1 The Concepts ‘Formal’ and ‘Functional’

1.1 Introduction

The terms *formal* and *functional* are used in more than one sense in current literature in linguistics. Three senses of the terms have a bearing on the topic of this paper, relating to approaches in linguistics, types of explanation, and the nature of linguistic knowledge. The main aim of this paper is to plug a particular standpoint with regard to the two concepts, namely, that they should be seen as complementary rather than contending, and that an explanatory linguistic theory integrates formal and functional approaches rather than merely combines them. It is thus claimed here that linguistic knowledge integrates, rather than merely combines, both formal and functional properties. The main implications of the claim, as will be obvious after a reading of §1 is that linguistic knowledge is both *innate* and dependent on *use*, or, in other terms, *computational* and *adaptive*, and thus needs to be investigated in terms of both formal and functional methods. The paper takes up in §2 one aspect of phonological systems, namely, phonological segments, to clearly show the interaction of formal and functional principles in them. The implication of the position taken here, and briefly discussed in §3, for language education is that children construct the form and content of phonological knowledge in terms of both the innate, computational disposition as well as adaptive skills dependent on their experience with language.

1.2 Linguistic Approaches

Current approaches to the study of language can be broadly classified into two types: Formal and Functional (see e.g. Newmeyer 1998). An exhaustive list of the distinct approaches should include the formalist theories of Glossematics (Hjelmslev 1943),
American Structuralism and Generative Linguistics, on the one hand, and theories of Functional Sentence Perspective (e.g. Firbas 1992), Systemic Functional Grammar (e.g. Halliday 1985) and Functional Grammar (e.g. Givon 1979, Dik 1989, and Foley and van Valin 1985), on the other. A good analytical review of functional approaches is available in Nichols (1984). Comparative studies of the two approaches are available in Darnell et al. (1998), and Newmeyer (1998).

A comparative understanding of the two approaches is of interest not only to linguistics but also to epistemology in general, as they represent two different conceptions of linguistic knowledge. They have analogues in other areas (see e.g. Gellner 1998), such as rationalism and empiricism in philosophy and cognitive psychology, capitalism and communism in economic thought, and individualism and socialism in social action. The basic premises and methods of the formalist and functionalist approaches in linguistics are often stated, largely assumed, although not always agreed upon (see e.g. Haspelmath 2000). The following may be stated as the main theoretical issues on which the two approaches arguably differ (see also Pandey 1999).

1. **THEORETICAL GOALS:** explanation of linguistic knowledge *versus* explanation of linguistic texts.
   - **NATURE OF LINGUISTIC KNOWLEDGE:** innate and autonomous *versus* acquired through experience/use and interactive with other domains of knowledge.
   - **RESEARCH METHOD:** Natural science *versus* Social science.
   - **RELEVANT DATA:** Native speaker’s intuition *versus* linguistic structures.
   - **NATURE OF EXPLANATION:** Deductive-Nomological *versus* teleological and statistical.
   - **NATURE OF ARGUMENTATION:** Deductive and based on negative evidence *versus* inductive and textual.
   - **THEORETICAL CATEGORIES (CONCEPTS AND DEVICES):** Formal (defined in terms of formal structure of linguistic units and/or their distribution) *versus* Functional (defined in terms of grammatical/semantic function).
   - **LINGUISTIC EVIDENCE:** Internal *versus* External (Use, Damage, Change, etc.).

Some of these distinctions are maintained by the practitioners of the two approaches. For instance, we find extended discussions of methodological issues in the work of Chomsky (1986, 2001), on the formalist side, and in the work of Halliday (1988) and Givon (1979), on the functionalist side. Extreme views on formalism and functionalism and their mutually opposed goals and methods have been the focus of a lot of writing, both in
linguistics and in philosophy. Thus, Itkonen (1978) argues for hermeneutics as the proper method of linguistics because of language being primarily a social semiotic system, while Dummet (1993) defends the philosophical position concerning cognition rather than communication to be the primary property of language.

