just like autosegmental representations, and the same label-counting measure applies to both. It is well known that in morphology there is a tradeoff between the number (and complexity) of rules that we have to apply and the size of the inventory we are left with. For example,

[51] Of course, it is possible to say that vA means only 'participle forming suffix' and An means 'continuous'. The 'active' meaning of vA can be derived by the suffixation of a zero element serving no other purpose whatsoever. This way, the original vA vs. vAAn opposition is expressed as vA+0 vs. vA+An. The original suffixes are attached to ordinary verb stems, while 0 and An can be attached only to verb+vA forms. This has to be part of their lexical entries, so the analysis involving zero will result in three relatively complicated lexical entries in the place of two relatively simple ones.

the irregular plurals *algae, larvae, ...* can be stored in the lexicon as such, or we can derive them by a minor rule of plural formation. But as the rule is quite idiosyncratic (cf. *dilemmas* vs. *antennae*) we will have to list every element to which it applies in the structural description of the rule. Therefore the only redundancy we can exploit by the rule-based solution comes from the fact that this way we need not repeat the suffix *-e* with every word. If the number of items involved is small enough, this gain might be offset by the labels we have to employ in the description of the rule itself.

For most cases, the break-even point will be with two items: listing 1+1 item in the lexicon will be slightly less expensive than listing 1 item and 1 rule, while listing 2+2 items in the lexicon will be slightly more expensive than listing 2 items and 1 rule. The reason for this is that storing control structure is just as expensive as storing any other data. When we find a part common to several lexical entries (or rules), the Principle of Brevity demands that we 'factor out' this part and retain only one token in the lexicon. But this process has its own costs because we have to equip every 'simplified' entry with some kind of pointer to the part factored out. Hungarian *ik*-verbs (discussed in more detail below) are a case in point: if we do not store the morpheme *ik* in their lexical entries, we must mark each *ik*-stem by some diacritic +IK. Obviously, a single diacritic feature would be cheaper than a whole morpheme composed of several features such as I, St, and Ob,[52] but it is not at all clear that Universal Grammar provides us with an unlimited supply of arbitrary diacritics.

One might object to this simplistic notion of simplicity on the grounds that according to it, a grammar describing, say, the paradigmatic forms of a thousand nouns will be more complex than a grammar describing the paradigmatic forms of only a hundred nouns, even if the thousand items in the former set were chosen from a single (and simple) class while the hundred items in the latter set give a representative cross-section of noun-classes. The response to this objection is that comparing grammars of such different empirical coverage is next to impossible: in the final analysis, there is no need to compare grammars (morphologies) that do not cover the whole 'irregular' word-stock. And if we have two competing grammars covering this whole domain, their complexity will also depend on the manner in which they handle the flow of productive (regular) items: the simpler the representations we can set up for these, the better the grammar we have. This follows from the definition, since, in the limiting case, the overall simplicity of the grammar will be dominated by the items contributed by this (potentially infinite) flow. Of course, taking the limit might not be a trivial operation -it might be impossible to separate the 'grammatical' from the 'lexical'.[53]

Needless to say, the complexity of a rule will depend on the semantics associated with it to a con-

[52] Actually, these three features, in the obvious geometrical arrangement, are all we need: the other features will be supplied by the universal convention of adding – values.

[53] In practice, one always deals with a finite corpus, and 'productivity' is the only guideline in separating the regular from the irregular. But even productivity has to be treated very cautiously: for instance irregular formations (e.g. verbal umlaut in English) can be productive rules in the grammar of children (see Berko 1958). Once such a rule is acquired, the usual strategy is not to drop the rule later, but rather to restrict its scope. Therefore, it is possible that even irregular forms will have to be handled by rules in grammars aiming at psychological reality. On the other hand, regularities present in the language might be left unanalyzed by language learners (see e.g. Bowerman 1982), so the system will have to be set up in such a manner that brute-force listing remains a viable alternative.

siderable extent. In the case of inflectional morphology, the semantics (as given by the morphosyntactic features) is purely compositional, and therefore relatively easy to take into account. In the case of derivational morphology, the situation is much more complex, and simply listing the derived items might be a viable alternative to a rule-based solution. Finally, in the case of compounding, with its extremely irregular semantics, the evaluation measure proposed here will actually favor listing to any rule-based solution. For instance, the semantics proposed for English N_1+N_2 compounding by Kiparsky (1983) can be paraphrased as 'an N_2 that is V-ed by N_1'. The V in the formula is indeterminate: for instance a *ropeladder* is a 'ladder *made of* rope', while *manslaughter* is 'slaughter *undergone by* man'. Even if we find the appropriate V, the resulting semantics is usually extremely weak. Though *testtube* is indeed 'a tube used in test', componential analysis (in the sense of Nida 1975) of its semantics will reveal it to be 'made of glass', 'closed at one end', etc. Since this unpredictable part (roughly corresponding to the 'distinguisher' of Katz and Fodor 1963) is also part of the meaning of the compound, we have to store it somewhere in the lexicon. There seems to be no better place for this than with the compound itself, and the same goes for the particular choice of V we make. But once we have entered the compounds to the lexicon, the rule becomes redundant, and we can actually simplify the grammar by leaving it out.

After these rather general considerations, let us return to the task at hand, namely to describing the forms in (1-3). For the most part, I will take the 'agglutinating' hypothesis as my starting point, and will show step by step that a somewhat more 'inflecting' solution is in fact simpler according to the evaluation measure proposed above. First, let us isolate tense/mood from person/number.

```
(5)

       V<+U>C          VC
X11=     |  |    X21=  ||    X31=  V'-->0   A(pres)= 0
         A  k          Al

                                                      -U

       V<+U>C        V<+U>C                           CV
X12=     |  |    X22=  |  |   X32=  0       B(past)= |
         A  m          A  d                           tA
```

```
          U                 U                                  -U
          |
          VCC               CVC               CVC              CV
X13=       ||       X23=    |||       X33=    |||    C(imp)=   |
           nk               tAk               nAk              IA

           U                  U                                  -U
           |
           VC               VVCVC              VVC               CVV
X14=        |       X24=      |||     X34=     \/|    D(cond)=   ||
            k                 tAk              A k               nA
```

The operation of the system is relatively simple, once the rules of vowel harmony given in 2.3 are taken into account. X11, X12, and X22 are treated as quaternary suffixes on a par with the accusative *-at/et/ot/öt*. The (irregular) *a*-form of these suffixes shows up only after the overt tense/mood markers (which therefore bear a floating –U). The verbal system differs from the nominal system inasmuch as verbal stems bearing –U are lacking. In a few cases, e.g. *zár* 'lock', this makes it possible to distinguish conjugation from declension: *zárak* 'lock.nom.pl' vs. *zárok* 'lock.1st.sg.pres.indef'; *záram* 'my lock' vs. *zárom* 'lock.1st.sg.pres.def'; *zárad* 'your(sg) lock' vs. *zárod* 'lock.2nd.sg.pres.def'. The noun *zár* is lexically marked with a floating –U, but we have to prevent this feature from triggering Core-Neg Annihilation (2:25) in the verbal system. The easy solution would be to posit an empty element as the present indicative suffix that separates –U from the domain of person-number inflection – this could be done e.g. by specifying this element as +U. The solution I prefer makes use of zero-derivation instead: the rule converting noun-stems to verb-stems will delete the –U. The net effect is the same, but this way the aforementioned lack of –U marking on verbal stems becomes a dynamic well-formedness condition (in the sense of 2.4).

Binary suffixes are best exemplified by X13 *unk/ünk* and X33 *nak/nek* (the latter is homophonous to the dative case-ending). Ternary suffixes are also present in the verbal system, the most notable example is X23 *tok/tek/tök*. Since its distribution parallels that of *hoz/hez/höz*, the rules in 2.3 will give the right results for X23 as well. In the ECH verbal system however, no vacillation of the *hölgyekhez/hölgyekhöz* 'lady.PL.ALL' type is present. After the –U suffixes, X23 never appears as **tűrtetök, *tűrjetök, *tűrnétök*. This supports the classification (proposed by Ádám Nádasdy (pc)) of the *ö*-variants in these contexts as substandard, i.e. sub-ECH. Since the superessive case ending *on/en/ön* never shows this substandard variation, while *höz* is also acceptable in forms like *%vizhöz* (ECH *vizhez*) 'water-SUE', *%Ferihöz* (ECH *Ferihez*) 'Frankie-SUE', Nádasdy is probably right in concluding that the marginal acceptability of *höz* in such cases results from dialect merger, and, as such, has little to do with vowel harmony.[54]

Second, the most 'unmarked' of the suffixes, namely the 3rd.sg.pres.indef X31 is actually less than zero: it is a rule deleting a 'free' vowel slot (i.e. one that has no material associated to it). In other words,

[54] György Szépe (pc) notes that the *-tök* forms are regular in certain dialects other than ECH.

X31 has to be subtracted from the skeleton, rather than added to it. This is best seen if we compare it to X32, which is taken to be zero here. The forms in B32 (resp. D32) are derived by adding to the stem the suffixes given as B (resp. D) in (5): the A floating under the empty V links up, the rules of vowel harmony apply, and we get the correct surface forms without any special rule. In the case of B31 (resp. D31), the rule identified with X31 is triggered, so the floating A cannot link up. Again, vowel harmony (I-spread) derives the correct result.

Before turning to the forms in A32 and C32, let us discuss the first and second person endings first. Taking the 1st.sg.subj.2nd.obj suffix to be *lAk*, the forms in X42 are derived from /stem + X + lAk/ in a straightforward manner (X = A, B, C, D). It is noteworthy that the semantics associated with these rules is equally simple: the meaning of the inflected form is the meaning of the stem + the meaning of X + the meaning of *lAk*. Now, it is certainly possible to analyze *lAk* as *l* + *Ak* (as proposed by Szépe 1969). If we add a floating –U to *l, Ak* can in fact be identified with the 1st.sg marker X11 in (5) without any problems in the phonology. But if we want to keep the semantics as simple as the 'agglutinating minimum' used above, we will have to say that *l* is a second person object marker. This way, the meaning of /stem + X + *l* + X11/ will be the meaning of the stem + the meaning of the tense/mood marker X + 2nd person object + 1st person subject. In the light of the form X21 it is not unreasonable to say that *l* is a second person suffix, but we have to note in the lexicon the fact that this time it is an object marker, rather than a subject marker. Moreover, we have to list its distributional limitations as well, and this is precisely the sort of information that can get lost in the course of time. According to this solution, the lack of **várl, *kérl* (meaning 'he waits for you, he begs you') forms is somewhat puzzling. A more serious problem is that X11 belongs to the indefinite conjugation, so the 'agglutinating' semantics predicts *lAk*-forms to be indefinite, while second person objects are semantically definite. Syntactically, various second person pronouns govern *lAk:* of these, *te* 'you (sg)', *ti* 'you (pl)' and *bennetek* 'you (pl)' govern indefinite, while *mindnyájatok* 'you all' governs definite conjugation.

The Hungarian linguistic tradition is divided on the issue of the definiteness of *lAk:* Lotz (1962, 1967a) lists Riedl (1858), Simonyi (1895), Szinnyei (1912), Lotz (1939), Hall (1944), Majtinskaja (1955), Szabó (1955), and Tompa (1961) among those who take *lAk* to be definite, and Sylvester (1539), Szenczi Molnár (1610), Komáromi Csipkés (1655), Pereszlényi (1682), Földi (1795), Fogarasi (1843) among those who take it to be definite. This dissension is probably due to the fact that the most common second person pronouns are syntactically indefinite, though semantically definite: this makes it extremely hard to classify *lAk* with respect to definiteness. Agglutinating semantics predicts it to be indefinite, since X11 is in the indefinite conjugation, but structural considerations indicate that its place is in the definite conjugation, because, unlike the endings in the indefinite paradigm, it is governed by lexical properties of *objects* (for further discussion, see Kornai 1988, 1989). Thus we are forced to conclude that in the definite/indefinite opposition, the indefinite is truly unmarked: X11 carries no information whatsoever regarding definiteness, and it is this lack of information that is the constitutive factor of indefiniteness.

The operation of the system is a bit more complicated with vowel-initial suffixes: here we must suppose that the free final vowel in B, C, and D coalesces with the initial vowel of the quaternary suffixes X11, X12, and X22. If we take this rule

```
(6) Free vowel coalescence

            V'+V --> V       (m.i.)
```

into account, the relevant forms will be derived correctly, with the exception of B11 and D11. For these, the system would predict *vártak/kértek/tűrtek* and *várnák/kérnék/tűrnék*, respectively: of these, only *kérnék* and *tűrnék* are correct. Perhaps significantly, all the hypothetical forms appear in the paradigms elsewhere. It is tempting to speculate that other forms are used "in order to avoid homophony", and in the case of *%várnák*, this is supported by the fact that this form is problematic for the language learner, and in some (substandard) dialects it is acceptable as 1st.sg.cond.indef. However, the "avoid homophony" argument appears to be somewhat flawed in the light of the fact that the B11 forms actually surfacing are identical to B12 forms. But no matter what the cause is, the effect has to be captured in some rule. One solution would be to posit a rule turning word-final *k* into *m* in the context appropriate for B21. Since the rule will have to make reference to morphosyntactic features anyway, it is simpler to put a 1st.sg.past.indef entry *tAm* in the lexicon. This entry, being more specific than either tense or person/number suffixes, preempts the application of the suffixation rules in the manner proposed by Kiparsky (1983). Similarly, the D11 pattern can be captured by adding the 1st.sg.cond.indef entry *nék* to the lexicon: as it is fully specified for the feature I, it will not show harmonic alternation.

X13 will also undergo Free Vowel Coalescence: much as in the case of Rounding Harmony, the floating A will fail to link up with the resulting V slot. In conditional mood, the A is already linked to the first V, so the result could just as well be **várnaunk, *kérneünk*. The fact that the result is *várnánk, kérnénk* suggests that in (6) it is actually the free vowel that is preserved: the A feature in D(cond) will automatically spread and the U of X31 (if it is still there after the application of (6)) cannot link up. The ternary suffix X23 has already been discussed: X24 differs from it only in containing two extra Vs at the beginning. One of these will coalesce with the final V's of B, C, and D, but the other will have a lengthening effect: this gives us the pattern of alternation observed in B23-B24, C23-C24. The VVV cluster in D24 is simplified because the feature A cannot spread onto the third V (Hungarian has no extralong vowels) — this gives us D24 (=D23). B14 and C14 require a rule (stored with X14) that drops the free vowel (together with its floating A) of the preceding suffix.

The forms in A32 show an alternation similar to that in A24 and A34. After grave vowels, A32 appears as *ja*, and after acute vowels, it appears as *i*. Describing these facts poses a serious problem for any 'monotonic' theory of phonology in which features are never deleted, because the distribution of the alternants is clearly conditioned by the same factor that governs vowel harmony (namely, the absence vs. presence of the feature I). But the harmonic pair of *ja* would be **je*, and there is no monotonic mechanism that could derive *i* from **je*. At least one timing unit has to be deleted *together with its featural content*, which would otherwise link up with the remaining vowel slot.

If we want to provide a unified account of the *ja/i* (A32), *játok/itek* (A24), and *ják/ik* (A34) alternations (an analysis I will argue against on semantic grounds later), we are forced to the conclusion that there are *three* timing units in the underlying representation of *ja/i* of which two must go away in acute contexts. To be more precise, A24 can be taken as *ja/i* + X23, and A34 as *ja/i* + the plural *k*. Since in general X24 contains two extra V slots before *tok/tek/tök*, we should assume that in the special case A24 these slots are filled with the melodic content of *ja/i*. On the basis of the grave alternant (where nothing spreads) the melodic content must be I followed by A, and the skeleton must be CVV.

By leaving the melody unassociated we can make the Vs in the skeleton subject to (6) and derive the grave alternant of A24 without any special rule the following way. One of the Vs coalesces with the vowel of *ja*, and the other triggers lengthening. This means that the feature A will spread on the remaining slot and we derive the correct *játok*. In acute contexts, first the initial free V' of X24 coalesces with the free V' of *ja/i*, and then a rule deleting the first C of *ja/i* is triggered by the harmonic context. This rule is given (with two V slots for the time being) in (7) below.

```
(7) ja/i harmony

timing tier          CVV     V     / VC*
                        --> |    /  |  ____
segmental tier       IA      I   /    I
```

This rule will fully specify the *i* and delete the V slot following it. The full derivation, given in (8), will also involve a step of spreading I further, in accordance with the rules of ordinary vowel harmony.

```
(8)   context:        grave        acute

                        U                 U

               CV'  V'V'CVC      CV'  V'V'CVC
UR:               +    |||                 |||
               IA      tAk       IA        tAk


                     U                    U

               CV'V'CVC             CV'V'CVC
Free V Coal:        |||                  |||
               IA   tAk             IA   tAk


                     U                   U
                     |
               CV'V'CVC              V'  CVC
ja/i harmony:  ||   |||              |   |||
               IA   tAk              I   tAk

spread:         ja'tok                 itek
```

This solution trivially extends to A34 and derives *ják/ik* from /*ja/i* + *ak/ek/ok/ök*/, at least if we suppose that *ja/i* is represented with a floating –U. Therefore, from a purely phonological point of view it seems to be possible to analyze A24 and A34 as containing A32. However, as I will argue shortly, the semantics of such a rule system cannot be kept compositional and therefore we need separate lexical entries for A24 and A34. This means that the skeleton of A32 can be taken as CV (and we should formulate (7) so that it contains only one V in its structural description).

Let us begin our discussion of the semantics with the putative plural element *k* that appears in final position in every plural form. Though *k* also appears in the *lak/lek* forms discussed above and in 1st.sg.indef (X42 and X11, respectively, with the exception of B11), the existence of the nominal plural *ak/ek/ok/ök* makes it very likely that we have the same (or similar) suffix here. Vágó (1980: 3.2.7) in fact isolates this *k* from the rest of the personal suffixes *in 3rd person.* This means that we have a rule turning (actual) [–ME –YOU] forms into +PL ones by *k*-suffixation of some sort. But how do we get the 1st and 2nd person forms? Extending the scope of the rule to these would leave us with a great number of cranberry morphs, as can be seen from comparing X11 to X13, X12 to X14, X21 to X23, and X22 to

X24. Therefore, we have to have other rules that turn every item potentially subject to person-number marking into forms with actual suffixal marking.

How can a rule achieve this? Because of the Lexical Integrity Hypothesis we must assume that the operation of any rule of inflection is driven by its semantics. For instance, whenever the syntax tells us that we have to put a stem in the plural form, or whenever we wish to express the idea of plurality, we concatenate the morpheme carrying plurality with the stem and unify the morphosyntactic information of the suffix (the actual feature +PL) with the morphosyntactic information (potential +PL feature) of the stem. In the formalism proposed here, the same mechanism of graph unification, namely the recursive collapsing of adjacent tokens of the same type into a single token, can be used both for the 'syntax' and the 'semantics' of a rule, as they are both represented as labelnode graphs.

In the case of phonological representations, the temporal structure is of high importance: there 'adjacent' means 'linked to adjacent slots on the timing tier'. Unification acts on concatenated representations as the Obligatory Contour Principle (Leben 1973), of which (6) is an instance. In the case of morphosyntactic representations, however, temporal structure is of little importance: there is no way to argue that, say, definiteness has to follow number or vice versa.[55] Morphosyntactic labels will be called adjacent if they are linked to the same node, i.e. if they are sisters. This way, the dominance structure of morphosyntactic trees becomes a decisive factor in rule application. The agglutinating hypothesis means that the morphosyntactic tree is flat. For instance, if person and number features appear under a node that dominates tense features as well, we would expect personal suffixes to fuse with number suffixes and tense suffixes with the same ease. But if person/number features occupy a node of their own, with tense under a separate node, only person and number suffixes are expected to fuse.

