Phonology and Language Use

JOAN BYBEE

University of New Mexico

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York NY 10011–4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

© Joan Bybee 2001

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2001
First paperback edition 2003

Typeface Times Roman 10.25/13 pt. *System* QuarkXPress [BTS]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication Data
Bybee, Joan L.
Phonology and language use / Joan Bybee.
p. cm. – (Cambridge studies in linguistics ; 94)
Includes bibliographical references and index.
ISBN 0 521 58374 8 hardback
1. Grammar, Comparative and general – Phonology. 2. Linguistic change.
3. Grammar, Comparative and general – Morphology. 4. Universals (Linguistics)
I. Title. II. Series.
P217.3 .B93 2001
414 – dc21
00-045525

ISBN 0 521 58374 8 hardback
ISBN 0 521 53378 3 paperback

Transferred to digital printing 2003

Contents

List of Figures *page* xiii
List of Tables xv
Acknowledgments xvii

1 Language Use as Part of Linguistic Theory 1
1.1 Substance and Usage in Phonology 1
1.2 Some Basic Principles of a Usage-Based Model 6
1.3 The Creative Role of Repetition 8
1.4 Frequency Effects 10
1.5 Phonology as Procedure, Structure as Emergent 14
1.6 Organization of the Book 16
1.7 Language as a Part of Human Behavior 17

2 A Usage-Based Model for Phonology and Morphology 19
2.1 Introduction 19
2.2 The Rule/List Fallacy 20
2.3 Organized Storage 21
2.4 Morphological Structure Is Emergent 23
2.5 Rules and Schemas Compared 26
2.6 Frequency Effects 28
2.7 Units of Storage 29
2.8 Phonological Units 31
2.9 From Local to General Schemas 31
2.10 Conclusion 33

3 The Nature of Lexical Representation 35
3.1 Introduction 35
3.2 The Phonemic Principle 35

3.3 A Cognitively Realistic Model of Phonological Representation 37
3.4 Linguistic Evidence for Detailed and Redundant Storage 40
3.5 Usage-Based Categorization versus Phonemic Representation 49
3.6 Phonetic Detail in the Lexicon – Variation and the Early Involvement of the Lexicon and Morphology in Change 54
3.7 A Model for Sound Change 57
3.8 Special Reduction of High-Frequency Words and Phrases 60
3.9 Conclusion 62

4 Phonological Processes, Phonological Patterns 63
4.1 Introduction 63
4.2 Phonetic Etiology and Its Limits 65
4.3 Articulatory Gestures 69
4.4 Patterns of Change and Constraints on Processes 77
4.5 Segments as Emergent Units 85
4.6 Generalization over Syllable-Initial and Syllable-Final Position 86
4.7 Phonotactics 88
4.8 Conclusion 95

5 The Interaction of Phonology with Morphology 96
5.1 Introduction 96
5.2 Morphological versus Phonological Conditioning 97
5.3 Lexical Storage of Complex Forms, Both Regular and Irregular 109
5.4 Lexical Strength 113
5.5 Paradigmatic Relations Expressed as Lexical Connections 117
5.6 Lexical Classes: Productivity Due to Type Frequency 118
5.7 The Interaction of Lexical Strength and Lexical Connection 124
5.8 Product-Oriented Schemas 126
5.9 Phonological Similarity in Gangs 130
5.10 Conclusion 135

6 The Units of Storage and Access: Morphemes, Words, and Phrases 137
- 6.1 Introduction 137
- 6.2 Phonological Representations of Words 138
- 6.3 Morphemes within Words 144
- 6.4 Phrases and Constructions with Alternations 157
- 6.5 Conclusion 166

7 Constructions as Processing Units: The Rise and Fall of French Liaison 167
- 7.1 Introduction 167
- 7.2 Final Consonant Deletion in French 168
- 7.3 Grammatical Constructions and Liaison 171
- 7.4 Loss of Liaison as Regularization 177
- 7.5 Syntactic Cohesion as Frequency of Co-occurrence 185
- 7.6 Taking the Phonology Seriously 185
- 7.7 Conclusion 187

8 Universals, Synchrony and Diachrony 189
- 8.1 Universals and Explanation 189
- 8.2 Searching for Universals 191
- 8.3 Phoneme Inventories 197
- 8.4 Two Main Mechanisms for Phonological Change 199
- 8.5 Syllable Structure 204
- 8.6 More Evidence against Universals as Purely Synchronic 211
- 8.7 Diachronic Sources for Formal Universals: The Phonemic Principle and Structure Preservation 212

References 217

Author Index 231

Subject Index 235

Languages Index 238

1

Language Use as Part of Linguistic Theory

1.1 Substance and Usage in Phonology

This book introduces into the traditional study of phonology the notion that language use plays a role in shaping the form and content of sound systems. In particular, the frequency with which individual words or sequences of words are used and the frequency with which certain patterns recur in a language affects the nature of mental representation and in some cases the actual phonetic shape of words. It is the goal of the present work to explore to the extent possible at the present moment the nature of the relation between the use of linguistic forms on the one hand, and their storage and processing on the other.