However, notwithstanding the professed and acknowledged differences between the two approaches, there has been a growing realization of their mutual complementariness, apparent in at least three ways.

a. Both formal and functional approaches have been found to be valid in different domains. Thus the account of the grammatical principles of syntactic processes has been found to be immensely deepened by generative linguistic accounts, while at the same time certain syntactic phenomena, such as word order universals (e.g. Tomlin 1986), have been found to be best explainable in functionalist terms (e.g. Comrie 1981)\(^1\).

b. As some philosophers, especially Pateman (1989) and Gellner (1998), have shown, language, as an object, is neither a natural kind nor a social kind, but both natural and social kind, and is thus amenable to complementary approaches to research and explanation.

c. Attempts to unify the two approaches have been on the rise on both sides, in the formalist work of North American linguists (see e.g. Newmeyer 2001, for a review), and in the functionalist work of European linguists (e.g. Dik 1989 and Sperber and Wilson 1995). This trend is remarkably noticeable in the core formalist work, where construction-specific rules have been replaced by general principles. As it turns out, many of these principles use functional concepts, such as the universal ‘Economy Principle’\(^2\) and the principal of ‘Greed’\(^3\) in Minimalist Program (Chomsky 1995), determining the constituent structure of sentences and their movement. Within formal phonology, the rise of Optimality Theory (e.g. Prince and Smolensky 1993) represents a rather extreme step in the direction of the use of universal functional principles alone.

1.3 Types of Explanation in Linguistics

1.3.1 Apart from the overall approaches to the study of language, the terms ‘formal’ and ‘functional’ are also used for distinguishing two types of explanation in linguistics.
‘Formal’ explanations are unique to linguistics as they do not figure in a typology of explanations for natural (and human) sciences (see e.g. Nagel 1960). For an early use of the term ‘formal explanations’, consider the following quote from Chomsky (1971:44):

‘Where properties of language can be explained on such “functional” grounds, they provide no revealing insight into the nature of mind. Precisely because the explanations proposed here are “formal explanations”, precisely because the proposed principles are not essential or even natural properties of any imaginable language, they provide a revealing mirror of the mind (if correct).’

As is well-known, functional explanations are a valid type in natural sciences, especially life sciences. As observed in the above section, they are also being used increasingly in formal approaches to linguistics.

How are formal explanations distinguished from functional explanations? The distinction that Jayaseelan (1993:2) makes is useful to start with: ‘For a complete understanding of a thing, we need to see the thing in itself (which is the domain of formal explanations) and the thing in relation to other things (which is the domain of functional explanations).’

However, Jayaseelan’s elaboration of the distinction (1993:2) between formal and functional explanations in terms of “How Questions” (formal) and “Why Questions” (functional) is problematical, in the light of the well-known rejection by Chomsky (1986:10) of considering ‘knowledge-how’ as being ‘knowledge of language’, the ‘how’ meaning to do or behave. As is well-known, Chomsky has argued vehemently against the study of the behavioral aspect of language as a part of the study of Universal Grammar.

The main reason for the reliance of functional approaches on functional explanations is the critical relation of language to the external reality for communication. As should be obvious, this is also the reason for their general disparagement in formal approaches. More crucially, functional explanations, especially in functional approaches can be shown to be often not rigorous, and circular (see e.g. Newmayer 2001), and thus irrefutable, in the sense of Popper. There is thus hardly any example available of a revision of functional explanations in functional linguistics, unlike the case of formal explanations. They seem to gain acceptability when supported by rigorous formal analyses using common theoretical vocabulary, as, for example, in the case of the Minimalist principles cited above.
1.3.2 Basic Differences

Considering the rise in the trend for neutralizing the distinction between them, it appears that the employment of one is not treated as untenable alongside the other. There does, however, appear to be some sort of a demarcation, not strictly a fence, between the two types. What is the line of demarcation between formal and functional explanations, if any?

a. The basic difference between the two types of explanation can be attributed to the requirement for formal explanations to be stated, using a common conceptual vocabulary, so that they are falsifiable, but non-falsified. The distinction between the italicized expressions is crucial. A formal statement may include any type of concepts, formal or functional, but the statement should be refutable. If refuted, however, it must presumably be revised. Thus the Principle of Economy, referred to above, is a falsifiable and non-falsified principle. The notion ‘economy’ itself may be treated as formal or functional, depending on how it is defined. If it has the sense of being motivated by use, it may be treated as functional, but if it has the sense of simplicity that is verifiable, then it may be treated as formal. On the other hand, a principle such as Nocoda in OT, which disallows a coda (a formal term) in syllable structures, is considered a functional constraint because it is falsifiable, and is attested to be falsified. Markedness principles thus may be considered to be inherently a functional type. Falsifiability, then, is a strict criterion for a formal explanation, but not for a functional explanation.

There is abundant evidence in the literature to show the validity of this distinction between the two approaches to linguistic study, and can be shown to reveal some other distinctive aspects of the approaches. Thus, a formalist explanation typically places a form along with related possible and impossible forms and attempts to provide a unified account of them in terms of existing universal principles. By contrast, functionalists tend to focus on the relevant form and generalize about them in terms of the chosen descriptive framework, whose validity is never examined for universal significance. There is little attempt in the latter to explain what makes the relevant forms possible.

