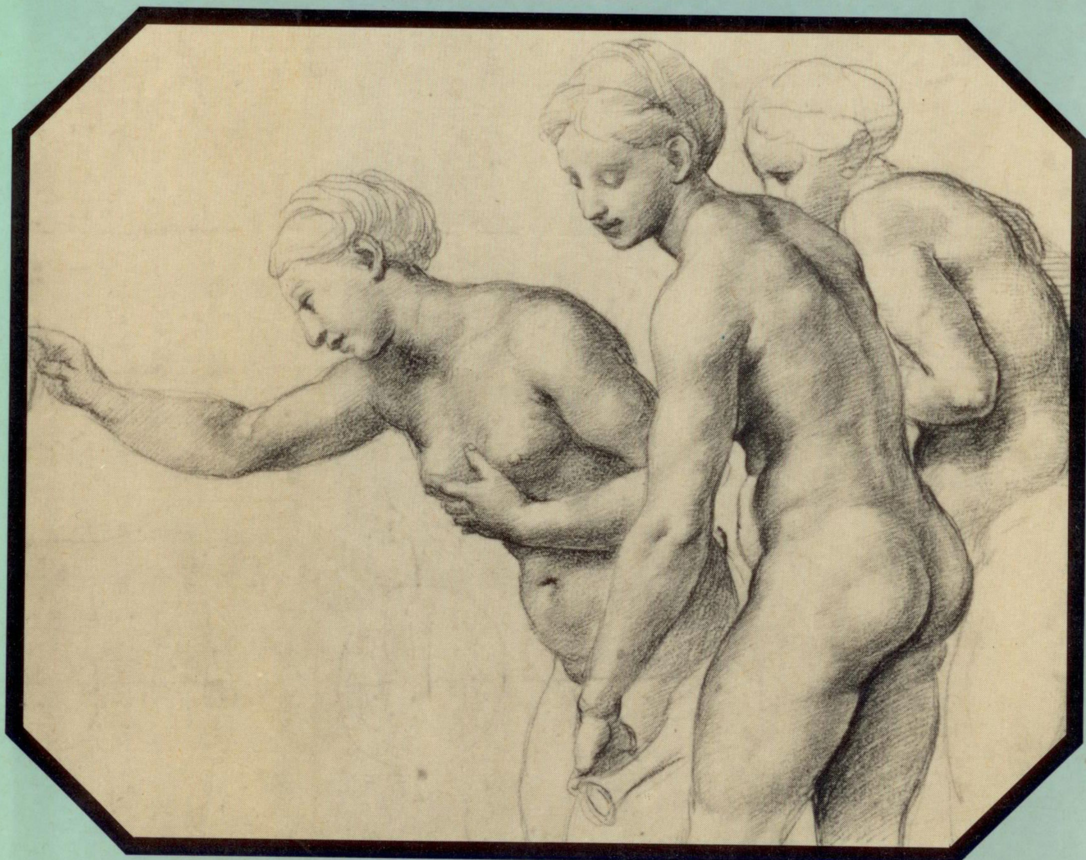
LANGUAGE AND NUMBER
JAMES R. HURFORD

Blackwell

LANGUAGE AND NUMBER

The Emergence of a Cognitive System



JAMES R. HURFORD

Language and Number
The Emergence of a Cognitive System
James R. Hurford

Language and Number is intended as a contribution to linguistic theory in the broadest sense. It offers a view of language (illustrated in detail through an examination of the linguistics of number) that brings together considerations of individual psychology and of communication within a speech community. These two strands, the psychological and the social, are put together to give an evolutionary perspective on language, which explains salient characteristics of its form.

The psychological considerations relate both to the invention and to the ordinary acquisition of language; the social considerations relate to the ways individuals negotiate common standardized expressions for their meanings. Languages, Professor Hurford argues, grow through the interaction of individual minds on the forms invented and socially negotiated by their predecessors.

The book also makes a contribution to the philosophy of number, arguing that our knowledge of number is a product of our possession of language and the faculty for constructing collections from aggregates. This sophisticated and original approach successfully maps out the various biological, social and cognitive factors that coalesce in the evolution of language.

James R. Hurford is Professor of General Linguistics at the University of Edinburgh.

Contents

Jacket illustration: *Study of the Three Graces*, c.1516, by Raphael, Royal Library, Windsor Castle, is reproduced by gracious permission of Her Majesty The Queen.