The structure of the oppositions in the verbal system of Hungarian makes it necessary to employ a morphosyntactic feature ±D(efiniteness). If Hungarian is truly agglutinating, we should be able to put this feature under a separate node — in other words, we should isolate a morpheme carrying +D. At first sight, *ja/i* is a plausible candidate, since it is the suffix A32 of the 3rd.sg.pres.def conjugation. Compared to the (less than) zero of 3rd.sg X31, which carries the unmarked choice 3rd.sg, we might conclude that *ja/i* adds the meaning [+PRES +D]. Further, since the 3rd.sg.imp.def C32 has the 'regular' form *ja/je*, we might wish to conclude that *ja/i* is in fact neutral with respect to tense/mood and carries only the meaning [+D].

But if we note that X23 and X24 differ only in definiteness, we must conclude that definiteness is expressed by the additional V slots (= vowel lengthening effect) of X24. In fact, A24 differs from X24 only in that in the former these slots are filled by *ja/i*. Now, the meaning of *ja/i* + the meaning of X24 = the meaning of A24, at least if we wish to retain the agglutinative hypothesis. However, this equation will hold only if we substitute [+PRES], rather than [+D], for the meaning of *ja/i*.

Thus, no separate morpheme carrying [+D] can be isolated, and the morphosyntactic tree of the verbal paradigm can contain only two class nodes immediately dominated by the root. The first class node

[55] This can be seen from the highly conventionalized, but cross-linguistically extremely varied manner in which natural languages handle the sequencing of such morphemes within words (cf. 3:1C).

dominates the tense/mood features, and the second dominates the person/number/definiteness features. Since tense and mood suffixes are in complementary distribution, the ordering (linear precedence) of these features is immaterial, and the same holds for the person, number, and definiteness features, as these are realized in portmanteau suffixes. The class nodes themselves can be ordered in two ways: tense/mood will either precede or follow person/number/definiteness. Since the order of the actual suffixes is fixed, the difference between the alternatives does not correspond to any difference in meaning, and we might settle on either of them. Perhaps it is better to choose the 'iconic' order reflecting the surface order of the suffixes, but this is a matter of convention.

It must be emphasized that this situation changes radically if we take derivational, rather than inflectional, suffixes into account. For instance, the 'frequentative' suffix *gat/get* and the 'factitive' *tat/tet* are interchangeable:

(9)

A	sajtó	cserélgetteti	az edzővel	a játékosokat
the	press	change.rep.fact.A32	the coach.INS	the player.PL.ACC

'the press makes the coach repeatedly change the players'

A	sajtó	cseréltetgeti	az edzővel	a játékosokat
the	press	change.fact.rep.A32	the coach.INS	the player.PL.ACC

'the press repeatedly makes the coach change the players'

The difference in meaning is slight (for better examples, see Baker 1985), but the phenomenon is clear: with derivation, different sequencing of the same constituent parts leads to different (but equally well-formed) lemexes, but with inflection, different sequencing leads to ill-formed words.

The asymmetry of the morphosyntactic tree is reflected in (5) which contains twelve person/number/definiteness suffixes and four tense/mood suffixes. The analysis would not be complete without the special suffixes that stand in an elsewhere relationship (see Kiparsky 1973) with the general person/number/definiteness markers. These are tabulated in (10):

(10)

```
         U
         |                      /Fr                U
        CVC              C     / x
A14=    | |      A21=    |    /  C        A24= CVVCVC
        I k              sz  /   x                |||
                            /   cor            IA tAk

       CV             CVVC           CVC         CVC
A32=           A34=      |      B11= |||    B33= |||
       IA             IA k           tAm         tAk

         U
                                                  I
        CVC           CVVCC          CVC          |\
C31=    |||     D14=  ||/||     4A=  |||    D11= CVVC
        IAn           nA nk          lAk         ||/|
                                                 nA k
```

Most of these suffixes have already been discussed in the text, the rest require no special rules aside from those of vowel harmony. A21 is a phonologically conditioned allomorph of X21. Its distribution is given in the table as the context of its rule of identity. A24 and A34 are the items that could not be reduced to A32 for semantic reasons. B11, B33, and D14 could well be analyzed as B + X12, B + X11, and D + X13, respectively, but here again the semantics stands in the way of further segmentation. For those who do not care about semantics, C31 should be a warning: phonologically, it is simply a concatenation of the imperative marker C and the superessive case ending *on/en/ön!*

The picture emerging from this analysis is not a very cheerful one. We needed 17 regular suffixes (those listed in (5) plus *lAk*) and 9 irregular ones, to cover 52 paradigmatic forms. In other words, the surface complexity of the system has only been halved, while the theoretical agglutinating minimum promised a tenfold reduction. Given the structure of the oppositions, 3 person markers + 4 tense/mood markers + 2 number markers + 2 definiteness markers + *lAk* (12 suffixes in all) should be taken as the realistic agglutinating minimum – the present system is twice as complex as that. This is in part due to the great number of irregular formations, and in part to the fact that only one significant division (namely, that between tense/mood and person/number/definiteness) could be established. This result can be interpreted typologically: Hungarian is roughly halfway between agglutinating and inflecting.

Let us now turn to the discussion of the alternations governed by the final phoneme of the stem. Again, I will start with listing sample paradigms. The stems used are *lát* ‘see’, *told* ‘lengthen’, and *zongoráz-* ‘play the piano’.

	1st	*2nd*	*3rd*	number	conjugation
(11A)	látok	látsz	lát	*sg*	*indef*
	látom	látod	látja	*sg*	*def*
	látunk	láttok	látnak	*pl*	*indef*
	látjuk	látjátok	látják	*pl*	*def*
(11B)	láttam	láttál	látott	*sg*	*indef*
	láttam	láttad	látta	*sg*	*def*
	láttunk	láttatok	láttak	*pl*	*indef*
	láttuk	láttátok	látták	*pl*	*def*
(11C)	lássak	láss(ál)	lásson	*sg*	*indef*
	lássam	lás(sa)d	lássa	*sg*	*def*
	lássunk	lássatok	lássanak	*pl*	*indef*
	lássuk	lássátok	lássák	*pl*	*def*
(11D)	látnék	látnál	látna	*sg*	*indef*
	látnám	látnád	látná	*sg*	*def*
	látnánk	látnátok	látnának	*pl*	*indef*
	látnánk	látnátok	látnák	*pl*	*def*

	1st	*2nd*	*3rd*	number	conjugation
(12A)	toldok	told%(a)sz	told	*sg*	*indef*
	toldom	toldod	toldja	*sg*	*def*
	toldunk	told%(o)tok	told%(a)nak	*pl*	*indef*
	toldjuk	toldjátok	toldják	*pl*	*def*
(12B)	told%(ot)tam	told%(ot)tál	toldott	*sg*	*indef*
	told%(ot)tam	told%(ot)tad	told%(ot)ta	*sg*	*def*
	told%(ot)tunk	told%(ot)tatok	told%(ot)tak	*pl*	*indef*
	told%(ot)tuk	told%(ot)tátok	told%(ot)ták	*pl*	*def*
(12C)	toldjak	toldj(ál)	toldjon	*sg*	*indef*
	toldjam	told(ja)d	toldja	*sg*	*def*
	toldjunk	toldjatok	toldjanak	*pl*	*indef*
	toldjuk	toldjátok	toldják	*pl*	*def*
(12D)	told%(a)nék	told%(a)nál	told%(a)na	*sg*	*indef*
	told%(a)nám	told%(a)nád	told%(a)ná	*sg*	*def*
	told%(a)nánk	told%(a)nátok	told(a)nátok	*pl*	*indef*
	told%(a)nánk	told%(a)nátok	told%(a)nák	*pl*	*def*

The % sign indicates that the part in parantheses is optional only in dialects other than ECH. Of the forms listed here, only the *ál/él* of C21 and the *Ja/Je* of C22 are truly optional in ECH, although in some of the forms not listed in (12), such as 4C *toldalak/toldlak* the optionality is present within ECH. The short and the long forms of C21 and C22 have different stylistic value within ECH, but they are both clearly grammatical.

	1st	*2nd*	*3rd*	number	conjugation
(13A)	zongorázok	zongorázol	zongorázik	*sg*	*indef*
	zongorázom	zongorázod	zongorázza	*sg*	*def*
	zongorázunk	zongoráztok	zongoráznak	*pl*	*indef*
	zongorázzuk	zongorázzátok	zongorázzák	*pl*	*def*
(13B)	zongoráztam	zongoráztál	zongorázott	*sg*	*indef*
	zongoráztam	zongoráztad	zongorázta	*sg*	*def*
	zongoráztunk	zongoráztatok	zongoráztak	*pl*	*indef*
	zongoráztuk	zongoráztátok	zongorázták	*pl*	*def*
(13C)	zongorázzak	zongorázz(ál)	zongorázzon	*sg*	*indef*
	zongorázzam	zongoráz(za)d	zongorázza	*sg*	*def*
	zongorázzunk	zongorázzatok	zongorázzanak	*pl*	*indef*
	zongorázzuk	zongorázzátok	zongorázzák	*pl*	*def*
(13D)	zongoráznék	zongoráznál	zongorázna	*sg*	*indef*
	zongoráznám	zongoráznád	zongorázná	*sg*	*def*
	zongoráznánk	zongoráznátok	zongoráznának	*pl*	*indef*
	zongoráznánk	zongoráznátok	zongoráznák	*pl*	*def*

The first thing to be noted in (11) is the form B31. Here, instead of *t*, we now have *ott*. Since the regular alternant appears after VC-final stems, where C is a liquid, nasal, or glide, it is reasonable to suppose that the new form is the result of a resyllabification rule that inserts VC to break up the obstruent + *t* cluster. The *t* spreads on the preceding C, and the quality of V is determined by vowel harmony (ternary alternation). This can be formalized as

```
(14)               Ob
          U     |
                |
     0-->VC   /C__t
          |
          A
```

and there is little doubt that the rule covers the majority of cases. However, there are exceptions in both directions. For instance, most stems ending in -V*d* (e.g. *marad* 'remain') take *t*, while some of those ending in geminate *r* or *l* take ott (e.g. *hall* 'hear'). Certain stems, e.g. *száll* 'fly', vacillate.[56] In SLH, this phenomenon is highlighted by the behavior of the past participle, which is formally the same as

[56] Ádám Nádasdy pointed it out that such stems are generally pronounced with a short final consonant in ECH. This observation naturally leads to the hypothesis that the vacillating behavior is in fact a reflex of cross-dialectal variation.

B13. But the 3rd.sg.past.indef form of *áll* 'stay' is *állt*, while the past participle is *állott* 'stale', and the number of such examples runs well into the hundreds. Another fact that casts doubt on the (synchronic) reality of (14) is the existence of syllable final obstruent + *t* clusters in present-day Hungarian (cf. 2.4).

Nevertheless, I think (14) can be maintained as part of a synchronic description of Hungarian. The last objection only shows that the rule has to be restricted to verbs — this is automatically achieved if we store it together with B31. Since past participles are frequently lexicalized as adjectives (i.e. their semantics is no longer compositional), they can be safely disregarded. Moreover, the very pattern of exceptions supports the functional explanation embodied in (14): it is highly significant that the exceptions appear in those cases where regular syllable structure can be maintained by other rules, namely, by devoicing (2:39), in the case of *d* + *t* clusters, and by degemination (2:43), in the case of geminate liquids.

The second thing to be noted in (11) is the alternation under C. Here we find *t* + *j* replaced by *s* + *s*. This is an instance of a much larger pattern, part of which can be attributed to natural assimilation rules. For instance, *n* or *ny* + *j* becomes *nny, l* + *j* becomes *jj, d* or *gy* + *j* becomes *ggy* in forms under C — this is as predicted by (2:41). But coronal fricatives spread on the following C slot *only* in conjugation. Vágó (1980:36) notes the difference between *mász* + *j* = [ma:ss] 'climb 2nd.sg.imp.indef' and *húsz* + *játék* *[hu:ssa:te:k] 'twenty games', but treats "*j*-stridency assimilation" (1980:2.7) among the natural assimilation rules. Here the rule is treated as morphophonemic:

```
(15)          C       C
              |    .·
          (Z)--       I
             /\
           Ob / \
          /  cor (ant)
         Fr
```

The feature I floating with the second C slot in (15) is a true morphophoneme: it appears in the suffixes A14, A24, A34, and in the suffixes in C. But the *t* + *j* → *ss* alternation under (11C) is conditioned by a larger sequence which is a proper superset of this morphophoneme. To wit, it is conditioned by the morphophoneme C(imp). To use the examples of Vágó (1980:72) again, there are minimal pairs

ütjük/üssük 'beat.1st.pl.pres.def/imp.def'
festjük/fessük 'paint.1st.pl.pres.def/imp.def'

In substandard dialects, this morphological conditioning is lost, and we have

%üssük/fessük 'beat/paint.1st.pl.pres.def'

More interestingly, the rule is conditioned not only by C(imp), but also by the element preceding the stem-final *t*: compare

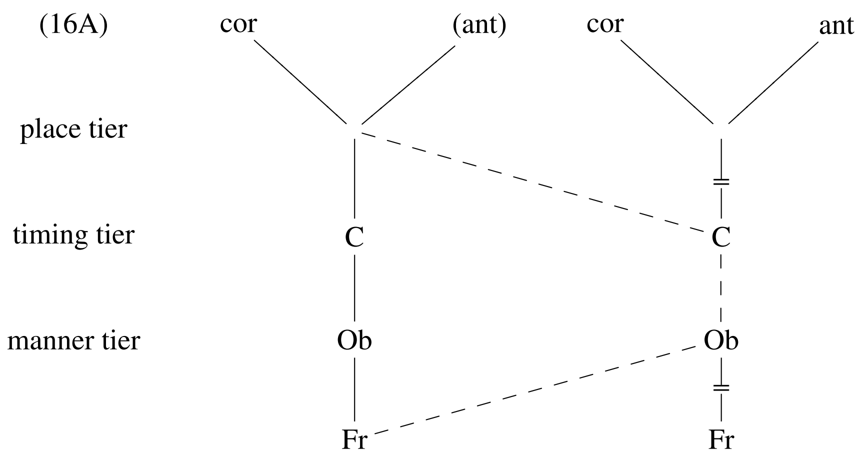
tarts [č] ‘keep.2nd.sg.imp.indef’ and
láss ‘see.2nd.sg.imp.indef’

Thus, if the element preceding *t* is a consonant, *t* + *j* stops short of complete assimilation to [š+š], rather, *t* + *š* affricates to [$t^s t^s$]. In the special case when the element preceding *t* is a coronal fricative [š] or [s], affrication is blocked and the resulting cluster looks as if the intervening *t* had dropped out and *j*-stridency assimilation had taken place:

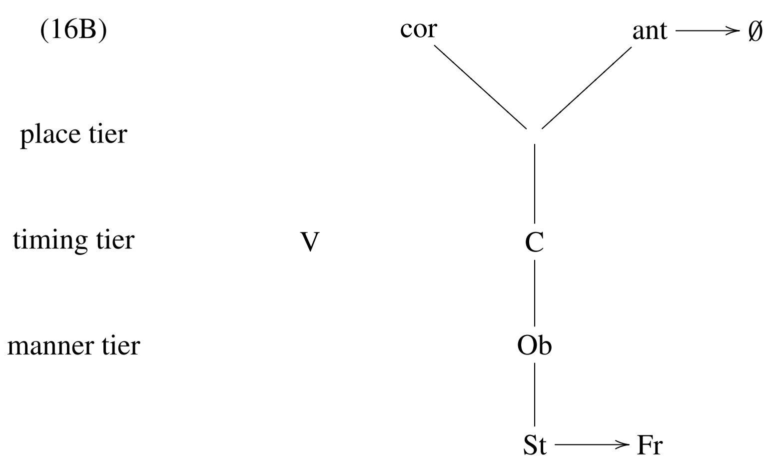
oszt+ j = ossz ‘deal’
fest+ j = fess ‘paint’

Therefore, the lexical entry of C(imp) contains not only rule (15), but also (16), which makes different adjustments in different contexts. Specifically, (15) ‘tests’ whether the preceding consonant is *t*: if it is not, the rule is inoperative. If it is *t*, (15) goes on to check the element preceding it. If this element is a vowel, or a coronal fricative, (15) replaces the feature St of *t* by Fr and spreads the place node of the preceding element on *t*. But if this element is some other consonant, (15) leaves *t* alone, and changes the first C of *ja/i* into an [š]. Diagrammatically:

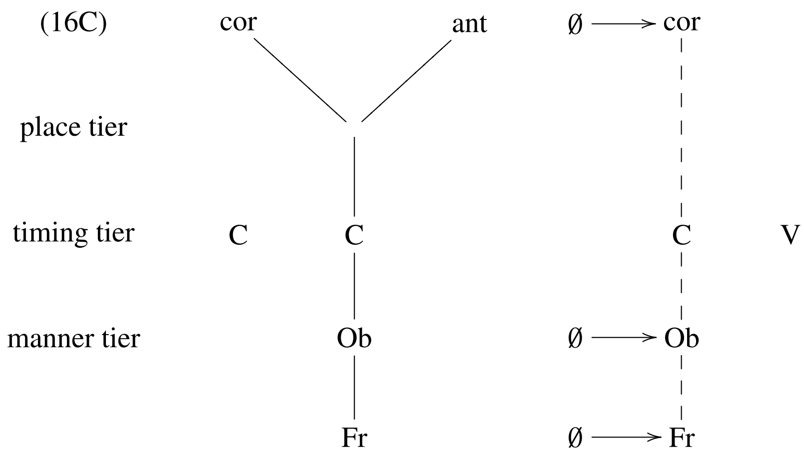
(16A)



(16B)



(16C)



(16A) is in an elsewhere relationship with (16C) — the ordering of (16B) is immaterial, since it does not compete with the other two rules. The (sub)rule(s) in (16) are stored in the lexical entry of C(imp), which also contains an instance of (15). But (15) is more general, since it makes reference only to the morphophoneme *j* of *ja/i* and the consonant immediately preceding it: therefore (16) is in an elsewhere relationship with it. Thus, (16) can feed (15), and we derive the relevant forms as follows:

(17)	gloss:	look	peek	climb	deal	paint	see	keep
	UR	néz+C	les+C	mász+C	oszt+C	fest+C	lát+C	tart+C
	(16A)	n.a.	n.a.	n.a.	ossz+C	fess+C	n.a.	n.a.
	(16B)	n.a.	n.a.	n.a.	n.a.	n.a.	lás+C	n.a.
	(16C)	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	tart+š
	(15)	nézz	less	mássz	osssz	fesss	láss	n.a.

(In the case of *oszt, fest,* and *tart,* degemination (2:43) and low-level affrication rules will eventually yield the correct output.) Some independent evidence can be adduced in favor of this analysis by taking the infinitive MPU into account. This form (and the whole infinitival conjugation) appears with an epenthetic vowel a/e after CC-final stems. This suggests a rule of A-epenthesis:

```
(18)          V     / CC   CV
         0--> |    /     __||
              A   /        nI
```

Again, this rule is best stored with MPU. Interestingly, certain stems trigger (18) although they end in a single consonant: an example is *tanít* ‘teach’, *tanitani* ‘to teach’ **tanítni*. The simplest way to account for this is to treat the last vowel of such stems as ending in C-slots, as suggested in Vágó (1987).[57] By representing *tanít* as

[57]For an alternative analysis that makes no use of C and V units see Levin 1985 and the discussion in Vágó 1989.

```
CVCVCC
||||/|
tani t
```

it will fall within the scope of (18) automatically. Since such a representation will also trigger (16C), we correctly predict that in C(imp) forms the *t* does not turn to [š]: compare *tanits* 'teach 2nd.sg.imp.indef' to *láss* (11C21) above.