Consider, as illustrations of this point of difference, the formalist treatments of the Noun
Phrase in Abney (1987) and of Incorporation in Baker (1988) with the functionalist accounts of the phenomena in Givon (1993) and in Dik (1980), respectively. Abbney’s treatment takes a comprehensive stock of all the possible and impossible types of NPs and how they are predicted by the revised X-bar theory of phrase structures. Baker’s study similarly attempts to explain the phenomenon of noun incorporation in terms of the existing universal principles of Movement, Government and Case Theory, and thereby provides a justification of the relevant principles. In contrast, Givon’s account of NPs and Dik’s of Incorporation provide descriptions of the structures involved in terms of their own versions of Functional Grammar.

The depth and range of formal accounts is offset by their avowed exclusion of communicative use. Aspects of structure which are dependent on use find a more revealing treatment in functional accounts. Thus an examination of the aspects of reference and definiteness of NPs, which is bound in discourse (such as the knowledge of the world, speaker’s intuition, etc.) is found to unfurl a deeper range of facts about the knowledge of NPs in Givon (1993) that a reader of Abbney must remain ignorant of.

b. A second requirement of a formal explanation, based on the work on formal properties of cognition (e.g. Pylyshyn 1984, Bromberger and Halle 1989), and related to the one proposed above, is that it concern the computational faculty of language rather than the adaptive faculty, and, by implication, to invariant rather than emergent phenomena. The latter, being bound to contexts, may be unpredictable and thus stand falsified. According to this criterion explanations of sociolinguistic variations cannot be treated as formal, since they address the adaptive rather than the computational linguistic faculty. On the same grounds, the explanation of gradient phonetic phenomena referring to adaptive or emergent speech using continuous mathematics (e.g. Pierrehumbert et al. 2000) may not be considered formal, unless shown to be related to the computational faculty of speech.

It is worth repeating that we are assuming that for an explanation to be formal or functional does not affect its validity for scientific research, contrary to the assumptions of hard core formalists (e.g. Hale and Reiss 2000, van der Hulst 2000). Whether or not
one of them is more valuable than the other in linguistic theorizing is controversial at this stage.

1.4 Phonological Knowledge

Burton-Roberts, Carr and Docherty (2000) have recently shown that there are at least four main views concerning the goals and methods of phonological theory, and thus of phonological knowledge. Given the generally held assumption that phonological knowledge is intrinsic to linguistic knowledge, there are at least four different ways in which the various issues pertaining to the nature of investigation of phonological knowledge are envisaged and elaborated upon. We can call these views as (a) Isolative, (b) Functional, (c) Formal, and (d) Integrational or Unified. The Isolative view, held by Bromberger and Halle (1989, 2000), and endorsed by Chomsky (1995: 163, 212) considers phonology to be different from the rest of the language faculty, in belonging to the Periphery rather than the Core. The Functional View, in contrast to the Isolative view, held by Burton-Roberts, and supported by the work of Ohala (e.g. 1983, 1990), Lindblom (1986, 1992), Bybee (1994, 2001), and Hayes (1999), among others, considers phonology to be grounded in the sensorimotor apparatus. The Formal view of phonology, represented by Hale and Reiss, and supported by the research on sign language phonology (see for example, Coulter 1993, Brentari 1998), considers phonology to be like the rest of language faculty, and to be grounded in the mind. While the extreme view represented by Hale and Reiss excludes phonetic substance entirely from phonology (see also Myers 2000), other versions of this approach debate concerning the inclusion of phonetic substance in phonology. Finally, the Unified view, held by Pierrehumbert, Beckman, and Ladd, considers phonology to be a laboratory science, and argues for the adoption of scientific methodology to test formal claims. This approach seeks a bind between phonology, psycholinguistics and sociolinguistics.

I assume here the superiority of the Unified view of phonology to the other views. This view is in consonance with the growing trend for a unification of formal and functionalist concerns in linguistics, as noted above. We thus assume that phonological knowledge has the two integrated aspects, namely formal and functional. By formal aspects of
phonological knowledge we mean the innate, (parametrically) invariant computational
disposition for human speech, and by functional aspects of phonological knowledge we
mean the adaptive and emergent aspects of phonological knowledge, dependent on our
organic (perceptual and articulatory) systems.