To someone approaching linguistics from other disciplines, it might seem odd that language use has not been taken into account in formulating theories of language. However, since language is such a complex phenomenon, it has been necessary to narrow the field of study to make it manageable. Thus we commonly separate phonology from syntax, synchrony from diachrony, child language from adult language, and so on, constantly bearing in mind that interactions exist that will eventually have to be taken into account. We then go on to formulate theories for these domains – a theory of syntax, a theory of phonology, a theory of language acquisition – knowing all the while that the ultimate goal is to encompass all these subfields in one theory of language.

Early in the twentieth century, a proposal was made to distinguish the shared knowledge that a community of speakers has from the actual uses to which that knowledge is put (de Saussure 1916). Many researchers then focused their attention on the structure of that shared knowledge (called 'langue' by Saussure and 'competence' by Chomsky

1965) and paid little attention to language use in real time. The focus on competence, or the structure of language, turned out to be extremely productive. Structuralism provided linguists with a workshop of analytic tools for breaking down the continuous speech stream into units, and these units into features; structuralism postulated hierarchical relations among the units and assigned structures to different levels of grammar, organizing language and the people who study it into subfields – phonology, morphology, syntax, and semantics.

The present work proposes to demonstrate that the focus on structure needs to be supplemented with a perspective that includes more than just structure, a view that includes two other important aspects of the language phenomenon – the material content or substance of language, and language use. The SUBSTANCE of language refers to the two polar ends – phonetics and semantics – that language molds and structures, the two ends between which language forms the bridge. Language USE includes not just the processing of language, but all the social and interactional uses to which language is put. For present purposes, in the context of phonology, the frequency with which certain words, phrases, or patterns are used will be shown to have an impact on phonological structure. I will return to a discussion of these two aspects of language and the role they play in past and future theories after describing some recent developments in linguistics and related fields that suggest a need for an enlarged perspective on language.

In the domain of morphosyntax, a substantial development beyond structuralism has already taken place. The content of grammatical categories has been studied as a substantive rather than a structural matter, for example, in crosslinguistic studies of subject, topic, noun, verb, tense, aspect (Comrie 1976, 1985, Dahl 1985), mood, and so on. Also use is being studied as a prime shaper of syntactic structure (Givón 1979, Haiman 1994, Hopper and Thompson 1984, and others) and morphological structure (Bybee 1985, Bybee et al. 1994, DuBois 1985). So far, no comparable development has occurred in phonology, but there are several indicators that it is time to open up the field to new questions and new sources of data and explanation.

Despite having looked carefully at matters of structure, having defined and redefined units such as phoneme and morpheme (or formative), having shifted and reshifted levels such as phonemic and morphophonemic, we find that problems and questions still remain. Units and levels do not submit to definitions that work for every case. We still do not have strict definitions of even the most basic units, such as segment, syllable, morpheme, and word. Instead we find variation and gradience commonplace in empirical studies, and we find phonological phenomena intimately bound up with lexicon and morphology, syntax, discourse, and social context.

Developments from outside linguistics also point to a new view of language. Studies of natural categorization by psychologist Eleanor Rosch and her colleagues have had an impact on the way that linguists view categories, including word meaning (Lakoff 1987), grammatical classes such as gender (Zubin and Köpcke 1981), verb classes (Bybee and Moder 1983), grammatical functions such as subject and topic, and phonetic categories (K. Johnson 1997, Miller 1994, and other 'exemplar' approaches to phonetic categories). In particular, these studies show that the way human beings categorize both nonlinguistic and linguistic entities is not by discrete assignments to categories based on the presence or absence of features, but rather by comparison of features shared with a central member. All category members need not have all of the features characterizing the category, but a member is more central or more marginal depending on the number and nature of shared features. Moreover, Nosofsky (1988) has shown that the perceived center of a category can shift toward the more frequently experienced members.

A second development important to linguistic modeling is the development of computer models that can reproduce apparent 'rule-governed' behavior as well as probabilistic behavior using parallel distributed processing (Daugherty and Seidenberg 1994, Rumelhart and McClelland 1986, and others). In such models, labeled connectionist models, structures are not given in advance (i.e., innate), but take their form from the nature of the input, just as neurological matter is structured by the input it receives. Connectionist models, then, are quite compatible with usage-based theories of language. Langacker (1987) and now Ohala and Ohala (1995) argue that storage of linguistic percepts should be like the storage of other mental percepts.