Jacket by Miller, Craig and Cocking Design Partnership
Printed in Great Britain

Basil Blackwell
Oxford and New York

ISBN 0 631 15568 6

To the three graces of the magic Number 14

LANGUAGE AND NUMBER

The Emergence of a Cognitive System

JAMES R. HURFORD

Basil Blackwell

First published 1987
Reprinted 1990

Basil Blackwell Ltd
108 Cowley Road, Oxford OX4 1JF, UK

Basil Blackwell, Inc.
3 Cambridge Center
Cambridge, Massachusetts 02142, USA

British Library Cataloguing in Publication Data
Hurford, James R.
Language and number : the emergence of a cognitive system.
1. Psycholinguistics 2. Grammar, Comparative
I. Title
401'.9 P37
ISBN 0-631-15568-6

Library of Congress Cataloging in Publication Data
Hurford, James R.
Language and number.
Bibliography: p.
Includes index.
1. Grammar, Comparative and general—Numerals.
2. Linguistic change. 3. Universals (Linguistics)
4. Numbers, Theory of. 3. Psycholinguistics. I. Title.
P275.H87 1987 401'.9 87 5167
ISBN 0-631-15568-6

Typeset in 11 on 12½ point, Bembo
by Photographics, Honiton, Devon.
Printed in Great Britain by
Athenaeum Press Ltd, Newcastle upon Tyne.

Contents

Preface

A working linguist who regards linguistics as a science shedding light on minds, languages, and language-communities as real objects, mental or social, should find this book congenial. Though I write from a linguist's background, I also hope that psychologists and philosophers (and perhaps even mathematicians) with an interest in language and number will find points in this book useful to them in the development of their own ideas, as I have drawn on theirs. But the 'hocus-pocus' linguist concerned only with constraining formalisms and the merely descriptive linguist, both of whom avoid questions of ontology and explanation, probably will not pick it up.

The book is intended as a contribution to linguistic theory in the broadest sense. It offers a view of language (illustrated in detail from one particular subsystem) which brings together considerations of individual psychology and of communication within a speech-community. These two strands, the psychological and the social, are put together to give an evolutionary perspective on language, which explains salient characteristics of its form. The psychological considerations relate both to the invention and to the ordinary acquisition of language; the social considerations relate to the ways in which individuals negotiate common standardized expressions for their meanings. Languages grow through the interaction of individual minds on the forms invented and socially negotiated by their predecessors. This growth of languages leaves traces from which the linguist, working like an archaeologist, a geologist, or indeed a nineteenth-century comparative linguist, can reconstruct the likely pattern of development. Turning the perspective around, he can use the pattern of development to explain why languages are the way they are now.

Somewhat to my surprise, the book also claims to make a contribution to the philosophy of number. The three central chapters, 3, 4, and 5, are at least as much about the nature of numbers as they are about language, and an argument is developed that our knowledge of number is in fact a product of our possession of language, plus a faculty for constructing collections from aggregates. An account is given of the rise of the number sequence and of constructions expressing the basic arithmetical operations of addition and multiplication, the latter by turning the techniques of denotational semantics onto numerals. While it may be convenient to talk about numbers as abstract Platonic objects, the account here shows a way of explaining the possibility of inventing and knowing such objects through linguistic devices. This in its turn could be fed back into arguments that languages, like numbers, are abstract objects, such as Katz has put forward. Once numbers, the Platonist's paradigm example of abstract objects, have been shown to develop *through* language, the argument that a language is a Platonic object becomes much harder to sustain.

Nevertheless, I do claim that languages are in some sense abstract objects, the results of the historical interaction of both psychological and social factors. But, unlike Chomsky, who also believes that that is what languages are, I believe that it is possible to say something interesting about the interplay of factors that gives rise to them, and thereby to begin to explain their form. A computational model of the social negotiation of standardized expressions is developed in the final chapter.

After I finished writing *The Linguistic Theory of Numerals* about twelve years ago, I thought I would not write another book on the same topic. And in a sense I haven't, even though this book is about linguistic theory with specific relation to numerals. I feel I have climbed the same mountain twice by completely different routes, seeing completely different landscapes on the way. The basic data for the two books are the same (the numeral systems of natural languages), but the books' objectives are completely different.