2 Phonological Knowledge

2.1 An account of a speaker/hearer’s phonological knowledge should include the
following aspects of speech sounds: a) segment inventory (in common terms, inventory
of vowels and consonants), b) phonotactic constraints (i.e. constraints on the occurrence
of speech segments, for example, words end in vowels in Japanese and in some Dravidian
languages), c) alternation (i.e. alternation between segments in related forms, for
example, f and v in wife and wives, d) prosodic organization (in common terms, the
organization of speech forms from lower to higher levels, segments → syllables →
words → phonological phrases → intonational phrases), and (e) relation of phonology
with syntactic, semantic and pragmatic structure. I propose to take up one of these for
discussion here, namely, segments. My main purpose is to argue for evidence in favour of
the integration of the formal and the functional in linguistic knowledge.

2.2 Segments

2.2.1 If there is one topic that has occupied the interest of linguists since the earliest
times, it is the phonological segment. Although some crucial issues remain unresolved,
we know a lot more about it to-day than before. I take up two issues concerning it, which
can be expressed in the form of the following questions:

- How are phonological segments internally structured?
- What factors determine phonological inventories?

The questions are interrelated, as will be obvious in the course of the discussion below.

2.2.2 Internal structure of phonological segments: distinctive features

It is now considered an axiomatic fact about speech sound segments that they are made
up of distinctive features, which commonly have binary (+/-) values, as shown for a five
vowel system below:
The vowels /i/ and /u/ share the [+] value of the feature ‘high’, and are contrasted from
the other vowels which are assigned the [–] value of the feature. The feature [+low]
contrasts the vowel [a] from all the other vowels. The high vowels /i/ and /u/ are
distinguished from each other by the feature ‘back’: /i/ is [–back], while /u/ is [+back].

This way a set of distinctive features groups sounds that share common phonetic
properties as well as distinguishes each sound from another within a phonological
system.

The theory of distinctive features has had a continuous history from the time of
Trubetzkoy (1939) and Jakobson (1968), leading figures of the Prague Linguistic
Circle, in the period preceding the Second World War. It has undergone various
revisions (see e.g. Jakobson, Fant & Halle 1952, Chomsky & Halle 1968, Clements 1985,
McCarthy 1988). The crucial point to note here is that the development of the theory
presents itself as an exemplar case of the integration of form and function in linguistic
knowledge. Let us look at some important points of the theory.

The concept ‘distinctive feature’ owes its origin to the functional orientation of Prague
School. Trubetzkoy (1939), and perhaps Jakobson (1968 [1932]) before him, proposed
it as a phonetic property constituting a unit of distinctive contrast. The phonetic property
combines with other distinctive phonetic properties to yield “the inner order or structure
of the phonemic inventory as a system of distinctive oppositions” (p. 71). For example,
the feature ‘voice’ is in ‘privative’ contrast in /b/: /p/; the feature is present in one and
absent in the other segment. The opposition between the Hindi front vowels /i/ and /e/
from amongst the set /i e ɛ/ is ‘gradual’, since /e/ is one degree lower in height than /i/, while /ɛ/ is two degrees lower, as is obvious from the diagram below:

(3)

<table>
<thead>
<tr>
<th></th>
<th>Front</th>
<th>Central</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>Close</td>
<td>i</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close-mid</td>
<td>e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open-mid</td>
<td>ɛ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

And the opposition between /s/ and /ʃ/ or between /l/ and /ɹ/ in English is neither privative nor gradual but ‘equipollent’.

ii. The development of the theory of distinctive features has taken place in the climate of formal linguistics, in the main, Generative Phonology (Chomsky & Halle 1968) and its sub-theories. The significance of the notion of distinctive feature in the phonological analysis of languages was found to extend itself to other aspects of phonological theory, such as a universal theory of phonological oppositions (e.g Jakobson & Halle 1956, Chomsky & Halle 1968), a universal theory of possible speech sounds (e.g. Chomsky & Halle 1968), and the characterization of natural classes of speech sounds (e.g. Chomsky & Halle 1968, Clements 1985). Some of the developments in the theory of distinctive features owing to the formal orientation can be summed up as below.

a. Phonological generalizations are best expressed in the form of a finite set of distinctive features. The term ‘best’ can be characterized in terms of economy of statements, natural classes of segments, and universality of the set that can be assumed to have different combinations to yield all the possible segments in world languages. Maddiesson (1984) lists some more than 760 segments from a database of 317 languages (out of a total stock of about 6000 languages). All of these, and more, can be claimed to derive from a finite set of features. For classic arguments in favour of the formal view presented above, see Halle (1962). Much research in the distinctive feature theory can be found to have been devoted to the question of the exact number and definitions of such a finite set.
b. The current view of the organization of the distinctive features can be represented in the following diagram (see Broe 1992):

(4)

(4) represents a non-linear organization of distinctive features. The features are envisaged as being on different nodes or ‘tiers’. Thus, there is the ‘voice’ tier and the ‘nasal’ tier, and so on. The capital letter nodes represent the class nodes, while the small letter nodes represent individual distinctive features.