This brings us to (12), which gives the paradigm of a typical CC-final stem: as can be seen, a rule very much like (14) is operative throughout (12B). This rule is given in (19) below:

```
(19)

          U

      0-->VC  /CC__t
          |
          A
```

The personal ending X21 is optional in imperative forms: this is probably due to the fact that in imperative mood it is the second, rather than the third, person which is unmarked. The definite conjugation also has a 'short form' in 2nd.sg.imp: this can be derived by an optional rule that deletes the last syllable except for its last consonant. In forms such as (12C22) *told(ja)d* the omission of the optional *ja* can also trigger a cluster simplification rule (the mirror image of (2:43)). In the dialects marked by % in (12), rule (19) is optional (cf. table 3.6 of Vágó 1980) — in ECH it is lexically governed.[58] In addition, we have a rule (20) which inserts a (binary) vowel after CC in (12D):

```
(20)


       0-->V    / CC__n
           |
           A
```

As before, (20) is best stored with D(cond), since it is inoperative outside the conjugation system *(parknál* 'park-ADE', **parkanál, *parkonál)*. The A21 alternant *sz* is equipped with a similar rule (also

[58] László Elekfi (pc) notes that the alternation pattern can differ substantially from what is presented in (12) or in Vágó 1980. For a more systematic (though somewhat normative) treatment, see Elekfi 1973. Hopefully, the ongoing *Budapest Sociolinguistic Survey* (see Kontra 1988) will provide a much more solid empirical basis for further investigations than any individual linguist's intuition could.

optional), and the same goes for the consonant-initial suffixes X23, X33. For the sake of completeness, let me tabulate here the remaining verbal suffixes:

```
(21)
                                                 U
     CV        CV        CVVC       VVC         /|
MPU= ||   MQU= ||   MRU= ||/|  MPV= \||   MQV=  VV
     ni        vA        vA n        As         \|
                                                 A


         U            U            U
         |\
MRV=  VCCVV    I1=   CVC   I21=   CVC   I31=  CVV
      ||||/          |||          |||         |||
      AndA           nAm          nAd         nIA


       U              U U              U
       |                               |
I12=  CVCC     I22=  CVCVC   I32=  CVVC
      | ||           |||||         ||  |
      n nk           nAtAk         nI  k
```

Here again, the forms of the infinitival conjugation cannot be analyzed as *ni* + person marker, since that would require definite markers for the singular elements, and indefinite markers for the plural ones.[59]

There are a great number of verbal stems that show exceptional behavior in some way — the most important classes are discussed in 4.3. From a phonological perspective, perhaps the most revealing are the ones that end in a surface vowel (*lő* 'shoot', *fő* 'cook' etc. — these are treated as 'abstract *w* stems' by Vágó (1980). Others involve some kind of dialect merger ([boča:jt/boča:t] 'forgive') or dual representation (*fürödni/fürdeni* 'bathe'). Yet others show a defective paradigm (*siklik* 'slide', *csuklik* 'hiccup') or an extreme degree of suppletion (*van/lesz* 'be')[60]. As the preceding discussion made clear, the major issue in the treatment of such irregular elements is the proper choice of diacritic marking. The limitations of the present study do not permit an individual treatment of every irregular verb — let me conclude this section with a discussion of the most characteristic, and certainly the most numerous, irregular class.

(13) above is an instance of the so-called *ik*-verbs (cf. Vágó 1980:3.3.2), which are peculiar in two respects. First, the 1st.sg.pres.indef suffix A11 is replaced by A12 (obligatorily in SLH, but only

[59]This does not mean that these forms cannot be analyzed at all: see 4.2.

[60]Following Antal (1981), in 4.3 the copula is analyzed as two independent stems.

optionally in ECH). Second, the 3rd.sg.pres.indef A31 is replaced by *ik* (this time, obligatorily both in SLH and ECH). It is trivial to capture these facts in rules — the real question is where do we store them. My suggestion is that we do *not* make use of a diacritic IK (as done by Vágó), but rather store the relevant rules with the morpheme *ik*. Let us define 'citation form' as the paradigmatic form in which all morphosyntactic features take the unmarked value (this will be 3rd.sg.pres.indef for Hungarian). The citation form will coincide with the stem for non-*ik* verbs, but will contain the *ik* for *ik*-verbs. If speakers of Hungarian store the citation form, *ik*-ness is an explicit feature of verbs.

In the diacritic-based solution, there is no reason to call the diacritic specifically IK — it could just as well be "the marker of class MCVXXI". Moreover, diacritical marking is frequently lost, while the class of *ik*-verbs appears to be quite stable. In addition to making the wrong predictions with respect to language change, the diacritic based solution fails to explain the dialectal patterning noted above. If there are two rules triggered by the same diacritic, why should we find only dialects where the first of these becomes optional (or is completely lost, as in certain sub-ECH dialects), while the second is retained, and no dialects where the second is dropped (or optional), while the first one is retained?

Storing the relevant A11→A12 rule with *ik* solves these problems completely. The *ik* class is stable because the presence of *ik* can be learned on the basis of a single instance of the 3rd.sg.pres.indef form (which is perhaps the most frequent paradigmatic form). But the A11→A12 rule is hard to learn (especially as most *ik*-verbs are intransitive and thus do not show the relevant contrast), and if someone fails to acquire it, he will automatically become a sub-SLH speaker. The optionality of this rule in ECH does not mean that ECH speakers necessarily 'have' the rule. In fact, most ECH speakers *always* use the SLH forms with the most frequent transitive *ik*-verbs like *eszik* 'eat', or *iszik* 'drink'. I suggest that this is due to the fact that during language acquisition they simply learn the paradigms of such verbs: since these stems are irregular in a number of other respects, this is all the more necessary. The result of such a piece-by-piece acquisition strategy superficially looks as if the child had the rule but applies it only in certain cases (cf. Bowerman 1982).

4.2 Declension

As can be seen from (3:4), the nominal paradigm has too many combinatorial possibilities to be tabulated in the manner of the preceding section. In order to reduce it to manageable proportions, I will first discuss the case system separately. This is made possible by the following facts:

(i) case endings are absolute word final

(ii) each paradigmatic dimension contains a zero suffix.[61]

The accusative is undoubtedly the most frequent case-ending — it is also the most complicated. In the discussion of its harmonic properties (see 2.3) it has already been mentioned that in addition to its vowel-initial forms *at/et/ot/öt*, it also has a vowelless alternant *t*. This appears regularly after stems ending in vowels or *n, ny, r, l, j, z, zs, s*, and *sz;* so the quaternary form is restricted to stems ending in noncoronal obstruents and *m*.

```
(22)

                 /lab    cor
                /   x      |
0<--VC         /    C _ ;  C _
    |         /     |      |
    t        /      Re     Fr
```

Unfortunately, there are a number of exceptions. At first sight, the pattern of exceptions seems to be predictable: for instance, CC-final stems are expected to take the quaternary vowel even if the final C is of the type where this vowel would otherwise be deleted by (22). But there are many exceptions to this subregularity: a few are listed below.

(23)

-n:	*kombájn, konszern, firn*
-l:	*bagatell, akvarell, drill, görl, moll*
-z:	*pénz* 'money',*vigyázz* 'attention'
-s:	*adjutáns, tangens, plüss*
-sz:	*fajansz, expressz, dzsessz, jassz, rassz*

Another subregularity, critically discussed in 2.6, is Esztergár's law that 'lowering' stems appear with the quaternary vowel in the accusative. One would hope that these two subregularities, albeit not perfect, will reinforce each other and in the case of lowering CC-final stems the quaternary vowel will always be present. Given that CC-final stems of the relevant type are usually fairly recent borrowings, while lowering stems generally come from the most archaic strata of Hungarian, there are only a handful of

[61] The only exception to this rule is that the suffix *-i* can be present in a few lexicalized forms. For instance, *nagybani* 'big+INE+LOC' appears only in the phrase *nagybani eladás/vétel* 'wholesale selling/buying'.

stems like *szörny* 'monster' that fall under this 'reinforced' generalization to begin with. Therefore, it is highly significant that there is at least one counterexample *hossz* 'lap' that is clearly lowering (cf. *hosszak, hosszam*) but nevertheless elides the quaternary vowel.

This, and the large number of vacillating items, show that lack of elision is lexically governed. I suggest that (22) is stored with the lexical entry of the accusative morpheme, and the exceptional forms are stored with the lexical entries of the stems in question. The latter, being more specific, will override the general rule.

Of course, it would be possible to treat this alternation with a rule of epenthesis, rather than with a rule of elision — there is nothing in the facts described above that necessitates an analysis in terms of elision. But the larger pattern of Hungarian suffixation involves a number of (quaternary) VC suffixes that do not lose the V after consonants. In addition to the (derivational) *as/es/os/ös* discussed in 2.4, we have already seen the (inflectional) X11, X12, X22, and will encounter others later in this section. From the perspective of internal reconstruction, the phenomenon of Low Vowel Lengthening also argues for a V slot in these suffixes.

```
(24) Low vowel lengthening

   V-->VV
   |   |/
   A   A
```

The scope of this rule is not entirely clear, inasmuch as not every short vowel appears in stem-final position in Hungarian. Stem-final *a* and *e* are lengthened to *á* and *é* if followed by quaternary suffixes of the VC kind, and stem-final *i, u,* and *ü* are *not* lengthened. This makes it necessary to associate A to the stem-final vowel in the structural description of (24). Hungarian has no stems ending in *ö*, so the rule simply cannot be tested. Though lengthening apparently works for o-final stems like *allegro* 'id', all such stems are loans, and the majority of ECH speakers pronounce these with a long vowel even if no suffix follows.

In addition, low vowel lengthening is also triggered by a number of consonant-initial case suffixes. In fact, the only case-ending that does not trigger it is the 'formalis' *ként*,[62] which is exceptional among cases since it contains a fully specified vowel. Thus we will have to store something like (24) with every case-ending (except for *-ként*) or we can store the rule with the CASE node dominating the cases. I opt for the latter solution, since marking one case exceptional is cheaper than repeating the rule fifteen times.

In fact, certain considerations show that *-ként* is peripheral in the case system of Hungarian. Vowel harmony is the least important of these, since there is no single alternation covering all the cases. The accusative is quaternary, the allative (ALL) and superessive (SUE) are ternary. Most other case suffixes

[62] *-kor* 'at the time', if admitted as a case ending, would also belong here.

involve binary alternation, namely *a/e* for the dative, instrumental, sublative (SBL) and inessive (INE), *ó/ő* for the delative (DEL), elative (EAL), and ablative (ABL), and *á/é* for the adessive (ADE) and translative (TRA), but the terminative (TER), causative (CAU) and formalis (FOR) do not alternate at all.

Most of the cases, however, have special pronominal forms (*engem, téged, ...* for ACC, *nekem, neked, ...* for DAT etc), the exceptions are FOR, TRA, and TER. Also, the majority of cases take part in the directional system discussed in 3.2: the list in (3:7C) can be extended to cover TRA, INS, and perhaps other cases as well. Here again, FOR is among the exceptions. Finally, for every case there are verbs that govern it: FOR appears in case-frames like *szerepel vmi-ként* 'acts as sg, plays the role of sg'. But for most cases we can find verbs where the case in question is obligatory — FOR is optional in every case-frame where it appears.

Thus, it is reasonable to conclude that FOR is 'peripheral' in the case system. But a synchronic description cannot wait until FOR becomes a full-fledged case — clearly, speakers of Hungarian were able to learn a grammar in which FOR does not trigger low vowel lengthening but cases in general do. Therefore, if we store a rule like (24) with the node CASE, the representation of *-ként* will have to be adjusted so as not to trigger this rule. Similarly, the general rule of case-marked pronoun formation has to be blocked for FOR, TRA, and TER, and has to be replaced by a special rule for ACC, which does not fit the pattern 'case-marked pronoun = case-marker + possessive suffix'. The exact nature of this adjustment is (as yet) underdetermined by the theory. The representation can be kept distinct from that of 'ordinary' elements in a number of ways, and there are only a few instances where principled choices can be made.

An alternative analysis of Low Vowel Lengthening, due to Hetzron (1972), would be to treat *a/e*-final stems as ending in underlyingly long vowels that get shortened according to the boundary information stored with the suffix. Not only does such a solution run counter to the tendency in autosegmental phonology to do away with 'segment-like' boundaries, it also creates problems for those stems that end in surface *á/é* such as *hajrá* 'finish', *burzsoá* 'bourgeois', *kávé* 'coffee', etc.[63] Since such stems are clearly in the minority, in principle there is nothing wrong with marking them as exceptions. In practice, however, we do not see such stems losing this exceptional marking at the usual high rate, which suggests that the exceptional marking is easy to learn. The best candidate for such a transparent exception marker is, of course, the long final vowel itself — the analysis based on boundary information requires some other, more *ad hoc* diacritic.

In addition to adjustments on the stem-suffix boundary, the accusative suffix triggers some stem-internal changes as well. On the one hand, ACC-suffixation can trigger vowel shortening (e.g. *nyár* 'summer' vs. *nyarat*), and, on the other hand, it can result in the loss of a stem-internal vowel (e.g. *iker* 'twin' vs. *ikret*). Superficially, the two processes are alike: both mean the deletion of a vowel slot. But a more detailed analysis shows that shortening takes place only before *-at* and *-et*, while vowel drop can happen with *-ot* and *-öt* as well. Minimal pairs such as *gyár* - *gyárat* 'factory' (**gyarat* vs. *nyarat*), *siker*

[63] Vágó (1978) makes the same argument.

- *sikert* 'success' (**sikret* vs. *ikret*) show these processes to be lexically (rather than phonologically) governed. Therefore, we have to keep the representations of *gyár* and *nyár, siker* and *iker* distinct. This is one of the cases where the theory (autosegmental phonology) narrows down our choices considerably. Instead of enlisting some ad hoc diacritic, the timing tier can do the job.

In the case of vowel drop, the alternating element cannot be a C — therefore, either the vowel is unassociated, or it is not there at all. The second possibility would mean that the 'unstable' vowel in the stem surfaces as a result of some epenthesis rule of 'default vowel insertion' in the manner suggested by Vágó (1980:4.3.3). Though the majority of unstable vowels happen to be *o*, the following examples make it clear that there can be no default vowel: *cukor* 'sugar' *cukrot, sátor* 'tent' *sátrat, kazal* 'haystack' *kazlat, bajusz* 'mustache' *bajszot, kebel* 'bosom' *keblet, tükör* 'mirror' *tükröt*. In fact, *i* and *ü* are the only vowels which are never unstable. Given the rarity of -*ü*C and -*i*C stems, these might well be accidental gaps. (The phenomenon involves less than a hundred monomorphemic stems, so it is hard to draw definite conclusions.) Now, ACC and SUE contain a rule of V'-deletion, but the other case suffixes do not. This will lead to the correct surface forms *cukrot, cukron, cukorból* etc.

As for vowel shortening, the representation of the non-alternating *á* of *gyár* has to be (25A), therefore the alternating vowel of *nyár* can be (25B), (25C), or (25D).

```
(25)
            VV                VV                VC                 VC
       (A)  |/           (B)  |           (C)  |/            (D)   |
            A                 A                 A                  A
```

However, taking the second slot to be V' would lead to a unified treatment of vowel drop and shortening. This causes no problems with the accusative suffix, but SUE triggers vowel drop only. Therefore, we have to choose (25C) or (25D) in order to avoid deriving **nyaron*. Choosing (25C) makes the prediction that noun-verbs of the relevant type should show vowel insertion in the infinitive because they trigger (18). This prediction is clearly false: *teret* 'place-ACC', *térni* 'return-infinitive' *(*téreni)*, so we have to exclude (25C) as well. This leaves us with (25D), i.e the accusative morpheme will also contain a rule of free C deletion. (Perhaps the two can be collapsed as X'→0.) [64]

With that, the description of the accusative, or rather, the description of the morphological rules stored with the accusative, is nearly complete. One could perhaps add a rule of floating U deletion to derive alternations of the *varjú/varjat* 'crow', *borjú/borjat* 'calf' type — we will see a more general form of this rule later on. I assume that the semantics associated with the accusative rule/representation, and in general the semantics of nominal inflection is essentially trivial. The meaning of the accusative form of a stem is simply the meaning of the stem plus the meaning 'accusative'. What I have said in 4.1 concerning the 'plural' meaning holds equally well for case-marking: we generate the accusative form whenever we have to (because the syntax tells us we must) and whenever the idea we wish to express is

[64] There is one stem, *lélek* 'soul', in which both rules are operative: if we take its CV skeleton to be CVV'CV'C, we will derive the correct surface forms *lelket, lelken, lélekig* etc.

carried by the ACC morpheme.[65]

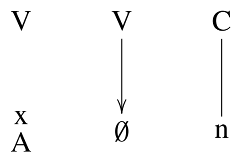
The timing tier can be used in the description of a number of other alternations. Following Vágó (1984), I will assume that the translative *vá/vé* and the instrumental *val/vel* begin in a C'. This slot will receive its specification from the stem-final consonant (cf. *házzal* 'house-INS', *kerttel* 'garden-INS', *ezredessé* 'colonel-TRA' etc). The default (floating) *v* specification will link up only after stems ending in vowels (cf. *hajóval* 'ship-INS', *békává* 'frog-TRA' etc). Vágó (1984) goes on to show that the abstract *w* of Vágó (1980) can be replaced by CV-level solutions in a number of cases, including the alternations *ló/lovat* 'horse/ACC', *falu/falvat* 'village/ACC'. In the same spirit, the 'w' class of verbal stems (e.g. *lő/lövök* 'shoot/X11' **lők, nő/növök* 'grow/X11' **nők*) can be supposed to end in a VC sequence, rather than the ordinary geminate VV. This is all the more likely in the light of the fact that verbal stems, aside from this class, all end in consonants. The 'vowel-drop' stems (*ugorj/ugrom* 'jumpC21/A32', *őrizz/őrzöm* 'guardC21/A32' etc) provide another place in the verbal system where nonstandard CV configurations can be fruitfully applied (see Vágó 1988).

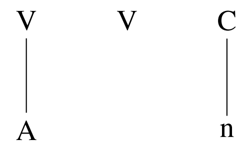
After C-final stems, the behavior of the superessive case-ending has been described by the rules of (ternary) harmonic spreading and vowel drop. After V-final stems, however, certain adjustments on the boundary are also necessary in order to derive the correct *babán* 'doll-SUE' from *baba* + *on, eskün* 'oath-SUE' from *eskü* + *ön* etc. In certain cases (e.g. *pincén* 'cellar-SUE' = *pince* + *en*) we need only a trivial rule that turns doubled vowels into geminate ones. If the final vowel of the stem is long, the superessive will preserve both the quantity and the quality of this vowel: this can be described simply by positing an alternant *-n* for SUE. However, if the stem-final vowel is short, the surface vowel preceding *-n* need not be long. In particular, it will be short if the stem ends in *i (zsenin* 'genius-SUE' *[i:]), *u*, or *ü* (as exemplified above). What these vowels have in common is the lack of specification on the A-tier: therefore the *n*-alternanat is selected in this context as well. Finally, in the case of *baba/babán*, it is the A of the preceding element, rather than the floating U of SUE that links up to the V-slot of *on/en/ön*. In other words, SUE first examines the quantity of the preceding vowel. If it is long, the V of *on/en/ön* is deleted. If it is short, SUE next examines the A tier. If there is nothing there, the V of *on/en/ön* is deleted. If there is a feature A there, then it is spread onto the next V, and the U floating over it is deleted. Finally, harmony (I and U spread) derives the correct result.