Yet a third recent development applicable to a large array of sciences is the study of complex systems and their emergent properties. The basic idea behind emergence as it will be applicable here is that certain simple properties of a substantive nature, when applied repeatedly, create structure. Lindblom et al. (1984) are, to my knowledge, the first to apply the notion of emergent structure in linguistics. They illustrate emergence in the following way:

> Termites construct nests that are structured in terms of pillars and arches and that create a sort of 'air-conditioned' environment. The form of these nests appears to arise as a result of a simple local behavioral pattern which is followed by each individual insect: the pillars and arches are formed by deposits of glutinous sand flavored with pheromone. Pheromone is a chemical substance that is used in communication within certain insect species. Animals respond to such stimuli after (tasting or) smelling them. Each termite appears to follow a path of increasing pheromone density and deposit when the density starts to decrease. Suppose the termites begin to build on a fairly flat surface. In the beginning the deposits are randomly distributed. A fairly uniform distribution of pheromone is produced. Somewhat later local peaks have begun to appear serving as stimuli for further deposits that gradually grow into pillars and walls by iteration of the same basic stimulus-response process. At points where several such peaks come close, stimulus conditions are particularly likely to generate responses. Deposits made near such maxima of stimulation tend to form arches. As termites continue their local behavior in this manner, the elaborate structure of the nest gradually emerges. (Lindblom et al. 1984: 185–186)

Lindblom et al. point out that the importance of this notion for linguistics is that structure can be explained without attributing a 'mental blueprint' to the creatures creating the structure – that substance and form are intimately related (see also Hopper 1987, Keller 1994). Note further that in this example and others of emergence in complex systems, substance and form are related via the PROCESS by which the structure is created.

If we apply emergence to language, the substance and use interact to create structure. The substance in question includes both phonetics and semantics. Phonetic substance has always been included in the field of phonology. Only a few phonologists have ever proposed that phonology is independent of phonetics (see Postal 1968). On the contrary, most phonologists see phonetics as motivating phonology (for a recent statement, see Hayes 1999). They have perhaps not always been serious enough about pursuing the phonetic facts, however. One prominent feature of generative phonology has been its disdain for the 'low-level' phonetic properties of speech, properties that presumably border on performance.

Semantics, on the other hand, has been considered irrelevant to phonology. This would not seem to be such a serious allegation to level at phonologists, except that phonological descriptions and theoretical works are full of references to notions such as morpheme and word boundaries – both of which delimit meaningful units – as well as to specific grammatical categories or specific morphemes. Generative phonologists and Optimality Theory phonologists have proceeded as though the content of these categories did not matter. I have shown in Bybee (1985) that the phonological fusion of morphemes reflects their degree of semantic fusion, and in the chapters of this book, I will explore further the relation between grammatical and lexical units and phonological structure. Generative theories have largely neglected such topics: even though morphological decomposition has played an important role in the development of generative theories from *The Sound Pattern of English* to Lexical Phonology and Optimality Theory, the semantic derivations that should parallel the phonological ones have never been attempted.

While substance has found its way into phonology from both the phonetic and semantic end, USE has been systematically excluded from structuralist theories altogether. As mentioned earlier, distinctions such as langue versus parole (de Saussure) and competence versus performance (Chomsky) were specifically designed to set up a mental object that is separate from the uses to which it is put and to designate the mental object as the proper domain for linguistics. Of course, there is some value in distinguishing mental representations from the social activities upon which they are based, but totally excluding factors of use from consideration ignores the potential relation between representation and use. It is certainly possible that the way language is used affects the way it is represented cognitively, and thus the way it is structured.

In fact, a good deal of progress in morphology and syntax has been made in explaining specific phenomena by making just this assumption. It has been shown that syntactic structures are the result of the conventionalization of frequently used discourse patterns (e.g., DuBois 1985, Givón 1979), and that grammatical morphemes develop from lexical morphemes in particular constructions through increases in the frequency of use and through extension in use to more and more contexts (Bybee et al. 1994, Haiman 1994). Greenberg (1966) has demonstrated that markedness effects are directly related to frequency of use, with unmarked members of categories being the most frequent, and Tiersma (1982) has shown that this hypothesis also explains cases of local markedness in morphology. Psycholinguists have long known that high-frequency words are accessed faster than low-frequency ones, and I have argued that high-frequency irregular morphological formations tend to maintain their irregularities precisely because of their high frequency (Bybee 1985, Hooper 1976b). In all of these findings we have a dynamic aspect – language structure is becoming or remaining because of the way language is used. Thus the emphasis on the static, synchronic language as the object of study has given way to the view of language as slowly, gradually, but inexorably mutating under the dynamic forces of language use.