The non-linear organization of the features is found to account for all the segment types in world languages and all possible processes of change due to assimilation (i.e. adjacent occurrence), which may involve a change at any of the tiers, the root node (for total assimilation, as in Hindi /kitna/ → /kitta/ ‘how much’, /n/ totally assimilates to the preceding /t/), or the class node (for partial assimilation, as in Sanskrit /sæm/ + /toʊ/ → /səntʊ/ ‘satisfaction’, where /m/ assimilates partially for place to the following /t/), or the feature node (for single feature assimilation, as in Sanskrit /ʊək/ + /i:/ → /uəgiː/ ‘god of speech’, where the voiceless /k/ assimilates to the following vowel in the feature ‘voice’ and is pronounced as /ɡ/).

The hierarchical organization of the features in (4) also implies their autonomous representation, along the lines of the autonomous representation of other phonological
units, as claimed by Autosegmental Phonology (see Goldsmith 1976, McCarthy 1979). (4) thus would not have been envisaged without the advent of the theory of Autosegmental Phonology. Both the theories support each other as theories of phonological representation.

c. For a fulsome account of how segments are represented in the lexicon and how they are articulated and perceived on the surface, some sub-theories of generative phonology assume that all the features are not present from the start. In other words, the features may be underspecified (see e.g. Archangeli 1984, 1988, Steriade 1987, 1995, Clements 1988). This assumption is motivated primarily by the consideration of economy in linguistic descriptions: simpler descriptions are preferred as being more highly valued. Apart from this general requirement, the notion is argued to be crucial to the explanation of certain phenomena such as vowel harmony, nasal spread, etc., and is also found to be useful in the explanation of exceptional behaviour of segments. Consider, for example, the fact that in Punjabi and Kashmiri and several other languages which do not permit more than one consonant in the beginning of words, there is a vowel inserted between the consonants, as in /prədi:p/ → /pərdi:p/ (‘a name’), /spəstʃ/ → /səpəstʃ/ ‘clear’. In all such cases, the vowel is /ə:/ Why should this be so? The answer that comes from a theory of Underspecification (e.g. Archangeli 1988) is that since in Hindi and Punjabi all the vowels are specified for some features, except for the vowel /ə/, which is lexically totally underspecified, it is inserted as a default case when a vowel per se is required to be inserted. As in the best of the scientific tradition, the development of the theory of Underspecification is divided with regards to the nature of underspecification. There are two main positions here, namely, Radical Underspecification (RUT) (Archangeli 1988) and Contrastive Underspecification (CUT) (e.g. Steriade 1987, Clements 1988). The basic difference between them is that RUT stands for maximum economy of feature specification in the lexical representation of words, subject to the preservation of lexical differentiation, whereas CUT allows for less restriction on economy in favour of lexical contrast of segments in given contexts. More concretely, given the full specification of
features in a five vowel system in (2), the RUT and the CUT feature specifications of the
five vowels for lexical representation will be as in (5) and (6) respectively (see also Roca
and Johnson 1999, Hall 2003):

\[
\begin{array}{|c|c|c|c|c|}
\hline
& i & e & a & o & u \\
\hline
\text{high} & - & - & - & - & - \\
\hline
\text{low} & + & + & + & + & + \\
\hline
\text{back} & + & + & + & + & + \\
\hline
\text{round} & - & - & - & - & - \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|c|c|}
\hline
& i & u & e & o & a \\
\hline
\text{high} & + & + & - & - & - \\
\hline
\text{low} & - & - & + & + & + \\
\hline
\text{back} & - & - & + & + & + \\
\hline
\text{round} & - & - & - & - & - \\
\hline
\end{array}
\]

(5) differs from (6) on two counts: one, it has only one value specified for each feature,
either + or -, and, two, one of the segments in it is fully underspecified, namely, /i/. (6)
contains opposite values of features for different segments in the lexical representation,
and lacks a fully underspecified segment.