(26A)	V	V	V	C
			↓	|
			Ø	n

[65]It is far from trivial to describe this 'idea' of the accusative in terms of independent semantic primitives. For a lexical semantic approach see Wierzbicka 1980, and for a model-theoretic approach see Kornai 1988.

(26B) (26C)





With that, the description of the case system is finished: the results can be tabulated as follows:

(27)

```
                                                     I
     CV        CV        CVVC         CVC           /\
SBL= ||   ILL= ||   ADE= ||/|    INS=  ||    FOR=  CVVCC   INE= CVC
     rA        bA        nA l         vAl          ||/||         |||
                                                   kA nt         bAn


       U           U              U              U             U
                   |\             |\             |\
SUE=  VC    ABL=  CVVC     DEL=  CVVC     EAL=  CVVC     ALL= CVC
      ||          ||/|           ||/|           ||/|          |||
      An          tA l           rAl            bA l          hAz


                    I            I
                    |            |\
DAT=  CVC     TER=  VC     CAU=  VVCC     TRA= CVV    ACC= V<U>C
      |||            |           |/||           |/         |   |
      nAk            g           A rt          vA          A   t
```

Now we can turn to the next to last column of (3:4). The suffix *-é* is anaphoric: its antecedent is something in the possession of the noun to which it attaches. ANP triggers low vowel lengthening but does not cause vowel drop. After *é, i* is a plural marker: it means that the antecedent of *é* is plural. In certain dialects, *é* can be iterated, and *%Péteréé* refers to 'something in the possession of something in the possession of Péter'. ECH generally employs a single *é* in these situations as well. In yet other dialects, *é* (and *éi*) can follow *éi*, but not *é*. Dialectal, or rather, idiolectal, differences of this sort are present at other points of the nominal paradigm as well – this makes it hard to describe a unified system.

The remaining suffixes and suffix-combinations involve plurality and possessive suffixation of another sort: essentially, the possessor agrees with the possessed element in person and number. The exact nature of this possessive agreement is a heavily debated matter: see Szabolcsi 1983, Kálmán 1983, Kornai 1989, Szabolcsi 1986, and Kenesei 1986. In the preceding section, the system of morphosyntactic features was used to guide the analysis. Given the relatively clearcut nature of verbal agreement, this could be done with a measure of objectivity. Unfortunately, the morphosyntactic features given in (3:5) for nominal (possessive) agreement are far from being obvious. Rather than imposing my own solution, I will first isolate as many suffixes as phonologically possible, and discuss the morphosyntactic features afterwards. In tabulating the data, the same matrix numbering will be used as in (1-3) above. In order to distinguish the two, here the class numbers I-V will be used instead of capital A-D, and the variable running over these will be Y instead of X.

(28)

I	*1st*	*2nd*	*3rd*	possessor	possessed
	uram	urad	ura	*sg*	*sg*
	uraim	uraid	urai	*sg*	*pl*
	urunk	uratok	uruk	*pl*	*sg*
	uraink	uraitok	uraik	*pl*	*pl*
II					
	sógorom	sógorod	sógora	*sg*	*sg*
	sógoraim	sógoraid	sógorai	*sg*	*pl*
	sógorunk	sógorotok	sógoruk	*pl*	*sg*
	sógoraink	sógoraitok	sógoraik	*pl*	*pl*
III					
	emberem	embered	embere	*sg*	*sg*
	embereim	embereid	emberei	*sg*	*pl*
	emberünk	emberetek	emberük	*pl*	*sg*
	embereink	embereitek	embereik	*pl*	*pl*
IV					
	hölgyem	hölgyed	hölgye	*sg*	*sg*
	hölgyeim	hölgyeid	hölgyei	*sg*	*pl*
	hölgyünk	hölgyetek	hölgyük	*pl*	*sg*
	hölgyeink	hölgyeitek	hölgyeik	*pl*	*pl*
V					
	őröm	őröd	őre	*sg*	*sg*
	őreim	őreid	őrei	*sg*	*pl*
	őrünk	őrötök	őrük	*pl*	*sg*
	őreink	őreitek	őreik	*pl*	*pl*

The stems are: *úr* ‘master’, *sógor* ‘brother-in-law’, *ember* ‘man’, *hölgy* ’lady’, and *őr* ‘guard’. These belong in Class I, ..., Class V of 2.3. Much as in the case of verbs, the first, second, and third columns contain forms that indicate that the possessor is first, second, or third person. In odd numbered rows, the possessed is singular, and in even numbered rows, it is plural. In the first two rows, the possessor is singular, and in the next two rows, it is plural.

The plural form of the stem was discussed at length in 2.3 and 2.6. As its vowel never elides, the only thing to be added to its description is that it triggers vowel drop. The distribution of the familiar plural *ék* 'people associated to, in the family/company of' after possessive suffixes will be discussed at length below (see also Lotz 1967b). As mentioned above, this suffix can be followed by the singular and plural possessive anaphors *é* and *éi* (which can be iterated in certain dialects). The *é* of *ék* behaves just like the *é* of ANP: it does not alternate harmonically or trigger vowel drop, but it causes low vowel lengthening. This makes it possible to analyze Y44 as *é* + *k*, where *k* is the vowel-less alternant of the plural that appears regularly after vowels.

It is easy to recognize X12 and X22 in Y11 and Y21, respectively. However, the 3rd.sg.def *ja/i* (A32) does not recur here. Instead, we have a form that shows harmonic alternation between *a* and *e*. With vowel-final stems, Y31 alternates between *ja* and *je* harmonically. For consonant-final stems, the presence or absence of *j* is lexically governed: for a thorough discussion, see Papp 1975:ch. 3. Since Y31 follows the stem immediately, yotization can be safely disregarded for the purposes of segmentation.

As for the third row, X13 can be recognized in Y13. The only difference is that Y13, like every possessive suffix, triggers low vowel lengthening and vowel drop, while the lack of underlyingly V-final verbal stems makes it impossible to test X13 for these properties. The first vowel of Y32 undergoes quaternary, and the second, ternary, alternations. Finally, the vowel of Y32 undergoes binary *(u/ü)* alternation — an additional *j* appears with roughly the same elements as in the case of Y31. The vowel-sandhi effect described for the superessive in (26) is the same for Y11, Y13, Y21, and Y23 (the presence of *j* bleeds the rule for Y31 and Y33). There is a closed set of stems (*idő* 'time', *borjú* 'calf', *ajtó* 'door', and perhaps a dozen others) that have to be represented with floating, rather than fully associated, U: with these, Y31 and Y33 trigger the deletion of the U' *(ideje, borja, ajtaja)*. The *borjú* class is extra special (in the sense of Páral 1976:13 passim) inasmuch as U-deletion is also triggered by the accusative and plural suffixes, and the *j* of Y31/Y33 collapses with the *j* of the stem. I suggest that in these stems the *j* is also unassociated: this makes it possible to ascribe the facts to an OCP-like effect, in which C'V' + C'V' becomes C'V'.

The same mechanism, namely, distinguishing between fully associated, unlinked, and empty vowel slots, can be employed in describing I11, I21, I31, I12, I22, and I32 as *ni* + Y11, Y21, Y31, Y13, Y23, and Y33, respectively. If we take the feature I of *ni* to be unassociated, the representation will be distinct from that of ordinary i-final stems, and we can exploit this in the rules affixing Y11, Y13, Y21, and Y23. If *ni* does not carry the feature –U, then the resulting forms will (correctly) show only ternary alternation. Moreover, Y13 and Y23 need not behave after this 'morphophonemic' *i* as they do with other V-final stems. Thus *j*-insertion can be avoided and the binary harmony of I31 and I32 will follow naturally from the binarity of Y13 and Y23. Since the infinitival marker is in complementary distribution with tense/mood markers, it is possible to subordinate the feature INF to the same node under which tense/mood features appear in the morphosyntactic tree of verbs. In other words, +INF forms are subject to person/number marking on a par with that of the nominals: this makes the system a great deal more agglutinative. This analysis is further supported by the fact that whenever infinitival conjugation appears in auxiliary + infinitive constructions, the subject of the infinitive takes the same dative ending that appears on the possessor in other (purely nominal) possessive constructions as well.

The forms in rows 3 and 4 can be derived from those in rows 1 and 2 with a single rule which delinks the U (if any) of the first V and adds an A, and infixes an *i* after it:

```
(29)         I        / U-->A
             |       /  |           +POS
         0-->V      /   V____
```

This is a truly non-concatenative rule and the formula in (29) reflects this fact: an element is added and, at the same time, some changes take place in the triggering context. The resulting forms trigger the same sandhi rules as before. To sum up, the analysis of the 12 forms listed in (28) requires 7 suffixes. Since the theoretical minimum is $\log_2 12 = 3.58$, we can conclude that roughly the same degree of agglutination is present in both the verbal and the possessive systems.[66] This is in sharp contrast to the "toldaléktömb" approach of Tompa (1961), who, like Majtinskaja (1955), simply begs the question by leaving the forms in (28) unanalyzed.

Moreover, the rules discussed above are fully compositional (with the possible exception of *ék* = *é* + *k*) — this was made possible by following the idea of Antal (1959, 1963) to treat *i* as an infix. Mel'cuk (1972) tries to go farther in the agglutinative direction, and isolates a marker of possession that has no reference to person and number whatsoever. According to his analysis, the possessive marker is Y31. This agrees with the semantic intuition that third person forms do not conjure up the idea of a third person pronoun, while first and second person forms are clearly related to first and second person pronoun possessors. The idea that nominal and pronominal possessors should be kept separate has a lot to recommend it: in particular, the agreement properties of the two are not the same. When the possessor is a plural noun, the possessed item takes 3rd.sg agreement (*Horváthék könyve /*könyvük* 'the book of the family Horváth'). But when the possessor is the 3rd.pl pronoun, it surfaces as the 3rd.sg *(az ő könyvük* 'their book', **az ők könyvük).* Mel'cuk can avoid infixation while retaining compositionality – the trouble with his analysis is that it cannot handle harmonic alternation properly.

Rácz (1974) argues that Antal has to posit e.g. *ank* and *enk* alternants in addition to the regular *unk* and *ünk* forms of Y13. This is a serious objection in the Item-and-Arrangement framework which is taken for granted by Rácz, but makes no sense in the Item-and-Process model adopted here. In fact, the tridirectional vowel feature system employed here made it possible to capture the relevant alternations in an especially compact manner. Another argument against Antal's solution (which caused some consternation at the time, witness Berrár 1960, Szépe 1960, Tompa 1960) runs as follows. Infixes are extremely rare cross-linguistically. Hungarian has no infixes aside from this putative *i*. Therefore *i* is not an infix. Mel'cuk (1972) does not take this argument seriously (he notes that certain Indonesian languages do have infixes within other suffixes) but repeats it nevertheless. I must admit I am somewhat baffled by this sort of argumentation. In general, Hungarian contains no prefixes. The only counterexample is the

[66] Counting the combinations with FAM, ANP(PL), and CAS, the overall nominal system is of course a great deal more agglutinative than the verbal system.

superlative *leg-* (and *leges-*, which is formed by suffixing *-as/es/os/ös* to *leg*). Does it follow that *leg-* is not a prefix? Would it follow if *leges-* were not present in the system?

Yet another counterargument to infixation is based on the Y32 forms of V-final stems. In general, we have *hajói* instead of **hajójai*. There is a great deal of vacillation (*ajtói %ajtai %ajtajai* 'its doors', *ekéi %ekéjei* 'his ploughs', *lépcsői %lépcsőjei* 'its stairs', *hajdúi %hajdújai* 'his attendants'), and the *jai* forms are acceptable only after *i*-final stems in SLH. These forms appear to be problematic regardless of infixation, and the best I can suggest is a lexically governed rule that deletes *ja*.

Let us now turn to the distribution of the familiar plural. My own informal survey indicated a two-way split in judgments — I will call one variant 'dialect S', and the other one 'dialect D'. While similar 'idiosyncratic dialects' are frequently reported in the literature, Labov (1975) questions the validity of any 'dialectal' split that is not corroborated by areal or social factors. Although my data were quite clear, the reader should keep it in mind that the dialect split reported here has not yet been subjected to extensive field testing.

Table (3:4) gives the distribution of *ék* in dialect D. In dialect S there are two differences: first, *ék* can appear after the SG possessor PL possessed markers Y12, Y22, Y32 and second, *ék* cannot appear after the PL possessor SG possessed markers Y13, Y23, Y33. This is summarized in table (30) below:
(30)

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

The morphosyntactic tree given in (3:5) is adequate only in dialect D, where the familiar plural *ék* is in complementary distribution with the plural possessed forms. There it is reasonable to suppose that *i* is simply the realization of the +PL node on the second branch of (3:5), and *ék* is the realization of a feature +FAM dominated by this PL. In forms like *urainkét* 'our masters'.ACC', Y14 = *unk* spells out the first branch +POS⟨+ME +PL⟩. The infixed *i* spells out second branch, the +ANP on the the third branch is spelled out as *é*, and the ACC on the fourth branch is marked by the final *t*. From this point of view, dialect S appears to contain a surface filter that prevents *ék*-suffixation after suffixes ending in *k*. Since the *k* of the possessive suffixes cannot be isolated from the rest, just as it could not be isolated in the verbal system, this is a purely phonological rule, especially as the putative *k* would refer to the plurality of the possessor anyway.

But there is a very real sense in which dialect S is more regular than dialect D. From the perspective of (3:5), forms like *sógoraidék* are simply inexplicable: they contain a double plural which should be blocked by the same principle that blocks double datives like **Péterneknek* (see Aronoff 1976). Yet if we accept these forms, but reject those like *sógorunkék*, *ék* and *éi* will appear in the same contexts, at least if we disregard the iteration of *é(i)*, which is also subject to dialectal variation. Therefore *k* and *i* will appear as plural markers of the possessor and the possessed, respectively, and the former can even be taken as a pure plural marker that does not refer to possession at all. In this situation, *k* means 'more than one of', *é* means 'something possessed by', and *i* means 'more than one of something possessed'. This latter reading is strongly supported by the pure *i* form of Y23 after the majority of V-final stems.

Thus, *Péteréi* means '$(_1$ more than one of $(_2$ something possessed $)_1$ by $)_2$ Péter', at least if we admit the improper bracketing that depicts the collapsing of identical parts. On the other hand, *Péterék* means '$(_1$ more than one of$)_1$ $(_2$ something possessed by$)_2$ Péter'. Therefore, the paraphrases are identical but their derivations are not. In a better meaning representation we might be able to derive the meaning of *éi* without recourse to improper bracketing.

But the point of the paraphrase is clear: the different structures assigned to these suffix-combinations correspond to different derivational histories, so the difference in their meaning can be attributed to a source other than the difference of the elements. Thus, it might well be possible to assign compositional semantics to the rules generating the forms of dialect S as well, where we end up with a simpler inventory of suffixes (at the price of more complicated phonological rules). The limits to this enterprise are set by those aspects of the meaning of *ék* that cannot be derived compositionally: insofar as *ék* is more specific than *é* + *k* (e.g. because it can only refer to humans), we have to list the complex form in the lexicon.

In sum, dialect D is in no way 'deeper' than dialect S: they simply represent different solutions to the morphological puzzle of possessive forms. At present, both systems are burdened with a high number of lexically governed rules: this means that the child acquiring Hungarian will have to learn a great number of idiosyncratic forms (especially with respect to yotization). What is remarkable about this situation is that both systems generate the same forms (with the same meaning) for frequently encountered items. Therefore, they can coexist in spite of the fact that they differ markedly at the (ever fuzzy) boundary region. This is certainly a good starting point for further speculations, but I fear that until the interaction of *é(i)*-iteration, yotization, and the S/D dialectal variation have been investigated in greater detail, little more of lasting value can be said.

4.3 Implementation

The description of Hungarian morphology developed in this study is currently implemented in two computational systems using the formalism of Two-Level Morphology (Koskenniemi 1983). The first implementation, using the DKIMMO system (see Dalrymple et al 1987), was restricted to the same set of stems that were used as examples in this dissertation: see (2:17) for nouns and (3:4) for the nominal paradigm, (4:1-4) and (4:11-13) for verbs and the verbal paradigm, (2:27) *passim* for adjectives, (2:19) for numerals, and 3.2 for pronouns and other minor categories. The second implementation, using the TWOL system (see Karttunen 1989), retains a great deal of the structure of the first one, but has a much wider coverage: at present, there are more than 20,000 entries in the lexicon.

The formalism of Two-Level Morphology is very different from the one used throughout this study. Standard generative phonology takes the lexicon to be a single homogeneous entity that simply lists all kinds of unpredictable information, such as phonological form, meaning, or subcategorization, for every morpheme of the language. In more contemporary approaches, especially in Lexical Phonology (Kiparsky 1982), the lexicon has a rich internal structure, and it is possible to store words and morphemes at various locations (strata) within this complex structure. Two-Level Morphology provides a highly flexible method of organizing the lexicon into sublexicons, but does not allow the rules to take part in the same (e.g. stratal) organization as the lexical entries. In this section, I will first survey the structure of (sub)lexicons, and then discuss the rules separately.

Conceptually, the whole lexicon is organized as a large finite automaton. The lexical entries are the arcs of this automaton. A sublexicon (which is called a `LEXICON` by the system) will be a collection of arcs having a common starting point. This way, the beginning point of an arc need not be listed in the lexical entry itself, since it can be inferred from the location of the entry within the system of sublexicons. The endpoint of an arc, however, will be listed in each lexical entry. The endpoint is again a sublexicon: it is called the *Continuation Class* (abbreviated ContClass) of the entry. Notice that the arcs collected in a sublexicon need not all end in the same sublexicon — since there is a ContClass for each entry it is possible to specify different endpoints for different arcs within the same sublexicon.

The syntax of lexical entries is quite trivial. Each entry is listed as a separate record (terminated by a semicolon) which is composed of three parts: underlying form, ContClass, and gloss. Of these three parts (which must be listed in this order and must be separated by blanks), only ContClass is obligatory. In the case of zero morphemes and other empty arcs the underlying form will be missing entirely, and the entry starts with the ContClass. Glosses can also be omitted entirely (when they are present, they must be enclosed in double quotes). Thus, typical lexical entries will look like the following:

```
borju1becsina5ltQ NCON "borju1becsina1lt CVcvccvvcvcvcvvcc F0 O2 C7 D1 CZ1";
```

The underlying form, `borju1becsina5ltQ`, can contain several things, such as the abstract vowel `a5` and the diacritic `Q` in the example, which do not appear on the surface. As an aid to reading the lexical entries, the phonological encoding of vowels is given in (31) below. The ContClass `NCON` is typical, but by no means universal, for noun stems: for example, the lexical entry for the stem **gázmű* will have ContClass `PL`, thereby ensuring that that the stem will always surface with the plural suffix (cf. 3.1). The third and final part of the lexical entry, the gloss `"borju1becsina1lt CVcvccvvcvcvcvvcc F0 O2 C7 D1 CZ1"`, is completely arbitrary as far as Two-Level Morphology is concerned.[67]

(31A)

code	CVtier	Atier	Utier	Itier
a	v	a		
a1	vv	a		
u	v		u	
u1	vv		u	
i	v			i
i1	vv			i
o	v	a	u	
o1	vv	a	u	
e	v	a		i
e1	vv	a		i
o2	v	a	u	i
o3	vv	a	u	i
u2	v		u	i
u3	vv		u	i

These are the regular Hungarian vowels. As a rule, adding 4 indicates a floating –U. If there was a *U* on the U-tier, that gets moved to the CV tier — this will be indicated by the 'core-specified' V-symbol `vu` that goes to the CV tier.