Very little attention has been given to phonology in this usage-based approach to language, yet these same ideas CAN be applied to phonological phenomena with very interesting results. It is the purpose of this book to explore the phenomena that have traditionally been studied as phonology, reevaluating structural notions in terms of use and substance. The successes of structuralism in its various guises are not being discarded. Rather structural notions will first be empirically evaluated to ascertain their viability, then the basis of such notions will be considered, and the role that substance and especially, use, plays in the phenomena will be discussed. The phenomena discussed here point to a deep involvement of phonology with lexicon and grammar, and a role for both token and type frequency in shaping phonological structure. A dynamic view of language is taken here, one that integrates both synchronic and diachronic sources of explanation.[1]

1.2 Some Basic Principles of a Usage-Based Model

The ideas that I will apply to phonology are for the most part already present in the literature and are now shared by a number of linguists, phoneticians, and psychologists. A brief statement of these ideas follows.

1. Experience affects representation. The use of forms and patterns both in production and perception affects their representation in memory. High-frequency words and phrases have stronger representations in the sense that they are more easily accessed and less likely to undergo analogical change. Low-frequency words are more difficult to access and may even become so weak as to be forgotten. The lexical strength of words may change as they are used more or less in different contexts. Patterns (represented as schemas, see below) that apply to more items are also stronger

---

[1] The phonological theory developed here is quite different from Natural Generative Phonology (NGP) (Hooper 1976a). For while NGP had very concrete lexical representations, much involvement of morphology and the lexicon with phonology, and the same view of the relation of synchrony to diachrony, it was a structuralist theory and provided no means of representing the impact of language use on language structure.

and more accessible, and thus more productive than those applying to fewer items. This is in contrast to modular approaches in which representations and rules or constraints are all static and fixed, and in which all rules or representations in the same component have the same status (for instance, all being equally accessible no matter how many forms they apply to).

2. Mental representations of linguistic objects have the same properties as mental representations of other objects. Of course, this is the simplest assumption we can make – that the brain operates in the same way in different domains. One consequence of this assumption is that mental representations do not have predictable properties abstracted away from them, but rather are firmly based on categorizations of actual tokens. As Langacker (1987) and Ohala and Ohala (1995) have pointed out, if predictable properties are taken away from objects, they become unrecognizable. (See Chapter 2 for further discussion.)
3. Categorization is based on identity or similarity. Categorization organizes the storage of phonological percepts. What form this categorization takes is an interesting question and one that can be approached through phonetic and psychological experimentation as well as through analogies with findings in other perceptual domains. From structural linguistic analysis we can already identify many different types of relations among linguistic objects – for example, the relation between two phonetic tokens of the same word, that between tokens of the same morpheme in different words, and that between two similar phones in different words in the same or different contexts.
4. Generalizations over forms are not separate from the stored representation of forms but emerge directly from them. In Langacker's terms, there is no 'rule/list separation' (see Chapter 2). Generalizations over forms are expressed as relations among forms based on phonetic and/or semantic similarities. New forms can be produced by reference to existing forms, but most multimorphemic words are stored whole in the lexicon.
5. Lexical organization provides generalizations and segmentation at various degrees of abstraction and generality. Units such as morpheme, segment, or syllable are emergent in the sense that they arise from the relations of identity and similarity that organize representations. Since storage in this model is highly

redundant, schemas may describe the same pattern at different degrees of generality (Langacker 2000).

6. Grammatical knowledge is procedural knowledge. Anderson (1993) and Boyland (1996) distinguish declarative or propositional knowledge (e.g., 'Washington, DC is the capital of the United States') from procedural knowledge (how to drive a car, tie your shoelaces, and so on). While linguistic knowledge is in part declarative (in the sense that we can cite the meanings of words, for instance), much linguistic knowledge is procedural (Boyland 1996). A native speaker can form an acceptable sentence quite automatically, yet be unable to explain how this was done or to list what the properties of an acceptable sentence are. Thinking of grammatical constructions as procedural units has profound consequences for our view of phonology. Phonology then becomes a part of the procedure for producing and decoding constructions, rather than a purely abstract, psychological system.