For both (5) and (6) a set of redundancy rules are required to fill out the blank features.
RUT requires two sets- one, (7), for the vowels other than the fully underspecified vowel
/i/, and the other, (8), for /i/.

\[
\begin{align*}
\text{[ ]} & \rightarrow [-\text{high}] / \ [\ldots] \\
& \quad +\text{low} \\
\text{[ ]} & \rightarrow [+\text{back}] / \ [\ldots] \\
& \quad +\text{low} \\
\text{[ ]} & \rightarrow [-\text{round}] / \ [\ldots] \\
& \quad +\text{low} \\
\text{[ ]} & \rightarrow [\alpha \text{ round}] / \ [\ldots] \\
& \quad +\text{back} \\
& \quad -\text{low}
\end{align*}
\]

13
(8) \[\, \] \to [+high]  \\
\[\, \] \to [-low]  \\
\[\, \] \to [-back]  \\
\[\, \] \to [-round]

CUT requires a single set of redundancy rules for specifying the missing feature values in (6):

(9) \ [+low] \to [-high]  \\
\ [+high] \to [-low]  \\
\ [+low] \to [+back]  \\
\ [-back] \to [-low]  \\
\ [\alpha \ back] \to [\alpha \ round]

When are the underspecified features filled out? The theories assume that the features are specified for the segments when required by phonological processes. Do the features remain underspecified when not required by phonological processes? There is more than one answer to that question. The theory of Lexical Phonology (e.g Kiparsky 1982, Mohanan 1986) assumes that segments must be fully specified for the forms in the postlexical module, which follows word formation and syntax. There is contrary evidence, however, for the persistence of underspecification in the postlexical module (e.g., Keating 1988, Lahiri and Reetz 2001).

Not all phonological theories incorporate the notion of underspecification. Optimality Theory (OT) (Prince and Smolensky 1993, McCarthy and Prince 1993), a dominant theory of phonology since the last decade, has no scope for underspecification, as it disallows any constraint on the underlying representation of forms (see Hall 2001 for a discussion). The latter are assumed to be freely generated on account of the Richness of the Base Hypothesis, according to which there are no restrictions on the generation of the lexical form. Constraints are applicable on the surface representation of forms in OT. All changes that differentiate the surface forms from their underlying representations take place in order to respect the constraints, which are in principle violable and ranked with respect to each other.
On the whole, the debate regarding the formal account of underspecification is far from settled. What the formal accounts try to achieve through rigorous argumentation and evidence is the precise nature of the computational system governing our phonological knowledge.

Certain questions concerning phonological inventories have been found to be of greater interest in the functional approach than in the formal approach. One such question concerns possible principle(s) of organization of segment inventories.

2.2.3 Phonological Inventories

A universal fact about phonologies is that every language has a unique inventory of speech sounds (see e.g. Maddieson 1984). The inventories must be posited at least at two levels, namely, underlying (or lexical) and surface phonetic. The former are assumed to represent words in the lexicon, while the latter represent the sounds on the surface, better known as allophones. Lexical Phonology (see e.g. Kiparsky 1982, Mohanan 1986) assumes a third level of representation, namely, lexical phonological level, which is intermediate between underlying level and surface phonetic level, and is roughly equivalent to the classical phonemic level. Thus the words leaf and leaves have the underlying representations /li:f/ and /li:f+z/. This level represents the knowledge about the underlying oneness of forms that may have different realizations on the surface; the common form /li:f/ is found for both the words. The lexical phonological representations of the words are /li:f/ and /li:vz/. This level represents the forms in terms of the contrasting sound units, for example the sounds /f/ and /v/ which occur on the surface and which contrast lexical items such as fan and van. Finally, the surface phonetic representations are [liːf] and [liːvz] in American English (leaving out other details of pronunciation). This level represents the non-contrastive and predictable sounds found in the pronunciations of forms, for example, the long vowel is pronounced shorter in leaf than in leaves since it precedes a voiceless consonant, namely, /f/.

Consider the restricted phonemic inventories (roughly equivalent to the lexical phonological level as discussed above) for plosives in languages selected at random from
the four main language families in India, namely, Austro-Asiatic (Bhumij), Dravidian (Toda), Indo-Aryan (Kashmiri), and Tibeto-Burman (Angami):

(10) **Angami** (Giridhar 1980)

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosives</strong></td>
<td>p, b</td>
<td>pʰ</td>
<td>t, d</td>
<td></td>
<td></td>
<td>k, g</td>
<td></td>
</tr>
</tbody>
</table>