(31B)

code	CVtier	Atier	Utier	Itier
a4	v	a	−u	
a5	vv	a	−u	
u4	vu		−u	
u5	vvu		−u	
i5	vv		−u	i
o4	vu	a	−u	
o5	vvu	a	−u	
e4	v	a	−u	i
e5	vv	a	−u	i
o6	vu	a	−u	i
o7	vvu	a	−u	i
u6	vu		−u	i
u7	vvu		−u	i

[67] The KIMMO1 system described here is built for research purposes and stores the surface orthographic form, the CV skeleton, frequency, etymology, and other potentially useful information in the gloss of major category stems. In a commercial system, these could be replaced by (English) glosses, semantic representations, or whatever is expedient.

It should be emphasized that no arithmetic is performed on these numbers: the system treats e.g. `u3` as a single (atomic) symbol *ű*. The other abstract vowels (see 2.3) are as follows:

```
(31C)   code  CVtier Atier Utier Itier
        i3    vvi
        i5    vvi         -u
        e3    vvi    a
        e5    vvi    a    -u
```

In the encoding of the harmonizing suffixes, we find several archiphonemes. As a rule, archiphonemes are denoted by capitals (usually the back alternant) followed by a number that indicates the length, arity, and/or roundedness of the archiphoneme.

```
(31D)   code  CVtier Atier Utier Itier
        O4    v      a     u   !This is the ternary vowel o/e/o2
        A4    vu     a         !This is the quaternary vowel a/e/o/o2
        A2           a         !Binary a/e
        A3    vv     a         !Binary a1/e1
        O2    v      a     u   !Binary o/o2
        O3    vv     a     u   !Binary o1/o3
        U2    v            u   !Binary u/u2
        U3    vv           u   !Binary u1/u3
```

For surface vowels and consonants the codes follow Hungarian (computer) orthography (see Prószéky 1985). The underlying segments will be discussed together with the rules later on.

Before turning to the description of the individual sublexicons, let me first explain in general terms how the transitions from one lexicon to another contribute to the description of morphotactics. The lexicon has only one initial state, `LEXICON ROOT`, and one final state, `LEXICON END`. In the process of morphological analysis, the aim is to describe a path from `ROOT` to `END` in the lexicon which corresponds to a parse of the surface form being analyzed. At any point in the automaton, some outgoing arc must be selected. We say that the arc was traversed successfully, iff the phonological rules in the system permit the realization of the underlying form stored with the arc as a string which matches the part of the surface form under scan. In this case, we move to the point in the lexical automaton which is listed in the ContClass of the arc successfully traversed, and advance the point of scanning to the end of the just matched part of the surface form. If the matching was not successful, we try some other arc until the whole surface string is consumed and the `END` state is reached.[68]

As an example, consider the form *magasan* 'highly'. The analysis starts at the beginning of the form, with the lexical automaton in initial state at `LEXICON ROOT`. In the Hungarian system, this

[68]In the actual implementation, backtracking happens symbol by symbol, rather than arc by arc. Although this is quite important for the efficiency of the implementation and for the proper treatment of internal sandhi, for the discussion of morphotactics it can be safely disregarded.

lexicon contains no real lexical entries, only empty ones. The ContClasses of such empty entries act as pointers to other sublexicons. In KIMMO1, these sublexicons are the following:

```
(32) N; A; V; VBAR; RE; PROF; PROP; PREV; ULTRA0; ADV; GEN; IDIO;
      ARGH; CONJ; SPECNUM; 2NUM; XNUM; XORD; CNUM; DNUM; MNUM;
```

Some of these names are self-explanatory: `N` contains the nouns, `A` the adjectives, `V` the verbs, `RE` the reflexive pronouns, `PREV` the preverbs, `ADV` the adverbials and so on. In our case, the first interesting arc is `mag` 'seed' in `LEXICON N`. Now *as/es/os/ös* is a possible continuation for nouns, but in this case a phonological rule (`"Lowering only"` see (98) below) blocks the derivation. Thus we move on to LEXICON A, which contains entries like the following:

```
(33) LEXICON A

arab BCON "arab CVvcvc F7 O6 HA02 AD02 BB02 J08 FK023 SZF20";
ferde ACON "ferde CVcvccv F6 O1 T02 HA00 AD05 BB02 J04";
fess ABCON "fess CVcvcc F1 S32 HA00 AD07 BB03";
fiatal ACON "fiatal CVcvvcvc F8 O1 D1 HA03 AD02 BB02 J07 FK032 SZF20";
maga4s ACON "magas CVcvcvc F8 O9 C1 D1 HA31 AD02 J11 FK032 SZF20";
o2reg ACON "o2reg CVvcvc F8 O9 HA04 AD04 BB45 J19 FK032 SZF20";
zulu BCON "zulu CVcvcv F0 HA03 AD03 BB02 J02 FK023";
```

This is, of course, only a random sample: the actual `LEXICON A` presently contains 2,727 entries from `abesszin` to `zso2rto6s`. The only successful match will be with `maga4s`, and we move on to its ContClass, `LEXICON ACON`.

```
(34) LEXICON ACON

COMP; AN; ATON; NINFL;
```

Here again we encounter empty lexical entries, which point us to further sublexicons. The one that will work in this case is

```
(35) LEXICON AN

An # "-ly";
```

which contains the adverbial suffix `An`, and nothing else. Since the stem has back vowels, the rules of vowel harmony permit the matching of underlying `An` with surface *an*, and the arc can be successfully traversed. The continuation class # is just an abbreviation for `END`, so the analysis is complete. However, the same analysis would not have worked for **araban*, since the ContClass of `arab` is

(36) LEXICON BCON

```
COMP; UL; ATON; NINFL;
```

which differs from ACON crucially in that it has UL instead of AN. Those stems that permit both adverbial suffixes *an/en* and *ul/ül*, such as *fitten/fittül* 'neatly', are listed with ContClass ABCON.

(37) LEXICON ABCON

```
COMP; AN; UL; ATON; NINFL;
```

In general, the Continuation Class will fully determine (modulo phonology) the set of morphemes that can be added to a lexical entry (hence the name), so there is no need to store 'rightward' subcategorization information separately. For instance, LEXICON ATON contains the category-changing (A to N) suffix sA1g, and LEXICON NINFL takes care of the cases when an adjective bears the inflection of a gapped noun as in *kékeket* 'blue.PL.ACC'.

'Leftward' subcategorization, e.g. the fact that degree suffixes can go on adjectives but not on nouns (cf. (3:8)), is encoded in a more indirect fashion. Rather than being explicitly stated in the lexical entry of the suffix, leftward subcategorization will have to be inferred from the structure of the lexicon as a whole. LEXICON COMP, which contains the suffix Abb, is accessible from the Continuation Classes of adjectives such as LEXICON ACON, LEXICON BCON, or LEXICON ABCON. However, it is *not* accessible from the ContClass of nouns, which is listed in (38).

(38) LEXICON NCON

```
FAM; # N; sA1g N "-ness"; NEND; NTOA; NTOV; PL; POS; PPP; SG;
```

The descriptive power of such a system of interlinked sublexicons will be illustrated on another set of ROOT lexicons which, taken together, handle the numerals (cardinal, ordinal, and fractional numbers) as well as their paradigmatic forms. The basic structure of numeral expressions follows a three-level pattern hundreds/tens/units as in *hat - száz - kilenc - ven - hét* 'six - hundred - nine - ty - seven'. The numbers that can appear in the first slot of this three-digit pattern are listed in (39).

(39) LEXICON MNUM

```
# 100CON "naught";
#egy 100CON "one";
#ketto3 100CON "two";
#ke1t 100CON "two";
#ha1ro4mG 100CON "three";
```

```
#ne1gy’ 100CON "four";
#o2t 100CON "five";
#hat 100CON "six";
#he1t’ 100CON "seven";
#nyo4lc 100CON "eight";
#kilenc 100CON "nine";
```

The first entry is actually empty,[69] allowing for numbers containing only two digits. Once a digit is analyzed (or generated), we move into the Continuation Class which is given by (40).

```
(40) LEXICON 100CON

#sza5z CNUM "hundred";
#sza5z DNUM "hundred";
#sza5z X2NUM "hundred";
#sza5z NUMFCON "hundred";
BIL;
BILFRAC;
```

Ignoring `BIL` and `BILFRAC` for the moment, we have to pass over `sza5z`, and move to the lexicon for the second digit:

```
(41) LEXICON CNUM

#tizen X2NUM "ten";
#huszon X2NUM "twenty";
#harminc X2NUM "thirty";
#negyvAn X2NUM "forty";
#o2tvAn X2NUM "fifty";
#hatvAn X2NUM "sixty";
#hetvAn X2NUM "seventy";
#nyolcvAn X2NUM "eighty";
#kilencvAn X2NUM "ninety";
```

At this point, we must go into

```
(42) LEXICON X2NUM

XNUM; 2NUM; XFRAC;
```

[69] The # word boundary is present in order to delimit the domain of vowel harmony.

Following the first alternative, we get into

```
(43) LEXICON XNUM

#egy NUMCON "one";
#ke1t # "two";
#ha1ro4mG NUMCON "three";
#ne1gy NUMCON "four";
#o2t NUMCON "five";
#hat NUMCON "six";
#he1t' NUMCON "seven";
#nyo4lc NUMCON "eight";
#kilenc NUMCON "nine";
```

Following the second alternative, we get into

```
(44) LEXICON 2NUM
#ketten # "two-ly";
#ke1tszer I "twice";
#ke1tszeres ACON "double";
#ke1t BIL "two";
#ketto3 2CON "two";
#kettenke1nt I "by twos";
```

The reason for treating the number 2 separately will become apparent when we turn to ordinals. When we follow the last alternative of `X2NUM`,

```
(45) LEXICON XFRAC

#egyed FRACON "oneth";
#ketted FRACON "two-th";
#harmad FRACON "third";
#negyed FRACON "fourth";
#o2to2d FRACON "fifth";
#hatod FRACON "sixth";
#heted FRACON "seventh";
#nyo4lcad FRACON "eighth";
#kilenced FRACON "ninth";
```

we get fractions conforming to the three-digit pattern such as *ezernégyszáznegyvenegyed* '1/1441'. (Notice that XFRAC is not a `ROOT` lexicon, so we will not get **egyed* as a free form.) From `100CON` we can go directly to

```
(46) LEXICON DNUM

#ti1z' NUMFCON "ten";
#hu5sz' NUMFCON "twenty";
#harminc NUMFCON "thirty";
#negyvAn NUMFCON "forty";
#o2tvAn NUMFCON "fifty";
#hatvAn NUMFCON "sixty";
#hetvAn NUMFCON "seventy";
#nyolcvAn NUMFCON "eighty";
#kilencvAn NUMFCON "ninety";
```

which corresponds to the pattern in which the last digit is 0. At this point, we exit the system of numerals by going into

```
(47) LEXICON NUMFCON

NUMCON; Ad FRACON "th";
```

The following lexicons simply list the suffixes that numerals (cardinal numbers) and fractions can take:

```
(48) LEXICON NUMCON

POS; PPP; SG; PL; FAM; NTOA; AN; BIL;
szC3r I "times";
szC3rA3s ACON "-fold";
A3nke1nt I "?an";
```

```
(49) LEXICON FRACON

NCON;
ik BCON;
re1sz NCON "part";
szC3r I "times";
```

Ordinal numbers will differ from cardinals only in the addition of the suffix *ik* ‘th’. The number 2 is irregular: we have *második* ‘second’. However, the regular *kettedik* ‘two-th’ appears in larger numbers such as *negyvenkettedik* ‘forty-two-th’, rather than **negyvenmásodik*. Thus we have

```
(50) LEXICON XORD

#elso3 BCON "first";
#ma1sod BFCON "second";
#harmad BFCON "third";
#negyed BFCON "fourth";
#o2to2d BFCON "fifth";
#hatod BFCON "sixth";
#heted BFCON "seventh";
#nyo4lcad BFCON "eighth";
#kilenced BFCON "ninth";

LEXICON BFCON

ik BCON "ord";
re1sz NCON "part";
szC3r I "times";
```

Let us now turn to `LEXICON BIL`. The three-digit pattern captured in the above system of lexicons is part of a larger pattern in which numbers are read off in groups of three digits as in *háromszázötvenhat- milliárd- kétszázhetvennégymillió- száztizenháromezer- ötszáznegyvennyolc* '356,274,113,548'. This recursion is captured by looping through

```
(51) LEXICON BIL

#ezerG XCMCON "thousand";
#millio1 XCMCON "million";
#millia1rd XCMCON "billion";
#billio1 XCMCON "trillion";
#trillio1 XCMCON "quadrillion";
#kvadrillio1 XCMCON "quintillion";
```

From such a 'round number' we can go directly to tens or units as well. This can be seen in:

```
(52) LEXICON XCMCON

X2NUM; DNUM; CNUM; MNUM; POS; PPP; SG; PL; FAM; NTOA; AN;
szC3r I "times";
Anke1nt I "?an";
```

Round numbers will also have their own fractional forms:

```
(53) LEXICON BILFRAC

#ezred FRACON "thousandth";
#milliomod FRACON "millionth";
#millia1rdod FRACON "billionth";
#billiomod FRACON "trillionth";
#trilliomod FRACON "quadrillionth";
#kvadrilliomod FRACON "quintillionth";
```

Finally, for the treatment of exceptional numbers such as 0 and 1/2 we need

```
(54) LEXICON SPECNUM

null NUMCON "zero";
nulla NUMCON "zero";
ze1ro1 NUMCON "zero";
iksz NUMFCON "x";
```

```
(55) LEXICON 2CON

POS; PPP; SG; PL; FAM; NTOA; BIL;
szC3r I "times"; szC3rA3s ACON "-fold";

LEXICON I

i2 # "locative"; #;
```

Let us now turn to the remaining `ROOT` lexicons in KIMMO1. Two of these, `ULTRA0` and `PREV`, handle prefixes. `ULTRA0` takes care of (ultra)superlatives (described in 3.2) by looping through `ULTRA` zero or more times and picking up one *leges* each time.

```
(56) LEXICON ULTRA0

ULTRA; SUP;

LEXICON ULTRA

leges ULTRA0 "ultra";
```

```
(57) LEXICON SUP

leg A "superlative";
```

When we get out of the loop, we go to `SUP`, and pick up the *leg*, at which point we move on to the adjective. Needless to say, this is all optional, inasmuch as `A` is a `ROOT` lexicon itself, so we can start there without any (ultra)superlative prefix. The optionality of verbal prefixes, which are listed in the prefix lexicon `PREV`, is handled in a slightly different manner.

```
(58) LEXICON PREV

agyon PREVC0 "38";
ala1 PREVC0 "55";
a1ltal PREVC0 "18";
a1t PREVC0 "385";
be PREVC0 "765in";
bele PREVC0 "294";
egybe PREVC0 "48";
el PREVC0 "1211";
elo3 PREVC0 "91";
elo3re PREVC0 "53";
fel PREVC0 "795up";
fe1lre PREVC0 "52";
felu2l PREVC0 "13";
ha1tra PREVC0 "66";
haza PREVC0 "61";
helyre PREVC0 "13";
hozza1 PREVC0 "71";
ide PREVC0 "64";
kereszt u2l PREVC0 "38";
kette1 PREVC0 "17";
ki PREVC0 "1232out";
ko2ru2l PREVC0 "76";
ko2zbe PREVC0 "19";
le PREVC0 "866down";
meg PREVC0 "1525perf";
melle1 PREVC0 "10";
neki PREVC0 "56";
```

```
oda PREVC0 "187";
o2ssze PREVC0 "434";
ra1 PREVC0 "273";
rajta PREVC0 "5";
sze1jjel PREVC0 "20";
szembe PREVC0 "14";
szerte PREVC0 "10";
sze1t PREVC0 "168";
tele PREVC0 "36";
tova1bb PREVC0 "34";
tu1l PREVC0 "56";
u1jra PREVC0 "12";
uta1na PREVC0 "40";
ve1gig PREVC0 "140";
vissza PREVC0 "337";
```

```
(59) LEXICON PREVC0

# V; #;
```

The alternative in (59) is set up this way because preverbs, unlike *leg*, need not be followed by a stem. When they are free forms, the second alternative of `PREVC0` is taken, and the analysis/generation is finished.

(58) lists *all* preverbs that are treated as productive in KIMMO1 (the numbers replacing the glosses show how many lexical entries contain these preverbs in the Debrecen Thesaurus). Those preverb-like elements which cannot be called productive are treated with the aid of two different `ROOT` lexicons:

```
(60) LEXICON VBAR

e1szrevesz SZCON "recognize";
ve1ghezvi SZCON "complete";
zokonvesz SZCON "resent";

LEXICON IDIO

e1szre #;
ve1ghez #;
zokon#;
```

The first of these, `VBAR`, lists the complex verb as an unanalyzed whole, while the second, `IDIO`, comes into play when negation (or an intervening auxiliary) separates the verbal stem from the prefix. `IDIO` also contains a number of other idiom fragments, such as *hetedhét*.