1.3 The Creative Role of Repetition

Usage-based functionalism emphasizes language as a conventionalized, cultural object. In order to understand the nature of language, we need to understand what it means for behavior to be conventionalized. Haiman (1994, 1998) discusses grammar as ritualized behavior and points to various properties of both ritual and grammar that are the result of repetition. It is useful here to distinguish between a ritual and a convention: though both represent repeated behavior, a ritual can be individual and idiosyncratic, but a convention is agreed upon socially and evokes a consistent response in other members of a society (Tomasello et al. 1993). What both concepts have in common is that their structure is shaped by repetition. The following is a summary of some aspects of language that are shaped by repetition.

Through repetition we get lexical strength – strong, easily accessible representations, such as a greeting when you see someone you know or responses such as 'thank you' and 'you're welcome'; that is, any kind of learned automatic response. It is repetition that ritualizes these responses and makes them readily available. These are just extreme examples of a general phenomenon that pervades linguistic representation – repetition leads to strength of representation (Bybee 1985).

Repetition also leads to reduction of form. This is true of nonlinguistic gestures such as making the sign of the cross. It is true in nonhuman rituals: among chimpanzees (according to Plooij 1978, cited in Haiman 1994) the original gesture of lying down is reduced to just leaning slightly backwards. And it is true of language in many obvious cases. Greetings become reduced, (*how are you* becomes *hi*), grammaticizing phrases with increasing frequency reduce and compress (*going to* becomes *gonna*), and, in less obvious cases, there is a general frequency effect in reductive sound changes (see Section 1.4).

Repetition also leads to the reduction of meaning. This reduction or bleaching of meaning can be related to what Haiman calls habituation, or the loss of impact due to repetition. Habituation is also a general phenomenon, not restricted to language or to humans. It is 'a decline in the tendency to respond to stimuli that have become familiar due to repeated or persistent exposure' (Haiman 1994:7). We recognize habituation in the trivialization by repetition of great music (Beethoven's Fifth Symphony) or great art (Van Gogh's sunflowers). We also find it in language in cases where the emphatic becomes the normal. For instance, in the French negative construction *ne . . . pas*, *pas*, literally 'step', was once an emphatic added to the original negative *ne*, but is now obligatory and nonemphatic.

Finally, and perhaps most importantly, repetition leads to emancipation. In emancipation, instrumental actions are disassociated from their original motivation and are free to take on a communicative function instead. The military salute derives from the more instrumental gesture used in the Middle Ages when knights in armor greeted one another. They raised the visor of their helmet to show their faces as an indication of a peaceful greeting. The armor is gone, the visor is gone, but a reduced form of the gesture remains, though without its instrumental function. It no longer raises the visor, but it has been imbued instead with the function of communicating respect for the military hierarchy.

Applications of the principle of emancipation through repetition in language involve all sorts of cases of conventionalization, and most commonly, cases in which one communicative function is replaced by another. For instance, the inquiry into someone's current state of being, *how are you*, is not just reduced phonologically to *hi*, but also is emancipated from its original communicative value and now serves simply as a greeting. (A more conservative function of *hi* is found in some dialects of Black English where speakers commonly respond to *hi* with *fine*). Emancipation is also richly illustrated in the process of grammaticization during which words lose their categoriality. For instance, verbs become auxiliaries and sometimes affixes, and also become disassociated from their lexical meaning and take on pragmatic or grammatical functions, as when *be going to* loses its motion sense and becomes a future marker.

Haiman (1994) demonstrates that the development of ritual is a common process in the animal kingdom, and by no means restricted to humans, or even primates, as dog and cat owners can attest. He further argues (Haiman 1998) that ritualization is the basis for the development of grammar. The process of grammaticization depends upon repetition and is characterized by the reduction of both meaning and form, by strong entrenchment of patterns, and by emancipation in the sense that forms in their grammaticizing constructions often shift from propositional meaning to discourse-oriented functions (Traugott 1989). Our understanding of the ritualization process can be applied to syntax, as Haiman has shown, but also to phonology, as we investigate the role of repetition in the structuring of phonological patterns and lexical representations.

1.4 Frequency Effects

Much is already known about frequency effects in language, and much remains to be learned. In this section, I will lay out the basic notions and terminology that will be taken up again in later chapters.

There are two ways of counting frequency of occurrence that are applicable to language: token frequency and type frequency. *TOKEN FREQUENCY* is the frequency of occurrence of a unit, usually a word, in running text – how often a particular word comes up. Thus *broke* (the past tense of *break*) occurs 66 times per million words in Francis and Kučera (1982), while the past tense verb *damaged* occurs 5 times in the same corpus. In other words, the token frequency of *broke* is much higher than that of *damaged*.