**Bhumij** (Ramaswamy 1992)

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosives</strong></td>
<td>p, b, bⁿ</td>
<td>t, dⁿ</td>
<td>tʰ, dʰⁿ</td>
<td></td>
<td>k, gⁿ</td>
<td>kʰⁿ</td>
<td>gⁿ</td>
</tr>
</tbody>
</table>

**Kashmiri** (Kelkar & Trishal 1964, Bhat 1987):

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosives</strong></td>
<td>pʰ, bʰ</td>
<td>t, dʰ</td>
<td>tʰ, dʰʰ</td>
<td></td>
<td>k, gʰ</td>
<td>kʰ</td>
<td>gʰ</td>
</tr>
</tbody>
</table>

**Toda** (Emeneau 1984)

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Dental</th>
<th>Alveolar</th>
<th>Retroflex</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Plosives</strong></td>
<td>p, b</td>
<td>t, d</td>
<td>t, d</td>
<td></td>
<td>c, j</td>
<td>k, g</td>
<td></td>
</tr>
</tbody>
</table>
The blank slots in the charts show that the relevant consonants do not occur in the language. The consonants on the left in a slot are voiceless and on the right are voiced. Note that none of the inventories is identical with another. A crucial question regarding the inventories is: How are they brought about? What are the factors that give rise to them? At least two contending explanations have been proposed in the literature, namely, the principle of Maximal Dispersion or Sufficient Contrast (e.g. Lindblom 1986, 1992, Lindblom, MacNeilage and Studdart-Kennedy 1983), and the principle of Feature Economy (Clements 2001, 2003, Weijer and Hinskens 2003). Both employ functional concepts and both have formalized them.

According to Lindblom and his colleagues a combination of constraints on production and perception of speech sounds, called Sufficient Contrast, helps shape inventories of speech sounds in languages. We find the principle at work for the phonological systems in general, especially vowels. The vowel systems that are commonly found among world languages are optimally dispersed (Lilyencrants and Lindblom 1972). A typical eight vowel pattern is as in (11), attested to be found commonly among world languages (Maddieson 1984):

\[(11)\]

\[
\begin{array}{c}
 i \\
 e \\
 \varepsilon \\
 a \\
 \hline \\
 u \\
 o \\
 \circ \\
 \end{array}
\]

The inventory in (11) shows that the vowels are distributed symmetrically in terms of height and front-back parameters. Alternative symmetrical patterns are possible, such as the following:
There exists a fairly large body of literature in phonetics by now which accounts for such symmetrical patterns (see Boersma 1998, and De Boer 2001 for detailed discussions). While the theory of maximal dispersion accounts for vowel systems in general, it must leave as exceptions cases of asymmetrical patterns, such as the following found in Indian languages: ɪ e œ u o ɔ a (7 vowels in Gallong, see Dasgupta 1963), ɪ e u ø o ɔ ø (7 vowels in Nicobarese, see Radhakrishnan 1981). When it comes to consonants the theory of Maximal dispersion meets with less success. There are many cases of consonant patterns that seem to clutter up a given articulatory-perceptual area.

The crucial question at hand is: Do we have a theory to account for the organization of both symmetrical and asymmetrical patterns? Such a theory will be superior to the one that makes predictions only about symmetrical patterns. Clements in a number of recent studies (e.g. 2001, 2003) proposes such a theory, drawing from earlier work (e.g. Martinet 1955), and in consonance with a lot of contemporary work on representational economy. Clements argues for a general principle of representational economy “according to which representational elements are specified in a given language only to the extent that they are needed in order to express generalizations about the phonological system” (Clements 2001).

The theory of representational economy differs from earlier versions of feature theory, namely, full specification (Chomsky and Halle 1968), and underspecification theories. It assumes that only active features are specified. The conditions for feature activation vary
at different levels: lexical, phonological and phonetic. In short, features should be lexically distinctive, phonologically involved in alternations, and phonetically pronounceable. Given their frequency of activation, features can be ranked on a Feature Accessibility Scale (e.g. Calabrese 1994). Thus the feature [sonorant] is found to be the highest on the scale, because it is found to contrast segments at the lexical level, whereas [constricted glottis], which characterizes glottal stops, is a feature that is lower on the scale, since fewer languages need to activate it.