The Linguistic Theory of Numerals was an attempt, like many studies by people of my generation around that time, 'to extend and modify the detailed theory of generative grammar'. It shared 'the principal methodological assumptions about "doing linguistics" professed in such works as Chomsky (1965) and Chomsky and Halle (1968)'. There was a widely shared vision

in the 1960s of generative grammar as a monolithic cumulative corpus of propositions about universal grammar. But by 1975 work within generative grammar was splintering into many distinct subschools, all with different theoretical emphases and preoccupations, none of which happened to find the data from numeral systems particularly interesting. The data are still there and they are interesting to anyone who will take a close look at them.

The dissolution of this shared vision led many, including myself, to study the philosophical and methodological foundations of generative grammar, and there was a growth of interesting and provocative work in this philosophical and methodological vein. There were conferences and collections of papers on explanations in linguistics, evaluating linguistic hypotheses, the relation of data to theory in linguistics, and so on. Scholars as diverse as Itkonen, Katz, Fodor, and Chomsky himself produced a number of challenging monographs on the ontology and epistemology of linguistics. I published three small efforts in this vein (Hurford, 1977, 1979b and 1980). All of this work, though it made interesting and worthwhile contributions to thought, clearly offered no hope (if that is the right word) of reuniting generative grammar. And meanwhile, the practitioners in the splintered subschools were getting on with 'doing linguistics' in their own ways. I heard the opinion that all this discussion of the foundations of linguistics was of no use: the thing to do was to get on with the job of describing and explaining language(s).

Linguistics has grown too large and diverse for anyone to be able to articulate any uncontroversial set of foundational premises for it. As there are many metaphysical starting points, there will probably always be a wide range of schools of linguistic thought and research programmes in the subject. Still, there remains some consensus on what counts as illuminating and interesting in linguistics. Chomsky's most important influence, I believe, lies in his emphasis on how linguistic work should try to illuminate the subject matter, rather than merely describe it in tidy ways. We all differ widely in our views of what counts as an adequate explanation, but there is agreement that some kind of explanation is the goal of work in linguistics. In this book I offer a specimen of 'doing linguistics', that is a detailed discussion of data in a particular area, which is at the same time centrally concerned with explaining the data and considering the general form of explanations in linguistics. Thus this book attempts to promote a particular view of explanation in linguistics by working thoroughly with an example.

The previous book severely restricted the data it accounted for. I was aware of many impinging considerations which I resolutely kept at arm's length. In this book I open up these considerations, barely mentioned previously, but relevant to the wider task of explaining the structure of numeral systems. These topics include: the activity of counting, ordinal numerals, languages with no numeral systems, the integration of numerals into noun phrases and sentences, numeral classifier constructions, the denotational semantics of numerals, word-order universals, non-standardized numerals, the acquisition of numerals, psycholinguistic experiments on the perception of number, the evolution of numeral systems, and communicative interaction between speakers using numerals in a speech-community. I idealized away from consideration of these factors, as a way of searching where the light is brightest. Although I still value idealization and the present book contains many idealizations, I believe that taking these topics into consideration has produced a more satisfactorily explanatory account of why this particular subsystem of language has the characteristics that it does. Idealization is a useful research strategy, but progress is also made by trying to reach out into the dimness beyond our idealizations.

A preference for integration and synthesis is noticeable in this book. I try to bring things together: linguistics, psychology, and philosophy; language use and language system; synchrony and diachrony; what is valuable in Chomsky's ideas and in those of his critics. The result is not fusion (or, I hope, confusion), but statements of links and relationships between areas and approaches that are all too often isolated from each other.

A 'jilted lover syndrome' can be seen over the last decade in a number of books by former generative linguists. In these works, the adverse criticism of Chomsky's work is unconstructively shrill, and the condemnatory rhetoric often obscures the real issues. I have taken issue with Chomsky at several points in this book, but I have tried to avoid an irrational tone. I believe that a generative approach to language provides us with the most refined strategy yet devised for discovering linguistic structure, and I would not renounce this approach any more willingly than an experimental scientist would give up his laboratory. But simply working with sophisticated laboratory equipment does not make an experimental scientist, and simply describing linguistic

structure using the rigorous frameworks and argumentative structures associated with the generative enterprise does not in itself explain the nature of that structure. I do not believe that Chomsky's ideas on the innateness of certain linguistic principles are wrong, merely that they are not the whole story about language, and that an interesting additional story can in fact be told.

Acknowledgements

The Linguistics Department, the Centre for Cognitive Science, and the Centre for Speech Technology Research at Edinburgh University provided material and intellectual environments which made it possible and necessary, respectively, for me to do this work. The Edinburgh University Library proved to be a very good scholarly resource.