TYPE FREQUENCY refers to the dictionary frequency of a particular pattern (e.g., a stress pattern, an affix, or a consonant cluster). For instance, English Past Tense is expressed in several different ways, but the expression with the highest type frequency is the suffix *-ed*, as in *damaged*, which occurs on thousands of verbs. The pattern found in *broke* has a much lower type frequency, occurring with only a handful of verbs (depending upon how you count them: *spoke*, *wrote*, *rode*, etc.). The penultimate stress pattern of Spanish has a very high type frequency, occurring with about 95% of nouns and adjectives that end in vowels (*abuéla*, *camíno*, *pronómbre*), while antepenultimate stress has a much lower type frequency (*claúsula*, *fonológica*). One can also count the token frequency of such patterns – that is, how often the pattern occurs in running text.

1.4.1 Token Frequency

Token frequency has two distinct effects that are important for phonology and morphology. In one frequency effect, phonetic change often progresses more quickly in items with high token frequency. This effect is particularly noticeable in grammaticizing elements or phrases that undergo drastic reduction as they increase in frequency. Thus *be going to*, which is becoming a future marker in English, is reduced to ['gʌɾ̃ə] or even further reduced in phrases such as *I'm gonna* to ['aiməɾ̃ə]. Similarly, the conventionalized contractions of English are reduced due to their high frequency: *I'm*, *I'll*, *I've*, *can't*, *don't*, *won't*, and so on (Krug 1998). But the effect occurs on a more subtle level as well: regular sound change in many cases progresses more quickly in items of high token frequency. There is a tendency in American English for syllabicity to be lost in sequences of unstressed schwa + resonant, as in *every*, *camera*, *memory*, and *family*. This reduction is more advanced in words of higher frequency (such as those just named) than in words of lower frequency, such as *mammary*, *artillery*, *homily* (Hooper 1976b). The loss of final [t] or [d] after a consonant is also more common in words of higher frequency, such as *went*, *just*, and *and*. In fact, a general effect of token frequency on the rate of deletion has been found for 2000 tokens of final [t] or [d] (Bybee 2000b).[2]

If sound changes are the result of phonetic processes that apply in real time as words are used, then those words that are used more often have more opportunity to be affected by phonetic processes. If representations are changed gradually, with each token of use having a potential effect on representation, then words of high frequency will change at a faster rate than will words of low frequency.[3] The streamlining of high-frequency words and phrases has the effect of automatizing production. Any motor activity that is repeated often becomes more efficient. The first effect of frequency, then, is to automate production (Boyland 1996). (For further discussion, see Chapter 3.)

---

[2] Further discussion of these examples and further references for other similar examples can be found in Chapter 3.

[3] This suggestion is found in Moonwomon (1992). Factors other than simple frequency are important, too. These are discussed in Chapter 3.

The second effect of frequency seems to contradict the first, since it makes items more resistant to change, but it concerns change of a different kind. High frequency encourages phonetic change, but it renders items more conservative in the face of grammatical change or analogical change based on the analysis of other forms (Phillips 2001). For example, high-frequency forms with alternations resist analogical leveling: while English *weep / wept*, *creep / crept*, and *leap / leapt* have a tendency to regularize to *weeped*, *creeped*, and *leaped*, respectively, the high-frequency verbs with the same pattern, *keep / kept*, *sleep / slept* show no such tendency (Bybee 1985, Hooper 1976b). As a result, morphological irregularity is always centered on the high-frequency items of a language. This conservatism of much-used expressions can also be found on the syntactic level (Bybee and Thompson 2000). It has often been observed that pronouns show more conservative behavior than full noun phrases. English pronouns, for example, maintain distinct forms for nominative and oblique case, while nouns have lost these case distinctions. The position of pronouns sometimes reflects an earlier word order. Similarly, verbal auxiliaries, which are very frequent, often retain conservative syntactic characteristics. The English auxiliaries, for instance, retain the ability to invert with the subject, and they precede rather than follow the negative, both properties once shared by all verbs (Bybee to appear).

This conserving effect of frequency places some items outside the domain of the regular combinatorial patterns of the language. Their frequency gives them a high level of lexical strength. That is, they are so engrained as individual patterns that they are less likely to change even if general changes are occurring in the language. To account for this entrenchment effect, I have proposed (Bybee 1985) that representations are strengthened whenever they are accessed. This strengthening makes them subsequently easier to access and also more resistant to some forms of change.

1.4.2 Type Frequency and Productivity

Another major effect of frequency and thus of usage is the effect of type frequency in determining productivity. Productivity is the extent to which a pattern is likely to apply to new forms (e.g., borrowed items or novel formations). It appears that the productivity of a pattern, expressed in a schema, is largely, though not entirely, determined by its type frequency: the more items encompassed by a schema, the stronger it is, and the more available it is for application to new items. Thus, the English Past Tense *-ed* applies to thousands of verbs and is much more productive than any of the irregular patterns, which are highly restricted in the number of verbs to which they apply (Bybee 1985, 1995; MacWhinney 1978; see also Chapter 5).