In general, indeclinabilia could be stored in one big `ROOT` lexicon, with all entries having # as their continuation class. However, KIMMO1 has three separate sublexicons for these. The largest one is `LEXICON ADV`, which contains 835 adverbials from `abbeli` to `vulgo`. The next one is `LEXICON ARGH` which contains 214 exclamations, curses, and other performatives from `agyo3` to `zsupsz`. Finally, there is `LEXICON CONJ`, which lists 129 connectives, namely the following:

```
(61)
a1m a1mba1r a1mba1tor a1mde ahogy aka1rcsak aka1rha alighogy
amennyiben ami1g amint aminthogy avagy azaz azonban azta1n azuta1n
ba1rha csak csakhogy csupa1n de e1s e1spedig egyre1szt ellenben
ellenkezo3leg enne1lfogva ha ha1t haba1r hacsak hacsaknem hahogy
hanem hanemha hia1ba hisz hiszen hogy hogyha hogysem holott i1gy
illeto3leg illetve inka1bb is jo1llehet ko2vetkeze1ske1pp
ko2vetkeze1ske1ppen ku2lo2nben ma1rmint ma1rpedig ma1sku2lo2nben
ma1sre1szt majd me1ghozza1 me1gis me1giscsak me1gpedig me1gse
me1gsem meg menne1l mert merthogy mi1g mi1gnem mielo3tt mihelyst
mihelyt mind mindamellett mindaza1ltal mindazona1ltal mine1l
minekelo3tte minekuta1na mint mintha minthogy mintsem miszerint
miuta1n mivel mivelhogy ne nehogy nem nemcsak nemhogy noha pedig
pediglen plusz re1szben re1szint s sem semhogy semmint so3t stb
szemben szinte1n teha1t teha1tlan tekintve tudniillik u1gy u1gyde
u1gyhogy u1gyis u1gymint u1gyse u1gysem u1gyszinte1n ugyan ugyanakkor
ugyancsak ugyanis vagy vagyis vagyishogy valamint viszont
```

In the present version of KIMMO1, these are all treated as unanalyzed. Other connectives receive an analysis through the following pair of lexicons:

```
(62) LEXICON GEN

aka1r GENCON "ever";
ba1r GENCON "any";
se GENCON "no";
vala GENCON "some";
a GENCON "th-";
GENCON;
```

```
(63) LEXICON GENCON

#;
#ki NCON "PROANIM";
#mi NCON "PROINAN";
#mennyi NUMFCON "PRONUM";
#ha1ny NUMFCON "PRONUM";
#mekkora NCON "PROADJ";
#milyen NCON "PROADJ";
#mely NCON "PROADJ";
#mikor I "PROADV";
#melyik NCON "PROADJ";
#hol # "PROADV";
#honnan # "PROADV";
#honne1t # "PROADV";
#hova # "PROADV";
#hogy AN0 "PROADV";
#merro3l # "PROADV";
#meddig # "PROADV";
#merre # "PROADV";
#mennyire # "PROADV";
```

As for the remaining minor categories, reflexive and possessive pronouns are described with the aid of `LEXICON RE` (64), pro-forms with `LEXICON PROF` (65), personal and other pronouns with `LEXICON PPOP` (66).

```
(64) LEXICON RE

magam ANP0 "1SG REFL";
magad ANP0 "2SG REFL";
maga ANP0 "3SG REFL";
magunk ANP0 "1PL REFL";
magatok ANP0 "2PL REFL";
maguk ANP0 "3PL REFL";
enye1m ANP0 "1SG POSS";
tie1d ANP0 "2SG POSS";
tied ANP0 "2SG POSS";
o2ve1 ANP0 "3SG POSS";
mie1nk ANP0 "1PL POSS";
```

```
tie1tek ANP0 "2PL POSS";
tietek ANP0 "2PL POSS";
o2ve1k ANP0 "3PL POSS";
enye1im ANP0 "1SG POSF";
tieid ANP0 "2SG POSF";
o2ve1i ANP0 "3SG POSF";
mieink ANP0 "1PL POSF";
tieitek ANP0 "2PL POSF";
o2veik ANP0 "3PL POSF";
egyma1s ANP0 "RECIP";
```

(65) LEXICON PROF

```
i1gy # "this way";
u1gy # "that way";
emi1gy # "this way";
amu1gy # "that way";
itt AN0 "here";
ott AN0 "there";
emitt AN0 "here";
amott AN0 "there";
ez0 SPCON "this";
az0 SPCON "that";
emez0 SPCON "this";
amaz0 SPCON "that";
ilyen BCON "like this";
olyan BCON "like that";
emilyen BCON "like this";
amolyan BCON "like that";
ennyi NUMFCON "this much";
annyi NUMFCON "that much";
ne1ha1ny NUMFCON "PRONUM";
kicsoda NCON "who";
micsoda NCON "what";
mindenhol # "PROADV";
mindenhonnan # "PROADV";
mindenhova # "PROADV";
mindenhogy AN0 "PROADV";
```

(66) LEXICON PROP

```
nek PER "PROP DAT";
na5l PER "PROP ADE";
to7l PER "PROP ABL";
hoz PER "PROP ALL";
rajta PER "PROP SUE";
vel PER "PROP INS";
e1rt PER "PROP CAU";
ra PER "PROP SBL";
ro1l PER "PROP DEL";
ben PER "PROP INE";
belo3l PER "PROP EAL";
bele PER "PROP ILL";
engem # "1SG ACC";
te1ged # "2SG ACC";
o3t # "3SG ACC";
minket # "1PL ACC";
titeket # "1PL ACC";
o3ket # "3PL ACC";
ala1 PERV "to under";
ele1 PERV "to before";
fele1 PERV "toward";
fo2le1 PERV "to over";
ko2ze1 PERV "to among";
melle1 PERV "to beside";
ko2re1 PERV "to around";
mo2ge1 PERV "to behind";
alatt PER0 "under";
elo3tt PER0 "before";
felett PER0 "toward";
fo2lo2tt PER0 "over";
ko2zo2tt PER0 "among";
mellett PER0 "beside";
ko2ro2tt PER "around";
mo2go2tt PER0 "behind";
alo1l PER0 "from under";
elo3l PER0 "from before";
felo3l PER0 "from toward";
```

```
fo2lu2l PER0 "from over";
ko2zu2l PER0 "from among";
mello2l PER0 "from beside";
ko2ru2l PER0 "from around";
ko2ru2lo2tt PER "around";
mo2gu2l PER0 "from behind";
uta1n PER0  "after";
ellen PER0  "against";
szerint PER0  "according to";
ira1nt PER0  "towards";
miatt PER0  "because of";
helyett PER0  "instead of";
ne1lku2l PER0  "without";
a1ltal PER0  "by";
egymaga4 PER0 "by self";
o2nmaga4 PER0 "by self";
```

The only `ROOT` lexicons not discussed so far are `LEXICON V` and `LEXICON N`. Presently there are 2,498 verb stems from `abajgat` to `zsuppol`. (67) lists the stems used in the description of the copula:

```
(67) LEXICON V

va VACON "be";
vol VOLCON "was";
lesz PRESIL "will be";
let PASTI "will be";
len CONDINF "would be";
le SZIMPI "should be";
lehet INDCON "may be";
le1gy # "be IMP 2SG";
```

Some of these, such as `va`, require special Continuation Classes:

```
(68) LEXICON VACON

gyok # "PRES 1SGI";
gy # "PRES 2SGI";
n # "PRES 3SGI";
gyunk # "PRES 1PLI";
gytok # "PRES 2PLI";
```

```
nnak # "PRES 3PLI";
gyogat INDCON "freq";
nni # "be INF";
```

Other suppletive stems, such as `vol`, are more regular inasmuch as they point to Continuation Classes which are used for the description of regular verbs as well.

```
(69) LEXICON VOLCON

PASTI; CONDI;
```

These classes, which are the regular pres.indef and cond.indef lexicons, are listed below:

```
(70) LEXICON PASTI

tAm # "PAST 1SGI";
tA1l # "PAST 2SGI";
t # "PAST 3SGI";
tUnk # "PAST 1PLI";
tAtA3k # "PAST 2PLI";
tAk # "PAST 3PLI";
```

```
(71) LEXICON CONDI

ne1k # "COND 1SGI";
nA1l # "COND 2SGI";
nA # "COND 3SGI";
nA1nk # "COND 1PLI";
nA1tA3k # "COND 2PLI";
nA1nAk # "COND 3PLI";
```

The whole system of 60+ interlinked verbal lexicons cannot be presented here — suffice it to say that all inflectional and derivational suffixes discussed in 4.1 are handled by KIMMO1. There are 11,420 noun stems in `LEXICON N` from `ablak` to `zsuzsu`. Declension, as well as the most productive cases of nominal derivation, are handled with a somewhat simpler system containing some 30 lexicons. The most important of these is `LEXICON NCON`, given in (38) above. Other important ones are listed below.

```
(72) LEXICON ANP

e1 ANP0 "ANP";
e1i ANP0 "ANPPL";
```

(73) LEXICON ANP0

```
ANP; CAS;
```

(74) LEXICON CAS

```
# "NOM";
LnAk # "DAT";
LnA1l # "ADE";
LtO1l # "ABL";
LhA3z # "ALL";
A3n # "SUE";
A4t # "ACC";
v0Al # "INS";
Le1rt # "CAU";
v0A1 # "TRA";
LrA # "SBL";
LrO1l # "DEL";
LbAn # "INE";
LbO1l # "EAL";
LbA # "ILL";
Lig # "TER";
ke1nt # "FOR";
```

(75) LEXICON CAS0

```
CAS; I;
```

LEXICON CASRED

```
LnAk # "DAT";
LnA1l # "ADE";
LtO1l # "ABL";
LhA3z # "ALL";
v0Al # "INS";
Le1rt # "CAU";
v0A1 # "TRA";
LrA # "SBL";
LrO1l # "DEL";
LbAn # "INE";
```

```
LbO1l # "EAL";
LbA # "ILL";
Lig # "TER";
ke1nt # "FOR";
ke1pp #;
ke1ppen #;
LUl #;
kor #;
```

(76) LEXICON COMC

```
i4k ANP0 "SEL";
```

LEXICON COMC0

```
COMC; NCON;
```

(77) LEXICON COMP

```
Abb COMC0 "COMP";
```

(78) LEXICON FAM

```
e5k ANP0 "PLF";
```

LEXICON NINFL

```
FAM; NEND; NTOA; PL; POS; PPP; SG;
```

LEXICON NTOA

```
A4s ACON "having";
I ACON "located at";
tAlAn BCON "without";
szeru3 ACON "like";
fe1le ACON "like";
forma ACON "like";
```

LEXICON NTOV

```
A4l VCON "to";
A4z IKCON "n-to-v";
```

```
LEXICON NEND

A4nke1nt # "per";
A4nke1nti # "per";
```

(79) LEXICON PER

```
A4m # "1SG";
A4d # "2SG";
A # "3SG";
Unk # "1PL";
A4tA3k # "2PL";
Uk # "3PL";

LEXICON PERV

m PREVC0 "1SG";
d PREVC0 "2SG";
jA PREVC0 "3SG";
nk PREVC0 "1PL";
tC3k PREVC0 "2PL";
jUk PREVC0 "3PL";

LEXICON PL

A4k ANP0 "PL";
```

(80) LEXICON POS

```
A4m POSCON "1SGPOSSG";
A4d POSCON "2SGPOSSG";
YA POSCON "3SGPOSSG";
Unk POSCON "1PLPOSSG";
A4tA3k POSCON "2PLPOSSG";
YUk POSCON "3PLPOSSG";
YAim POSCON "1SGPOSPL";
YAid POSCON "2SGPOSPL";
YAi POSCON "3SGPOSPL";
```

```
(81) LEXICON POSCON

FAM; SG;

(82) LEXICON PPP

YAink ANP0 "1PLPOSPL";
YAitA3k ANP0 "2PLPOSPL";
YAik ANP0 "3PLPOSPL";

LEXICON SG

ANP0 "SG";

LEXICON SPCON

SG; PL;
```

As the reader can see, the system describes the inflectional possibilities of dialect 'D' (see 4.2). With certain minor modifications, chiefly in `LEXICON POS` and `LEXICON PPP`, dialect 'S' could also be described.

Let us now turn to the system of rules that prescribes the relationship between surface forms and the lexical entries. While generative phonology treats the elements of the phonological alphabet as complex structures (feature matrices in SPE, and labelled graphs in autosegmental phonology), KIMMO1 is a string processor operating on unanalyzed alphabetical symbols. Although there is nothing in the basic structure of Two-Level Morphology that would preclude the use of features, in the existing implementations phonemes are treated as atomic. As we shall see, many differences between KIMMO1 and standard generative phonology stem from this fact.

Another important difference, this time essential to the structure of Two-Level Morphology, is that KIMMO1 makes a systematic distinction between the *Lexical Alphabet* and the *Surface Alphabet*. The underlying representations of generative phonology correspond to strings over the Lexical Alphabet of KIMMO1, so the terms 'underlying' and 'lexical' will be used interchangeably. In KIMMO1, the 'Surface' level is orthographic, rather than phonetic or phonological, but Hungarian orthography is almost phonemic, so this causes only a few discrepancies. But while in generative phonology there is virtually no difference[70] between 'Lexical' and 'Surface', in KIMMO1 the distinction pervades the whole system: these are the two levels of 'Two-Level Morphology'.

Generally, members of the Lexical Alphabet correspond to members of the Surface Alphabet in a one-to-one fashion: this is achieved by a general convention that acts as a system of context-free default

[70]It used to be a standard assumption that lexical representations can be underspecified but surface representations must be fully specified, but this is no longer accepted universally (see e.g. Keating 1988).

rules. Correspondences are denoted by ⟨lexical⟩:⟨surface⟩ pairs separated by a colon. For example, if underlying `p` corresponds to surface `b,` this is denoted by `p:b`. This notation is extended to cases where only one member of the correspondence is given: `p:` means a lexical `p` corresponding to something on the surface, and, similarly, `:p` means a surface p corresponding to something in the Lexical Alphabet. If no default is given, the system interprets this as identity correspondence. By declaring `p` in the alphabet we have in fact declared `p` to be a lexical character which defaulted to the same surface character. In other words, adding `p` to the alphabet is the same as adding `p:p`. The alphabet used in KIMMO1 is listed in (83).

```
(83)
Alphabet
   A:a  A:a1   a2:a  a3:a1  a4:a  a5:a1                a8:a a9:a1
   A:e  A:e1   A3:o  A3:e   A3:o2 A3:0   A4:e  A4:o   A4:a
   A1:a1 A1:e1 e2:e  e3:e1  e4:e  e5:e1  e6:e   e7:e1
   I:i         i2:i  i3:i1  i4:i  i5:i1  i6:i   i7:i1
   O:o  O1:o1  O:o2  O1:o3  o4:o  o5:o1  o6:o2  o7:o3
   C3:o C3:e   C3:o2        u4:u  u5:u1  u6:u2  u7:u3
   U:u  U1:u1  U:u2  U1:u3
   v0:v J:i Y:j F:0 W:v H:h z0:z j0:j #:0 J:z J:s J:j J:0
   a a1 b c cs ccs d dz ddz dzs ddzs e e1 f g gy ggy h i i1 j ly
   lly k l m n ny nny o o1 o2 o3 p q r s sz ssz t ty tty u u1 u2 u3
   v w x y z zs zzs ;
Diacritics
' G - / L M N Q ;
!J and a8 a9 are for ja/i
!v0 is for val/vel
!F is an imperative assimilation diacritic
!W is for abstract w nouns
!H is for silent h nouns
!z0 is for ez az.
```

The remarks preceded by the comment character '!' are ignored by the system. Since they are used only as mnemonic, they tend to be somewhat cryptic. For example, the first one means that `J` and `a8` are used in the underlying representation of *ja/i* (A32), while `a9` is used in *játok/itek* (A24) and *ják/ik* (A34) (see (1-3) in 4.1 above). In front contexts, when `J` is realized as `i`, both `a8` and `a9` are deleted. In standard generative phonology, this would require a rule such as

```
(84)      |+syll|           /  |+syll|
          |+89  |  --> 0    / _ |+high|
          |  :  |          /    |  :  |
```

where `+89` is the diacritic that distinguishes the 'morphophonemic' `a8` and `a9` from ordinary *a* and *á*. Essentially the same is expressed by the following Two-Level rule:

```
(85) "ja/i adjustment"

          [ a8:0 | a9:0 ] <=> J:i _ ;
```

It must be emphasized, however, that the colon ':' is *not* a new notation for the rewrite arrow, nor is the iff statement <=> equivalent to the slash '/' used in generative phonology for the purpose of separating the context from the rule. For the time being let us just say that the iff introduces directionality in the context statements, so that => and <= will both be well-formed context statements (with <= being closest in meaning to '/'). The colon does not have the directionality of the rewrite arrow, so in => contexts `a8:0` means *both* `a8 --> 0` and `0 --> a8`.[71]

Let us illustrate these differences on a hypothetical rule that takes `t` into `c` before high vowels. The SPE formulation of this rule would be something like

```
(86)     [-syll ]
         [+cons ]
         [-son  ]                      [+syll]
         [-voi  ]                /     [+high]
         [-cont ]  --> |-ant| / _      [  :  ]
         [+cor  ]              /
         [-lab  ]
         [+ant  ]
```

As a first approximation, this can be replaced by a KIMMO1 rule

```
(87) "Palatalization"

           t:c <= _ HighVowels
```

where HighVowels refers to a predefined set of high vowels. The most important difference between these two is that the Two-Level rule can never be opaque. High vowels always trigger the change, no matter what the other rules are, while in standard generative phonology we might find instances where a lexical `t` before a high vowel does not surface as a `c`. This will happen for instance if the high vowel is created by a rule that applies after palatalization. In Two-Level Morphology, there is no rule ordering, so this kind of opacity is impossible.

Moreover, we can guarantee that lexical `t` corresponds to surface `c` *only* before high vowels: this is expressed by

[71] In <= contexts : is not perfectly symmetrical — for a discussion see Karttunen 1983.

```
(88) "Palatalization only"

        t:c => _ HighVowels
```

In generative phonology, relationships of this kind are expressed by 'structure constraints' – but these usually refer only to one level of structure. In Two-Level Morphology it is possible to state constraints that refer to both underlying and surface structure. In fact, the rules are best viewed as constraints of this kind. The => statement protects the rule from the 'missing source' type of opacity: a surface `c` followed by a high vowel will always come from an underlying `t`. It should be emphasized that the syntax of Two-Level morphology permits *both* kinds of rules. In addition, the <=> notation is used to express the fact that both the <= and the => implications hold.

The lack of opacity has a very desirable consequence: in KIMMO1, the rules operate in the surface to lexical direction in precisely the same manner as they operate in the more usual lexical to surface direction, and the whole problem of context-sensitive parsing disappears. The price that has to be paid is that the rules interact minimally: there is no such thing as 'feeding' or 'bleeding' in Two-Level Morphology. The closest we can come is when a rule refers to surface characters in its context statement, and these are licensed (or forbidden) by some other rule that refers to lexical characters in its context specification. Rules that have to be ordered in standard generative phonology are usually replaced by rules that take their context specification from different levels.

If we extend the palatalization rule to fricatives, the SPE formulation becomes simpler: all we have to do is omit the feature `+cont` from the structural description. The corresponding Two-Level rule will be more complex notationally:

```
(89) "Stop-Fricative Palatalization"

    Dx:Py <=  _ :HighVowel  ; where Dx in (t s),
                                    Py in (c s1)
                             matched ;
```

However, readers of the Halle-Householder controversy (see Householder 1965, Chomsky and Halle 1965, Householder 1966) will recall that the complexity of a rule has little to do with the volume of ink we need to write it down. Two-Level systems compile the rules into finite transducers, and take the size of the resulting transducer as the basic measure of complexity. Since (89) can compile into a simpler automaton than (87), the apparent complexity of the notation is irrelevant.

But in generative phonology the use of features for the purpose of simplifying the rules is only a special case of the more general idea that combinations of features should express natural classes (see 2.1). Viewed from this perspective, the lack of features in Two-Level Morphology is a more serious problem, since natural classes occur in Two-Level rules just as often as in standard generative rules. In practice, the problem is solved by the possibility of explicitly listing the natural classes, as in (90) below, but the theoretical problem remains, since an 'unnatural' class is just as easy to list as a natural one. In fact KIMMO1 uses a number of 'unnatural' classes such as `l n g t`. The reason for this is that the Surface Alphabet is orthographical: these symbols are grouped together because in Hungarian they can be followed by orthographic `y` as part of a digraph.