Given the theories of Feature Activation and Representational Economy, Clements proposes the principle of Feature Economy (Clements 2003) that says, “if a feature is used once in a system, it will tend to be used again.” “Feature economy applies to features that are commonly distinctive in the class of sounds under comparison, but not to features that tend to be redundant in that class” (Clements 2003). Feature Economy predicts the presence of regular patterns in phonological inventories, such as the ones for plosives in the languages presented in (10). The features used for the bilabial plosives /p pʰ b/ in Angami are [-sonorant, -continuant, +spread glottis (aspiration)], among others. Feature Economy predicts that these features may be used again. Thus the plosive series emerges. It is usually found that segments are patterned in phonological inventories. The principle has also been found to further make predictions about segments that result from modification types such as labialization, palatalization, breathy voice, etc. (van de Weijer 2003, van de Weijer and Hinskens 2004). Termed as Feature Economy of Segmental Modification (I), the principle says, “If a feature occurs as segmental modification, then it will also occur as a primary consonantal feature”. Thus languages having aspirated plosives, /pʰ tʰ / etc. also have the segment /h/. The feature common for aspiration and /h/ is [+ spread glottis]. Since [+spread glottis] has been used once in the system, as a primary feature for /h/, it gets used again in modified segment type, namely, aspirated plosives. I expect that some version of feature economy can also explain allophonic inventories in languages.

3. Discussion and pedagogic implications
The discussion of the issues relating to phonological segments in the above section underscores the view of phonological knowledge that integrates form and function. We
have seen that the notion ‘distinctive features’ is of functionalist origin, and its theoretical
development was guided by formal concerns. Questions of the minimal number of the
distinctive features, their organization, and the nature of their specification have been
taken up for rigorous examination, in relation to other formal aspects of phonological
knowledge. Thus the geometrical representation of the features was seen as possible in
the light of the independent theory of Autosegmental Phonology. In regard to the
question of universal principles for the patterning of phonological segments, we found
that the explanations offered are functionalist in orientation, showing their motivation in
language use. Both the theories of Maximal Dispersion and Feature Economy suggest
that the computational system (in this case the set of distinctive features) is constrained
by considerations of Use, such as discriminability and pronounceability (for Maximal
Dispersion) and parsimony (for Feature Economy).

It should be quite clear that as a theory of phonology that seeks to account for the internal
structure of phonological segments and their inventories in different languages cannot
strictly separate formal from functional considerations, the knowledge of phonology (and
of language) must also be seen as an integrated whole. The investigation of linguistic
knowledge must thus be pursued by a unified research programme rather than one that is
only formal or functional in nature.

A little thought should make it clear as to why this should be so. The computational
system has to work within certain constraints. In the case of spoken language, the
constraints have to do with the auditory structure of the brain, the articulatory means
available, the dimension of time, memory space for information processing at one time,
etc. The computational system obviously has to work within the given constraints. Its
nature and its operation are integrated.

The general implication for pedagogy that follows from this discussion is that both
computational and adaptive considerations must go into a child’s education. Pedagogic
methods cannot be different from the nature of knowledge. The linguistic curriculum
must have enough scope for the use of the specific linguistic structures aimed to be
acquired. If use and function are the ‘ground’ then form is the ‘structure’ that stands on it.
With respect to human knowledge, we should add further, the ground not only provides
the basis for the structure to stand on, but also contributes to determining its form. Form
and function are integrated. Unless the integrated nature of knowledge is perceived this way, curriculum planning is likely to remain distant from the reality of knowledge. At least that’s the lesson we want to draw from a consideration of the nature of phonological knowledge.

**FOOTNOTES**

1. To some extent, this is the strain in Darnell et al. (eds.) (1998) and Newmeyer (1998), cited in the above footnote. Kelkar (1997) is a noteworthy attempt at unifying the two approaches in an analysis of Marathi.

2. Economy Principle: A principle of grammar (discussed at length in Chomsky 1995) which requires that syntactic representations contain as few constituents and syntactic derivations involve as few grammatical operations as possible.

3. Greed: A principle of grammar (proposed as one of the “self-serving” principles in Chomsky 1995) which specifies that constituents undergo movement in order to satisfy their own morphological requirements.

4. Throughout this paper slanted lines are used to enclose phonemic transcription, square brackets [ ] to enclose phonetic transcription. The latter is closer to actual pronunciation. Thus the word “talk” will be transcribed phonemically as /tɔːk/ and phonetically as [tʰɔːk]. The first consonant is transcribed with closer detail in the square brackets. The vowel is the same in both, namely , ɔː for the orthographic “a”. And the written symbol “l” has nothing equivalent in either since it is not pronounced.
REFERENCES


COLOPHON:

I am grateful to the participants at the seminar for their comments on the paper. I would especially record a rather profitable experience of sharing some of the views expressed here with my students for a course on Issues in Linguistic Theory given in the Winter Semester 2004, especially the participation of Anish, Joyashri, Narayan and Rajakrishna.