Since type frequency is based on the number of items matching a particular pattern, it is also relevant for determining the relative strength of phonotactic patterns, stress patterns, and other phonological patterns applying at the word level. Recent studies have shown that subjects' judgments of the relative acceptability of nonce items with occurring and nonoccurring phonotactic patterns are based on the distribution of these patterns in the lexicon. Patterns with high type frequency are judged to be more acceptable than patterns with low type frequency (Pierrehumbert 1994b, Vitevitch et al. 1997, and others; see Chapter 4.7). In addition, for classes defined by morphological, phonotactic, or stress patterns, type frequency interacts with the degree of similarity of the members of the class such that a smaller number of highly similar items (a 'gang') can also evince limited productivity (Aske 1990, Bybee and Moder 1983, Frisch et al. 2001; see also Chapter 5).

The importance of productivity of both phonological and morphological schemas to our understanding of cognitive representations of language cannot be overstated. Productivity provides us with evidence about the generalizations that speakers make, and it is important to stress that speakers' generalizations are not always the same as those devised by linguists on the basis of distributional evidence. Distributional evidence often recreates past stages of a language and does not reveal the restructuring and reanalysis that the patterns might have undergone. Productivity can be used as a diagnostic to determine which patterns are fossilized, and which represent viable schemas accessible to speakers.

1.4.3 Frequency Effects in Other Theories

The proposal that frequency of use affects representation suggests a very different view of lexical storage and its interaction with other aspects of the grammar or phonology than that assumed in most current theories. Structuralist and generative theories assume that the lexicon is a static list, and that neither the rules nor the lexical forms of a language are changed at all by instances of use. Similarly, as Pierrehumbert (1999) points out, all versions of Optimality Theory (Hayes 1999, Prince and Smolensky 1993, 1997) posit a strict separation of lexicon and grammar that makes it impossible to describe any of the interactions of phonology with the lexicon that are attested in the literature, many of which have just been mentioned: for instance, the fact that many phonological changes affect high-frequency items first, and the fact that the strength of phonotactic constraints is directly related to the number of items they apply to in the existing lexicon. Hammond (1999) identifies an effect of frequency in cases of the application of the Rhythm Rule and proposes that an Optimality Theory account of the facts can include item-specific constraints inserted into the constraint hierarchy. But such a proposal neither fits well with other properties of Optimality Theory nor does it provide an account for why words and phrases of different frequencies of use behave differently.

1.5 Phonology as Procedure, Structure as Emergent

If we conceptualize phonology as part of the procedure for producing and understanding language, the phonological properties of language should result from the fact that it is a highly practiced behavior associated with the vocal tract of human beings. To move away from the more abstract views of phonology, it is perhaps useful to compare speaking to other fairly complex but repetitive neuromotor activities, such as playing the piano. When a person learns to play the piano, he or she learns not just to strike notes, but to strike notes in sequence. Each piece of music has its own sequence of notes that must be learned. Practice is essential; the motor patterns that lead to the fluent, striking of longer and longer sequences of notes must be automated for a piece to begin to sound like music. With practice, the transitions between the notes become more fluent, and the speed of execution automatically increases. In order to maintain the correct rhythm and tempo, the player must at times hold back and not play every note as fast as possible.

An important result of learning to play several pieces is that new pieces are then easier to master. Why is this? I hypothesize that the player can access bits of old stored pieces and incorporate them into new pieces. The part of a new piece that uses parts of a major scale is much easier to master if the player has practiced scales than is a part with a new melody that does not hearken back to common sequences. This means that snatches of motor sequences can be reused in new contexts. The more motor sequences stored, the greater ease with which the player can master a new piece.

Storage and access to motor sequences implies categorization in storage. Even after the A Major scale is mastered as an automated unit, subsequences of the scale can still be accessed for use elsewhere. Connections are made between intervals that recur in different pieces.

Some analogies with the acquisition and use of phonology are obvious here. Children learn phonological sequences as parts of words, never independently of words. Articulatory routines that are already mastered are called forth for the production of new words, leading to a tendency of children to expand their vocabulary by acquiring words that are phonologically similar to those they already know (Ferguson and Farwell 1975, Lindblom 1992). This tendency leads to the structuring of the phonological sequences across words and the limiting of the potentially immense phonetic inventory. Put another way, the repetition of gestures and sequences across words allows relations of identity and similarity to develop in stretches of speech, giving rise to segment, syllable, and foot-sized units.