There is another rather undesirable consequence of the lack of features. In generative phonology, features have *intrinsic content* because they serve as input for phonetic interpretation. In Two-Level Morphology, the symbols have no intrinsic content whatsoever, so the system is not specific to language the way feature-based systems are. In this connection it is worth emphasizing that the lack of features in KIMMO1 is accidental: there is nothing in the basic architecture of Two-Level Morphology that would preclude the use of standard features.

```
(90) Sets

ARaB   = A A1 O O1 U U1 A3 A4 I W ; ! lowering archiphonemes
Bu     = a a1 a4 a5 u u1 u4 u5 i2 i3 i4 i5 e2 e3 e5 o o1 o4 o5 ;
! underlying back vowels -- ce3l faze5k
!fi2ng zsi3r hami4s i5j and back +4 vowels
Bs     = a a1 u u1 o o1 ; ! surface back vowels
B5     = a o ;
B4     = a a1 ;
B3     = o 0 ;
B2     = o2 0 ;
B1     = a e o o2 ;
D      = a4 a5 u4 u5 u6 u7 i4 i5 e4 e5 o4 o5 o6 o7;!lowering vowels
E      = a4 a5 u4 u5 i4 i5 e5 o4 o5 ; ! vowels lowering to a
E1     = a a1 o o1 u u1 0 ;
T      = i i1 i2 i3 i4 i5 i6 i7 e e1 e2 e3 e4 e5 e6 e7 u u1 u2 u3
         u4 u5 u6 u7 o o1 o2 o3 o4 o5 o5 o7 a a1 a2 a3 a4 a5 A4 A3
         A1 A O1 O U1 U I ;
T1     = i i1 e e1 I ;
R      = o2 o3 u2 u3 ;
A33    = C3 A3 ;
V2     = a a1 e e1 o o1 o2 o3 u u1 u2 u3 ;
V      = a a1 i i1 e e1 o o1 o2 o3 u u1 u2 u3 y ; ! surface vowels
C      = b c cs d dz dzs f g gy h j k l ly m n ny p q r s sz t ty
         v w x z zs ccs ddz ddzs ggy lly nny ssz tty zzs
         W v0 Y H j0 J z0 ;
Cd     = b c d f g h j k l m n p r s t x z ;
Cspec  = r n j l m i1 i5 u3 F ;
```

```
Vspec  = a a1 i o o1 o2 o3 u u1 u2 e e1 ;
ConJ   = z s j 0 ; !surface realizations of J
Dias   = ' G - / # L M N Q ;
Digr   = cs dz dzs gy ly ny sz ty zs ;
DDigr  = ccs ddz ddzs ggy lly nny ssz tty zzs ;
Monogr = b c d f g h j k l m n p q r s t z ;
X      = l n g t ;
MonogZ = b c d f g h j k l m n p q r s t ;
```

In addition, Two-Level Morphology permits the use of regular expressions called *Definitions*. KIMMO1 uses the following:

```
(91)
Definitions

Clu = [C: | Dias:0 ]* ; ! consonant cluster
TrSys = [Clu [T:0 | T1: ] ]* Clu ; ! transparent syllable
GoClu = r | s | l | n | sz | z | z0: | zs | j | ly | ny | ssz | nny ;
! A4 deletes
!regularly after these, for the exceptional stems  ending in GoClu
!we use the A4-retention diacritic M
ShortenS  = A4: | U: n k | U: k | A: i | A: | Y:0 ;!stem vv shortens
DropS = A4: | A3: | U: n k | U: k | A: i | A: | Y:0 ; ! stem v drops
RetCon = b | d | f | m | n | p ; ! yot is retained
DelCon = :V | Digr | DDigr | c | g | :h | j | s | v | x | z |
         G:0 | ':0 | W:v ;
! yot is deleted
NoQ = [= - #: - Q:]* ;
NoN = [= - #: - N:]* ;
NoHa = [= -#:]* ;
```

On the one hand, the use of Definitions provides an elegant way of handling syllable structure. On the other hand, 'unnatural' Definitions are just as easy to formulate as 'natural' ones, so the objections against Sets carry over to Definitions as well. Let us begin the description of the KIMMO1 rule system with a case where the objectionable character of the rules is due entirely to the fact that we are dealing with orthography: this is the assimilation of the initial *v* of the translative and the instrumental case endings to the preceding consonant. Since the process is restricted to these two morphemes (see 4.2), we will use a special morphophoneme `v0` in the assimilation rules:

```
(92)
"v0 Assimilation"
        v0:Cx <=> \:Cx :Cx _ ;
          where Cx in Monogr ; !takes care of z0 as well

"v0 Deletion"
        v0:0  <=>  [Digr: | DDigr | :Cx :Cx ] _ ;
                where Cx in Cd ;

"v0 after  V v x w z0"
      v0:Cx <= :Cy _ ;  where Cx in (v v sz v z)
                               Cy in (V v x w z) matched ;

"v0 after W H"
      v0:v <= W: _ ;
              H: _ ;

"DDigraph Construction"
        Cx:Cy <=> _ v0:0 ;
                  where Cx in Digr
                        Cy in DDigr  matched ;
```

The same method is used in the treatment of the demonstrative pronouns *ez* 'this' and *az* 'that'. In case-marked forms, the final *z* assimilates to the following consonant. Since *z*-s in general do not assimilate, we use an underlying `z0` and the following rules:

```
(93)
"z0 to z"
         z0:z <=> _ [:V | #: | :z |A4:0 ] ;




"z0 assimilation"
        z0:Cx <=> _ Cx: ;
                    where Cx in MonogZ; !v0 shouldn't trigger as it
!is not in MonogZ
```

The complex phenomenon of alternation between *a/e* and *ja/je* in the possessive Y31 (yotization, see 4.2) is handled by a pair of rules:

```
(94)
"Yot deletion"           ! Y is deleted after DelCon or T1: C: and
                         ! between :V2 and A: unless blocked by the
                         ! retention diacritic Q.  Y is also deleted
                         ! when preceded by the deletion diacritic N.
Y:0 <= #: [NoQ & [=* [= DelCon - T1: C:]]] _ [= - U: - A:a1 - A:e1];
                    #: [NoQ & [=* :V2]]  _ [= - U: - A:a1 - A:e1 ] ;
                                    N:  _ ; ! yot deletion diacritic
"Yot retention"
   Y:j <=   #: [NoN & [=* RetCon ]]  _ ;
                                  Q: _ ;
```

The general idea is simple: Y is deleted in Deletion Contexts and retained in Retention Contexts (see DelCon and RetCon in (91) above), and can freely remain or get deleted in those few contexts which are neither. However, there are exceptions in both directions: these are marked on the exceptional stems by a deletion diacritic `N` which will force deletion even in a Retention Context, and a retention diacritic Q which will force retention even in a Deletion Context. The only noun stem mentioned so far, `borju1becsina5ltQ`, is marked by such a diacritic, because (unlike with *t*-final stems in general), here yotization is obligatory.

The rules of vowel shortening and vowel drop (see (25) above) will also use diacritics, since they are restricted to a closed set of stems.

```
(95)
"Vowel Shortening"
        Lv:Sv <=> _  Clu ’: ShortenS ;
         where Lv in (a1 a3 a5 e1 e3 e5 e7 i1 i3 i5 i7 u1 u3 u5 u7)
               Sv in (a  a  a  e  e  e  e  i  i  i  i  u  u2 u  u2)
               matched ;
       ! the ’ is a diacritic marking the shortening stem


"Vowel  Elision"
        Vx:0  <=>  _ Clu G: DropS ;
                   where Vx in V ;
!the G is a diacritic marking the stems undergoing elision
```

Since KIMMO1 aims at comprehensive coverage, it has a number of minor rules devoted to phenomena not discussed in 4.1 or 4.2. Some of these are listed in (96) below.

```
(96)
"Stem-final Lowering"
      Lv:Sv <=> _ Y:j A: ;
         where Lv in (o5 o7)
               Sv in (a e)
               matched ; !csiko5,erdo7

"W-drop"
    W:0 <=> _ [C | #: | e1 | v0:] ;



"W-stem lengthening"
    Vs:Vl <=>  _ W:0 ;
              where Vs in (a4 o4 e4 u6 o6)
                    Vl in (o1 o1 e1 u3 o3) matched ;
```

Some of the rules that deal with more general phenomena are listed in (97):

```
(97)
"Low Vowel Lengthening"
[ A:a1|A:e1|Lx:Ly ] <=> _ [ A4:0 | A3:0 | L: | Y:0 | v0:v | U:0] ;
                     where Lx  in  (a e o a8)
                           Ly in  (a1 e1 o1 a1) matched ;
!the lengthening diacritic L precedes certain suffixes like Lig but
!not others like ke1nt

"A4-deletion"                     ! A4 deletes after vowels but not
                                  ! after I:i when followed by k or s
                                  ! and after GoClu in the absence of
                                  ! certain blocking diacritics
  A4:0 <=>                                             I:i _ \[k | s] ;
                                                [:V - I:i] _  ;
           #: [[= - #: - M: - G: - ': ]* & [=* GoClu]]  _ t ;
```

Finally, there is a rather large set of rules dealing with vowel harmony:

```
(98)
"Lowering" ! A4 must not be realized as a mid vowel in a
               ! a lowering environment--it either deletes or
               ! becomes a low vowel (a or e).

[ A4:o | A4:o2 ] /<= #: [NoHa & [ =* [D: | ARaB: ] TrSys]] _ ;

"Lowering only" ! If A4 is realized as an a it must be in
                ! a lowering-to-a environment

  A4:a  =>  #: [NoHa & [ =* [E: | ARaB: ]  TrSys]] _ ;

"Back Harmony"
[ A:e | A:e1 | A1:e1 | O:o2 | O1:o3 | U:u2 | U1:u3 | A33:e | A33:o2 |
  A4:e | A4:o2 | J:i ] /<=  #: [NoHa & [ =* [Bu: | :Bs]  TrSys]] _ ;

"Back Harmony only"
      Vx:Vy   =>  #: [NoHa & [ =* [Bu: | :Bs]  TrSys]] _ ;
        where Vx in (A  A1 O O1 U U1 A33 A4 J)
              Vy in (B4 a1 o o1 u u1 o  B5 ConJ)  matched ;

"A3 to o2"
      A33:o2  <=>  :R :C Clu _ ;

"A4 to o2 only"
A4:o2 => [o2: | o3: | u2: | u3: | O:o2 | O1:o3 | A33:o2] Clu :C _;

"A4 to o2"
      A4:o2 | A4:0  <=  [o2: | o3: | u2: | u3: |
                       O:o2 | O1:o3 | C3:o2] Clu :C _ ;
!The rounding trigger must be a non-lowering underlying round vowel
!or a non-lowering archiphoneme realized as a round vowel
```

These rules are arguably more complex than the ones given in 2.3 or 2.6. The main reason for this is that the ‘if’ and ‘only if’ parts could not be collapsed because the archiphoneme A4 can be deleted in all harmonic contexts.

4.4 Conclusion

The driving force behind the analysis is provided by the twin principles of *parsimony* and *uniformity*. Parsimony is manifested in the fact that the description of Hungarian morphology presented here was based on an extremely sparse set of primitives. Inevitably, there are some atomic entities: these were called *labels*, and their range covered phonological and morphosyntactic features alike. From these entities, more complex ones can be built by arranging them into *trees*, or rather Immediate Dominance structures, since they do not involve temporal ordering, and *tiers*, which are linear structures that do have temporal structure. The parsimonious handling of primitives often resulted in solutions, such as the three-feature treatment of vowel harmony, that are unique in the sense that other solutions formulated in the same terms are necessarily more complex. In other cases, most notably in that of S/D dialectal variation, no single best solution could be found. If it turns out that the internalized grammars of speakers show the same variation in these cases, but are isomorphic to each other for the 'unique solution' cases, then autosegmental phonology is a powerful theory indeed.

With this, we come to the central claim of this thesis, namely the uniformity of rules and representations. Lexical Phonology argues that rules and lexical entries are arranged in the same (stratal) structure and that lexical entries should be treated as identity rules that stand in an elsewhere relation with other, more general, rules. Here the argument is extended to minor rules, such as Low Vowel Lengthening (see 2.3, (4:22)) or quaternary vowel elision (see (4:24)), showing that these should be stored together with the morphemes that trigger them. Further, it is argued that morphemes occupying a specific paradigmatic position, such as 3rd.pl.past.indef B33, take precedence over the less specific morphemes in the paradigm, such as 3rd.pl.indef X33 (see 4.1).

The traditional view is that representations are passive entities manipulated by the rules (active entities). Much of the debate concerning rule ordering can be interpreted as a discussion of the *control structure* imposed upon these rules. The views presented here call for distributed control: there is no external rule ordering supervising the derivations, nor are there general principles that select one particular ordering from the possibilities. Rather, the elements (morphemes) themselves carry the control information that directs their processing, much as in categorial grammar, and the only general principle that applies is 'elsewhere' ordering, which is supposed to be built into the very architecture of the system.

In must be emphasized that these views, though somewhat reminiscent of the ideas of non-procedural grammar, do *not* entail a monotonic conception of derivations, i.e. that structure is always added (unified in) and never deleted. On the contrary, 'elsewhere' ordering epitomizes the kind of non-monotonic inference mechanism where the specific rule overrides the default. Furthermore, the dissertation specifically argues that certain morphological processes, such as the harmony governed alternation of 3rd.sg.pres.def *i/ja* (see 4.1), or the infixation of the plural possessive *i* (see 4.2), are not amenable to a monotonic treatment.

The implementation discussed in 4.3 does not mirror the analysis developed in the body of the dissertation in every detail. But in spite of the different formalism, certain central ideas, such as the unified treatment of stem-triggered and suffix-triggered lowering in vowel harmony, as well as the basic structure of the paradigms, are preserved intact. Since the implementation now covers virtually every paradigmatic detail of a significant number of Hungarian stems, it can be said with a measure of confidence that the analysis, though originally developed on a very limited set of stems, carries over to Hungarian as a whole.

5 Bibliography

E. Abaffy E. 1975: Egy mai magyar hangváltozásról. *Magyar Nyelvőr* **99**, 170-174.

Abondolo, D.M. 1988: *Hungarian Inflectional Morphology*. Akadémiai, Budapest.

Ackerman, F. and Komlósi A. 1983: Some steps towards the understanding of Hungarian word order. ms, Institute of Linguistics, Budapest.

Ackerman, F. 1984: Verbal modifiers as argument taking predicates. *Groninger Arbeiten zur germanistischen Linguistik* **25**, 23-71.

Algeo, J. 1978: What consonant clusters are possible? *Word* **29**, 206-224.

Anderson, S.R. 1976: Nasal consonants and the internal structure of the segments. *Language* **52**, 326-344.

Anderson, S.R. 1981: Why phonology isn't "natural". *Linguistic Inquiry* **12**, 493-540.

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

Anderson, S.R. 1988: Inflection. In Hammond and Noonan (eds): `Theoretical Morphology`, 23-43. Orlando, FL: Academic Press.

Anderson, S.R. 2005: *Aspects of the Theory of Clitics* Oxford: Oxford University Press

Antal L. 1959: Gondolatok a magyar főnév birtokos ragozásáról. *Magyar Nyelv* **55**, 351-357.

Antal L. 1961: *A magyar esetrendszer.* Nyelvtudományi Értekezések **29** Akadémiai, Budapest.

Antal L. 1963: The possessive form of the Hungarian noun. *Linguistics 3*, 50-61.

Antal L. 1981: *Egy új magyar nyelvtan felé.* Magvető, Budapest.

Archangeli, D. 1985: Yokuts harmony: evidence for coplanar representation in nonlinear phonology. *Linguistic Inquiry* **16**, 335-372.

Aronoff, M. 1976: *Word Formation in Generative Grammar.* MIT Press, Cambridge, Mass.

Baker, M. 1985: The mirror principle and morphosyntactic explanation. *Linguistic Inquiry* **16**, 373-416.

Bakos F. 1974: *Idegen Szavak és Kifejezések Szótára.* Akadémiai, Budapest.

Basboll, H. 1973: Notes on Danish consonant combinations. ARIPUC **7**, 103-142.

Bánréti Z. 1983: A megengedő kötőszavak szintaxisáról és szemantikájáról. *Nyelvtudományi Értekezések* **117**.

Becker-Makkai, V. 1970: Vowel harmony in Hungarian reexamined in the light of recent developments in phonological theory. In: Becker-Makkai (ed): *Phonological Theory: Evolution and Current Practice.* Holt, Rinehart, and Winston, New York, pp 634-648.

Berko, J. 1958: The child's learning of English morphology. *Word* **14**, 150-177.

Berrár J. 1960: Hozzászólás Antal László cikkéhez. *Magyar Nyelv* **56**, 43-46.

Birkhoff, G. and T.C. Bartee 1970: *Modern Applied Algebra.* McGraw-Hill, New York.

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.

Böhtlingk, O. 1887 (ed. transl. and comm.): *Pāṇini's Grammatik.* Leipzig (reprinted Georg Olms Verlagsbuchhandlung, Hildesheim 1964).

Bowerman, M. 1982: Reorganizational processes in lexical and syntactic development. In: Wannen and Gleitman (eds): *Language Acquisition: the State of the Art*. Cambridge University Press, 319-346.

Cardona, George 1965: On Pāṇini's morphophonemic principles. *Language* **41**, 225-237.

Cherry, C. 1956: Roman Jakobson's distinctive features as the normal coordinates of a language. In: *For Roman Jakobson*, Mouton, The Hague.

Cherry, C. 1957: *On Human Communication*. MIT Press, Cambridge, Mass.

Chomsky, N. 1956: Three models for the description of language *I.R.E. Transactions on Information Theory* **IT-2**.

Chomsky, N. 1957: *Syntactic Structures*. Mouton, The Hague.

Chomsky, N. 1965: *Aspects of the Theory of Syntax.* MIT Press, Cambridge, Mass.

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

Chomsky, N. and M. Halle 1965: Some controversial questions in phonological theory. *Journal of Linguistics* **1**, 97-138.

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

Clements, G. N. 1976: Vowel harmony in nonlinear generative phonology. In: Dressler and Pfeiffer (eds): *Phonologica* 1976 (Innsbruck).

Clements, G. N. 1985: The geometry of phonological features. *Phonology Yearbook* **2**, 126-161.

Clements, N. and S.J. Keyser 1983: *CV Phonology.* MIT Press, Cambridge, Mass.

Clements, N. and E.V. Hume 1995: The internal organization of speech sounds. In: Goldsmith (ed): *Handbook of Phonological Theory*, 245-306.

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.

Coxeter, H.S.M. 1961: *Introduction to Geometry*. Wiley, New York.

Dalrymple, M., R.M. Kaplan, L. Karttunen, K. Koskenniemi, S. Shaio, and M. Wescoat. 1987: Tools for morphological analysis. Center for the Study of Language and Information Report No. CSLI-87-108, Stanford, CA.

Delattre, P., A. Lieberman, and F. Cooper 1951: Voyelles synthétiques á deux formantes et voyelles cardinales. *Le Maitre Phonétique* **96**, 30-36.

Dressler, W.U. 1985: *Morphonology, the Dynamics of Derivation.* Karoma, Ann Arbor.

Elekfi L. 1973: A magyar hangkapcsolódások fonetikai és fonológiai szabályai. ms, Institute of Linguistics, Budapest.

Esztergár, M. 1971: *A Generative Phonology of Nouns and Vowel Harmony in Hungarian.* PhD dissertation, UC San Diego.

É.Kiss K. 1981: Structural relations in Hungarian, a "free" word order language. *Linguistic Inquiry* **12**, 185-213.

É.Kiss K. and Papp F. 1984: A dz és a dzs státusához a mai magyar fonémarendszerben. In: Szépe Gy. (ed): *Általános Nyelvészeti Tanulmányok* **XV**, Akadémiai, Budapest, 151-160.

Farkas, D. and P. Beddor 1987: Privative and equipollent backness in Hungarian. In: Bosch et al. (eds): *Parasession on Autosegmental and Metrical Phonology* CLS **23**, 90-105.

Fogarasi J. 1843: *művelt magyar nyelvtan elemi része: a magyar nyelv szelleme*. Budapest.

Földi J. 1795: Magyar nyelvkönyv avagy Grammátika. In Gulyás K. (ed): *Régi Magyar Könyvtár* **28**, Budapest, 1912.

Fudge, E.C. 1967: The nature of phonological primes. *Journal of Linguistics* **3**, 1-36.

Gazdar, G. and G. Pullum 1982: GPSG: A theoretical synopsis. IULC, Bloomington, Indiana.

Gazdar, G., E. Klein, G. Pullum, and I. Sag 1985: *Generalized Phrase Structure Grammar.* Harvard University Press.

Goldsmith, J. 1976: Autosegmental Phonology. IULC (Published by Garland Press 1979).

Goldsmith, J. 1985: Vowel harmony in Khalkha Mongolian, Yaka, Finnish and Hungarian. *Phonology Yearbook* **2**, 253-275.