With practice, speakers become more fluent in stringing words together, and this fluency and automation is characterized by the smoothing of transitions and overlapping of movements constrained by the need to retain information value. Some repeated sequences become highly automated and reduced in form. At the same time, speakers must be able to access and recruit sequences into new combinations to express their thoughts and intentions. With practice, the ability to produce new combinations is also enhanced, probably by the storage of multiword constructional schemata (Bybee 1998, Pawley and Syder 1983). The use of novel combinations, however, does not constitute as large a percentage of spontaneous speech as one might suppose. Erman and Warren (1999) estimate that about 55% of the spoken and written texts they analyzed consisted of prefabricated multiword sequences.

Grammatical and phonological structure emerge from the facts of co-occurrence in language use. Words that commonly occur together – for instance, nouns and their determiners, or verbs and their objects –

begin to behave as constituents. The more commonly they co-occur, the tighter their constituency becomes (Bybee and Scheibman 1999; see also Chapter 6). Phonological structure is affected by use in that articulatory accommodations occur as the result of real language use. This is the sense in which grammar can be said to be emergent.

Emergence in language is much more complex than the emergence described earlier of the structure of the termite nest. The main difference is that human beings are much more intelligent than termites. First, the experiences of human beings in using language are registered in the brain, are categorized there, and gain some of their structure from categorizing capabilities of the mind. Second, the use of language by humans is goal-oriented or purposeful. The purpose is to communicate thoughts, perspectives, needs, desires, and so on. Note that the purpose of communicative acts is to communicate, not to create grammar. Yet the result of innumerable communicative acts is to change language and to create and recreate grammar (Bybee et al. 1994, Keller 1994).

Thus, functional constraints are manifested in specific languages through individual acts of language use. If there is a constraint comparable to the NO CODA constraint of Optimality Theory, it is a result of the phonetic tendency to reduce and coarticulate coda consonants more than onset consonants. This tendency manifests itself in every instance of language use in languages that have coda consonants, reducing these consonants by very small degrees. Eventually, coda consonants are lost in such languages, leaving a language with a reduced number of coda consonants or none at all. Since no comparable tendency operates on onset consonants, the result is that some languages lack codas, but that no languages lack onsets. Thus, the intersection of phonetic grounding and language-specific and typological properties takes place in the individual speaker in multiple instances of language use.

1.6 Organization of the Book

The goal of this book is to propose a unified account of some of the major empirical phenomena that have been examined in phonological theory in recent decades as well as less familiar phonological phenomena that provide evidence for the importance of usage in understanding phonological structure. In Chapter 2, I present the model of representation proposed for morphology in Bybee (1985) and show

how it can be used to accommodate phonological representation and to model various usage effects. Chapter 3 argues for the need for phonetic detail in lexical representations and shows how variation and sound change can be modeled in an exemplar representation. Chapter 4 discusses patterns of phonetic implementation in terms of articulatory gestures and argues for certain strong constraints on sound change. Here also, the acceptability of phonological patterns in the lexicon, or phonotactic patterns, are discussed in terms of their distribution in the existing lexicon. Chapter 5 presents what is known to date about the interactions of morphology with phonology; it includes a discussion of morphologization, of the differences between patterns based on morphology from those based on phonology, and of the role of type and token frequency and similarity in the formation of productive classes. Chapter 6 treats sequences larger than a word and demonstrates through the study of sound change in progress that words are the basic units of memory storage, but that longer sequences of high frequency as well as constructions can also be stored in memory and processed as chunks. Chapter 7 studies lexicalized sandhi alternations, using French liaison as the example, and demonstrates that the loss of liaison in certain contexts corresponds to morphological regularization in that high-frequency contexts maintain liaison longer. The implication of this finding is that phrases and constructions with liaison are stored in memory much as morphologically complex words are. Finally, Chapter 8 delves into the theoretical status of universals or crosslinguistic tendencies, arguing that there is an essential diachronic component to any attempt to explain linguistic structure or to account for universals of language.

1.7 Language as a Part of Human Behavior

A basic assumption of this book is that the cognitive and psychological processes and principles that govern language are not specific to language, but are in general the same as those that govern other aspects of human cognitive and social behavior. Our enormous memory capacity, fine motor control, the ability to categorize experience, and the ability to make inferences may be fine-tuned for language, but are all clearly used in other domains as well. This means that the principles underlying language could be studied using the methods and theories of psychologists or even biologists. It also means, however, that the discoveries of linguists could unveil principles useful in the more general understanding of human psychology. This book is a linguist's book: it applies the established methods and data of linguists to the understanding of language as an emergent system resulting from the general cognitive capacities of humans interacting with language substance over many instances of language use.