Guerrsel, M. 1986: Glides in Berber and syllabicity. *Linguistic Inquiry* **17**, 1-12.

Hall, R.Jr. 1944: *Hungarian Grammar.* Language Monographs **21**.

Halle, M. 1959: *The Sound Pattern of Russian.* Mouton, The Hague.

Halle, M. 1964: On the bases of phonology. In: Fodor and Katz (eds): *The Structure of Language.* Prentice-Hall, Englewood Cliffs, 324-333.

Halle, M. and G.N. Clements 1983: *Problem Book in Phonology.* MIT Press, Cambridge, Mass.

Halle, M. and J-R. Vergnaud 1982: On the framework of autosegmental phonology. In: van der Hulst and Smith (eds): *The Structure of Phonological Representation.* Foris, Dordrecht, 65-82.

Harris, Z. 1946: From morpheme to utterance. *Language* **22**, 161-183.

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

Hetzron, R. 1972: Studies in Hungarian morphophonology. *Ural-Altaische Jahrbücher* **44**, 79-106.

Hockett, Ch. 1958: *A Course in Modern Linguistics*. Macmillan, New York.

Householder, F.W. 1952: Review of Harris. IJAL **18**, 260-268.

Householder, F.W. 1965: On some recent claims in phonological theory. *Journal of Linguistics* **1**, 13-34.

Householder, F.W. 1966: Phonological theory: a brief comment. *Journal of Linguistics* **2**, 99-100.

van der Hulst, H. 1985: Vowel Harmony in Hungarian: a comparison of segmental and autosegmental analyses. In: van der Hulst and Smith (eds): *Advances in Nonlinear Phonology*. Foris, Dordrecht, 267-303.

van der Hulst, H. and N. Smith 1982: An overwiew of autosegmental and metrical phonology. In: van der Hulst and Smith (eds): *The Structure of Phonological Representation.* Foris, Dordrecht, 1-46.

Hyman, L. 1982: The representation of nasality in Gokana. In: van der Hulst and Smith (eds): *The Structure of Phonological Representation.* Foris, Dordrecht.

Itô, J. and A. Mester 1988: Feature specification and feature predictability. Lecture held at the Stanford phonology seminar.

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

Jaeger, J.J. 1986: Concept formation as a tool for Linguistic Research. In: Ohala and Jaeger (eds): *Experimental Phonology.* Academic Press, New York, 211-237.

Jakobson, R. and Fant, G. and Halle, M. 1952: Preliminaries to speech analysis: the distinctive features and their correlates. *MIT Acoustic Laboratories Technical Report* **13**.

Jensen, J. 1978: A reply to "Theoretical implications of Hungarian Vowel Harmony". *Linguistic Inquiry* **9**, 89-97.

Jensen, J. and M. Stong-Jensen 1984: Morphology is in the lexicon! *Linguistic Inquiry* **15**, 474-498.

Juhász J., Szőke I., O. Nagy G. and Kovalovszky M. 1982: *Magyar Értelmező Kéziszótár.*[5] Akadémiai, Budapest.

Kahn, D. 1976: *Syllable-based Generalizations in English Phonology.* PhD dissertation, MIT.

Karttunen, L. 1983: KIMMO: a general morphological processor. *Texas Linguistic Forum* **22**, 165-186.

Karttunen, L. 1989: User's guide to the TwoL compiler. ms, Xerox Palo Alto Research Center.

Kassai I. 1981: A magyar beszéd hangsorépitési szabályszerüségei. *Magyar Fonetikai Füzetek* **8**, 63-85.

Katz, J. and J. Fodor 1963: The structure of a semantic theory. *Language* **39**, 170-210.

Kaye. J., J. Lowenstamm, and J-R. Vergnaud 1985: The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook* **2**, 305-328.

Kálmán L. 1983: Anti-Dühring. ms, Institute of Linguistics, Budapest.

Kálmán L. 1985: Type free context change semantics. ms, Institute of Linguistics, Budapest.

Kálmán C. Gy., Kálmán L., Nádasdy Á, and Prószéky G. 1984: Topic, focus, and auxiliaries in Hungarian. *Groninger Arbeiten zur Germanistischen Linguistik* **24**, 162-177.

Kálmán C.Gy., Kálmán L., Nádasdy Á., and Prószéky G. 1989: A magyar segédigék rendszere. In: Kiefer (ed): *Általános Nyelvészeti Tanulmányok* **XVII**, 49-103.

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

Keating, P. 1988: Underspecification in phonetics. *Phonology* **5**, 275-292.

Kenesei I. 1986: Az alárendelő mondatok. *Nyelvtudományi Közlemények* **88**, 311-350.

Kiefer F. 1970: *Swedish Morphology.* Skriptor, Stockholm.

Kim, Ch-W. 1971: Experimental phonetics. In Dingwall (ed): *A Survey of Linguistic Science.* Linguistics program, University of Maryland.

Kiparsky, P. 1968: How abstract is phonology? In: Fujimura (ed): *Three Dimensions of Linguistic Theory*, TEC Corp, Tokyo, 1973.

Kiparsky, P. 1973: "Elsewhere" in phonology. In: Anderson and Kiparsky (eds): *A Festschrift for Morris Halle.* Holt- Rinehart and Winston, New York.

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

Kiparsky, P. 1983: Word-formation and the lexicon. In: Ingemann (ed): *Proceedings of the 1982 Mid-America Linguistics Conference* University of Kansas, Lawrence.

Kiparsky, P. 1986: Morphology and grammatical relations. ms, Stanford University.

Kisseberth, Ch. 1970: On the functional unity of phonological rules. *Linguistic Inquiry* **1**, 291-306.

Klavans, J. 1985: The independence of syntax and phonology in cliticization. *Language* **61**, 95-120.

Komáromi Csipkés Gy. 1655: *Hungaria Illustrata.* Ultrajecti.

Kontra M. 1988: Budapest sociolinguistic interview. ms, Institute of Linguistics, Budapest.

Kontra, M. and C. Ringen 1986: Hungarian vowel harmony: the evidence from loanwords. *Ural-Altaische Jahrbücher* **58**, 1-14.

Kontra M. and C. Ringen 1987: Stress and harmony in Hungarian loanwords. In: Rédei (ed): *Studien zur Phonologie und Morphonologie der Uralischen Sprachen*, Verband der Wissenschaftliche Gesellschafte Österrreichs, Vienna, 81-96.

Kontra M., C. Ringen and J. Stemberger 1987: The effect of context on suffix vowel choice in Hungarian vowel harmony. *Proc. XIV International Congress of Linguists*, E. Berlin, 1987.

Kornai A. 1985: Lexical categories and X-bar features. *Acta Linguistica* **35**, 117-131.

Kornai A. 1985a: The internal structure of noun phrases in Hungarian. In: Kenesei (ed): *Approaches to Hungarian* **1**, JATE, Szeged 79-92.

Kornai A. 1986: Szótári adatbázis az akadémiai nagyszámitógépen. *Műhelymunkák a Nyelvészet és Társtudományai Köréből* **2**, 65-79.

Kornai A. 1987: Hungarian Vowel Harmony. In: M. Crowhurst (ed): *Proceedings of the 6th West Coast Conference on Formal Linguistics.* Stanford Linguistics Association, 147-161.

Kornai A. 1988: Linking Theory and Model-Theoretic Semantics: the case of the Hungarian infinitive. ms, Stanford University.

Kornai A. 1989: A főnévi csoport egyeztetése. In: Kiefer (ed): *Általános Nyelvészeti Tanulmányok* **XVII**, 183-211.

Kornai A. 1990: The Sonority Hierarchy in Hungarian. *Nyelvtudományi Közlemények* **91** 139-146

Kornai A. 1991: Hungarian Vowel Harmony. In: I. Kenesei (ed): *Approaches to Hungarian* **III**, 183-240.

Kornai A. 1993: The generative power of feature geometry. *Annals of Mathematics and Artificial Intelligence* **8**, 37-46.

Kornai A. and Kálmán L. 1989: Hungarian sentence intonation. In: van der Hulst and Smith (eds): *Autosegmental Studies on Pitch Accent.* Foris, Dordrecht, 183-195.

Koskenniemi, K. 1983: *Two-Level Morphology: a General Computational Model for Word-Form Recognition and Production.* Department of General Linguistics, University of Helsinki, Publication No. 11.

Koutsoudas, A., Sanders, G., Noll, C. 1974: The application of phonological rules. *Language* **50**, 1-29.

Kuipers, A. 1976: Typologically salient features of some Northwest Caucasian languages. *Studia Caucasica* **3**, 101-127.

Labov, W. 1975: *What is a Linguistic Fact?* Peter de Ridder Press, Liss.

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

Ladefoged, P. 1971: *Preliminaries to Linguistic Phonetics.* University of Chicago Press.

Lapointe, S.G. 1980: *A Theory of Grammatical Agreement.* PhD dissertation, University of Massachusetts, Amherst.

Lapointe, S.G. 1989: Some extensions of the Autolexical approach to structural mismatches. In: Huck and Ojeda (eds): *Discontinuous Constituency*, Syntax and Semantics **20**, Academic Press, New York, 151-182.

Laziczius Gy. 1944: *Fonétika*. Budapest.

Leben, W. 1973: *Suprasegmental Phonology.* IULC, Bloomington, Indiana.

Levin, J. 1985: *A Metrical Theory of Syllabicity.* PhD dissertation, MIT.

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

Lieber, R. 1987: *An Integrated Theory of Autosegmental Processes.* SUNY Press, Albany.

Lotz, J. 1939: *Das Ungarische Sprachsystem.* Stockholm.

Lotz, J. 1962: The place of the implicative -LAK form in the conjugational pattern of Hungarian. *Commentationes Fenno-Ugricae in Honorem Paavo Ravila = Mémoires de la Société Finno-Ougrienne* **75**, 317-327.

Lotz, J. 1967a: A -LAK bennfoglaló alak helye a magyar igeragozásban. *Szegedi Tanárképző Főiskola Tudományos Közleményei*.

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

B. Lőrinczy É. 1979: *A magyar mássalhangzó-kapcsolódások rendszere és törvényszerüségei*. Akadémiai, Budapest.

Majtinskaja, K.E. 1955: *Vengerskij jazyk*. Moscow.

Marantz, A. 1982: Re reduplication. *Linguistic Inquiry* **13**, 435-482.

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

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

McCarthy, J. 1979: *Formal Problems in Semitic Phonology and Morphology.* PhD dissertation, MIT.

Mel'cuk, I. 1972: On the possessive forms of the Hungarian noun. In: Kiefer and Rouwet (eds): *Generative Grammar in Europe.* Reidel, Dordrecht, 315-332.

Mohanan, K.P. 1984: *Lexical Phonology.* PhD dissertation, MIT.

Muysken, P. and H. van Riemsdijk 1985: Projecting features and featuring projections. ms.

Nádasdy Á. 1985: Segmental phonology and morphophonology. In: Kenesei (ed): *Approaches to Hungarian*, 225-246.

Neary, T.M. 1978: Phonetic feature systems for vowels. IULC.

Nida, E. 1949: *Morphology.* University of Michigan Press, Ann Arbor.

Nida, E. 1975: *Componental Analysis of Meaning*. Mouton, The Hague.

Ohala, J. 1984: Consumer's Guide to evidence in phonology. *Berkeley Cognitive Science Report* **17**, UC Berkeley.

Országh L. 1977: *Magyar-Angol Szótár.*[5] Akadémiai, Budapest.

Papp F. 1968: *A Magyar Nyelv Szóvégmutató Szótára*. Akadémiai, Budapest.

Papp F. 1975: *A Magyar főnév paradigmatikus rendszere.* Akadémiai Kiadó, Budapest.

Páral, V. 1976: *A száz-százalékos nő.* Európa, Budapest.

Pereszlényi P. 1682: *Grammatica Linguae Ungaricae*, Tyrnavia.

Phelps, E. 1978: Exceptions and vowel harmony in Hungarian. *Linguistic Inquiry* **9**, 98-105.

Postal, P. 1969: Anaphoric islands. CLS **5**, 205-239.

Prószéky G. 1985: Automatizált morfológiai elemzés a nagyszótári munkálatokban. ms, Institute of Linguistics, Budapest.

Rácz E. 1974: A birtokos személyragozásnak a birtok többességét kifejező alakrendszere. In: Rácz and Szathmári (eds): *Tanulmányok a mai magyar nyelv szófajtana és alakrendszere köréből.* Tankönyvkiadó, Budapest, 135-151.

Rennison, J. 1984: Tridirectional feature systems for vowels. *Wiener Linguistische Gazette* **33-34**.

Riedl A.M. 1858: *Magyarische Grammatik*. Wien.

Ringen, C. 1977: Vowel harmony: implications for the alternation condition. In Dressler and Pfeiffer (eds): *Phonologica 1976.* Innsbrucker Beiträge zur Sprachwissenschaft, 127-132.

Ringen, C. 1980: A concrete analysis of Hungarian vowel harmony. In Vágó (ed): *Issues in Vowel Harmony*. Benjamins, Amsterdam, 135-154.

Ringen, C. and M. Kontra 1989: Hungarian neutral vowels. *Lingua* **78**, 181-191.

Rubach, J. 1982: *Analysis of Phonological Structures.* Panstwowe Widawnictwo Naukowe, Warsaw.

Sadock, J. 1985: Autolexical Syntax. *Natural Language and Linguistic Theory* **3**, 379-439.

Sadock, J. 1989: Discontinuity in Autolexical and Autosemantic Syntax. In: Huck and Ojeda (eds): *Discontionuous Constituency*, Syntax and Semantics **20**, Academic Press, New York, 283-301.

Sadock, J. 1991: *Autolexical Syntax: a theory of parallel grammatical relations* University of Chicago Press

Sapir, E. 1925: Sound patterns in language. *Language* **1**, 37-51.

Savin, H. and T. Bever 1970: The nonperceptual reality of the phoneme. *JVLVB* **9**, 295-302.

Scalise, S. 1985: *Generative Morphology.* Foris, Dordrecht.

Schein, B. and D. Steriade 1986: On geminates. *Linguistic Inquiry* **17**, 691-744.

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

Selkirk, E. 1984: On the major class features and syllable theory. In: Aronoff and Oehrle (eds): *Language Sound Structure*. MIT Press, Cambridge, MA, 107-136.

Simonyi Zs. 1895: *Tüzetes magyar nyelvtan*. Budapest.

Siptár P. 1979: Szóvégi mássalhangzó-kapcsolatok az angolban és a magyarban. In: Fülei-Szántó – Gereben (eds): *Nyelvpedagógiai irások* **II** MKKE Nyelvi Intézet, 104-121.

Siptár P. 1980a: A note on initial clusters in English and Hungarian. *Acta Linguistica* **30**, 327-343.

Siptár P. 1980b: Megjegyzések a magyar és angol szókezdő mássalhangzókapcsolatokról. *Nyelvtudományi Közlemények* **82**, 325-337.

Siptár P. 1984: Vita a magyar magánhangzó-harmóniáról. *Magyar Nyelv* **80**, 228-238.

Staal, J.F. 1962: A method of linguistic description: the order of consonants according to Pāṇini. *Language* **38**, 1-10.

Stanley, R. 1967: Redundancy rules in phonology. *Language* **43**, 393-436.

Steinhaus, H. 1960: *Mathematical Snapshots.* Oxford University Press.

Steriade, D. 1987: Redundant values. In: Bosch, Need, and Schiller (eds): *Parasession on Autosegmental and Metrical Phonology*, CLS **23**, 339-362.

Steriade, D. 1995: Underspecification and markedness. In: Goldsmith (ed): *Handbook of Phonological Theory*, 114-174.

Sylvester J. 1539: *Gram(m)matica Hungaro-Latina in ususm Poeroru(m) Rece(n)s.* Neanesi.

Szabó D. 1955: *A mai magyar nyelv.* Budapest, Jegyzetellátó.

Szabolcsi A. 1981: The semantics of topic-focus articulation. In: Groenendijk, Janssen, and Stokhof (eds): *Formal Methods in the Study of Language.* Mathematical Centre Tracts **136**, 513-541.

Szabolcsi A. 1983: A birtokos szerkezet. Próbafejezet egy magyar grammatikából. ms, Institute of Linguistics, Budapest.

Szabolcsi A. 1986: *A birtokos szerkezet és az egzisztenciális mondatok.* PhD dissertation, Budapest.

Szenci Molnár A. 1610: *Novae Grammaticae Ungaricae Libri Duo*. Hanoveae.

Szépe Gy 1960: Hozzászólás Antal László cikkéhez. *Magyar Nyelv* **56**, 46-49.

Szépe Gy. 1969: Az alsóbb nyelvi szintek leirása. In: Károly and Telegdi (eds): *Általános Nyelvészeti Tanulmányok* **VI**, 359-466.

Szinnyei J. 1912: *Ungarische Sprachlehre.* Sammlung Gösche **595**.

Thráinsson, H. 1978: On the phonology of Icelandic preaspiration. *Nordic Journal of Linguistics* **1**, 3-54.

Tompa J. 1957: *A névszói kötőhangzó szófaj-megkülňböztető szerepe.* Nyelvtudományi Értekezések **14**, Akadémiai, Budapest.

Tompa J. 1960: Hozzászólás Antal László cikkéhez. *Magyar Nyelv* **56**, 50-51.

Tompa J. 1961 (ed): *A mai magyar nyelv rendszere.* Akadémiai Kiadó, Budapest.

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

Varga L. 1982: Két szintaktikai pozicióról. *Magyar Nyelv* **78**, 159-169.

Vasu, Ś. Ch. 1891 (ed. and transl.): *Ashṭādhyāyī of Pāṇini*. Allahabad (reprinted Delhi, Motilal Banarsidass 1962, 1980).

Vágó, R.M. 1975: *Hungarian Generative Phonology.* IULC, Bloomington, Indiana.

Vágó, R.M. 1976: Theoretical implications of Hungarian vowel harmony. *Linguistic Inquiry* **7**, 243-263.

Vágó, R.M. 1978: The lengthening of final vowels in Hungarian. *Ural-Altaische Jahrbücher* **50**, 144-148.

Vágó, R.M. 1980: *The Sound Pattern of Hungarian.* Georgetown University Press.

Vágó, R.M. 1984: Degenerative CV level units. ms, CUNY Queens College and Tel Aviv University.

Vágó, R. 1987: On the representation of length. In: Dressler, Luschützky, Pfeiffer, and Rennison (eds): *Phonologica 1984*, Cambridge University Press, 319-324.

Vágó, R. 1988: Empty consonants in the moraic phonology of Hungarian. *Acta Linguistica* **38**, to appear.

Vágó, R. 1989: Elméleti szempontok a *t*-végű igék elemzésében. In: Szűts (ed): *A magyar nyelvtudomány története*. Akadémiai, Budapest (to appear).

Veenker, W. 1968: Verzeichnis der ungarische Suffixe und Suffixkombinationen. *Mitteilungen der Societas Uralo-Altaica* **3**, Hamburg.

Wetzels, L. 1985: The historical phonology of intrusive stops: a non-linear description. *Canadian Journal of Linguistics* **30**, 285-333.

Wheeler, M.W. 1972: Distinctive features and natural classes in phonological theory. *Journal of Linguistics* **8**, 87-102.

Wierzbicka, A. 1980: *The Case for Surface Case.* Karoma, Ann Arbor.

Yip, Moira, J. Maling, and R. Jackendoff 1987: Case in tiers. *Language* **63**, 217-250.

Zwicky, A. 1985: Clitics and particles. *Language* **61**, 283-305.

Zwicky, A. 1986: Agreement features: layers or tags? *Ohio State University Working Papers in Linguistics* **32**, 146-148.