Grev Corbett has provided me with so many titbits of information on Slavic languages that it would be tedious to include '(G. Corbett, personal communication)' everywhere where it is appropriate. Any unreferenced data on Slavic comes from him, except in a few cases where Jim Miller has also helped me. (If the data are wrong, however, blame me.)

Many individuals have kindly (and sometimes unwittingly) made suggestions leading to the improvement of various points in this book, or given me data, references, or relevant work of their own. They include (with apologies to anyone whom I may have overlooked, or who would have preferred to have been overlooked): Lawrence Adewole, Darwish Al-Amadidhi, Ron Asher, Gwen Awbery, Ellen Bard, Ewan Barker, Ronnie Cann, Phil Carr, Ray Carrigan, John Chinnery, Grev Corbett, Annabel Cormack, John Crossley, Bob Dixon, Michal Ephratt, Norman, Vonny, and Simon Fairclough, Kit Fine, John Foster, Gerald Gazdar, Joseph Greenberg, Bob Hale, Robert Hetzron, Martin Hughes, Eve Hurford, Charles Jenewari, Siew-Yue Killingley, Ewan Klein, Bob Ladd, Roger Lass, Glenn Lean, Sebastian Löbner, John Maule, David McCarty, Jim Miller, Dieu-donné Moadjidibaye, Bill Nicoll, Trevor Pateman, John Phillips, Steve Pulman, Christiane Ravelomananana, Denise Rokosz, Len Rolfe, Graham Runnalls, Geoff Sampson, Paul Schachter, Lesley Stirling, Magda Stroinska, Sandra Thompson, Nigel Vincent, Henk Zeevat. I thank them all.

1

The Object of Study

1.1 Interest in Numerals

Some subjects hold an intrinsic fascination for their students because of the intricacy and beauty of their structures. For some people who successfully grasp an intricate system, the satisfaction of doing so is in itself sufficient. But whatever the intrinsic fascination of a subject, it gains in interest if it can also be shown to provide evidence and arguments relevant to debates in other subjects; and it gains even greater interest if it can be shown to bear on enduring philosophical issues. The set of subjects which are potentially interesting by virtue of their intrinsic structure and the light they shed on wider issues is far larger than the set of subjects which people have, to date, found interesting for these reasons. Within the set of potentially interesting subjects, what actually attracts attention is affected by such factors as accessibility of the relevant data and the possibly haphazard courses of intellectual and external history.

Natural language numeral systems have not figured largely in any of the (major or minor) intellectual debates of the twentieth century, but this seems to be an accidental omission. In fact, the numeral systems of natural languages, taken as a whole, show enough intricacy of morphosyntactic/semantic structure to be of purely intrinsic interest (although admittedly this is a matter of individual taste and judgement). More importantly, the structure of natural numeral systems turns out to yield a rich vein of evidence that can be brought to bear on central questions of the nature of language, the relation of language to mind and society, and the nature of number. To argue this convincingly, it is necessary to show a linguist's appreciation of the structural details of this particular type of subsystem of languages, along with a philosopher's concern for issues of general significance.

There are two popular stereotypes of the academic. One is the scholar who likes to delve into masses of detail, and values thoroughness and the exhaustive treatment of relatively narrowly circumscribed areas. The other is of the person concerned with ideas relating to matters of great significance, such as Freedom, Human Nature, the Universe, and so on, and apparently capable of sustaining discourse on these matters with very little reference to factual details. Most real academics are aware of these pernicious stereotypes and try to distance their own practice from either model. But steering between the rock and the whirlpool is not easy. In linguistics and related philosophical work one can find many examples either of work which describes particular (groups of) constructions in great detail with no attempt to relate the facts described to wider issues, or of work which philosophizes about the possibility of innate linguistic principles with little or no reference to detailed facts of language. On purely academic, intellectual grounds, both kinds of work are often of impressively high quality, whether in disciplined meticulousness, or in the firm command of abstractions. But there is a sociological problem: the two styles seldom interact with each other. Nitty-gritty linguists are sceptical of any philosophical discussion which does not 'get its hands dirty' by dealing with linguistic facts at something like the level of detail they are used to. And many philosophers who discuss questions involving language are not sufficiently aware of the sheer depth and complexity of linguistic problems, an awareness which only comes from grappling with particularities, at least for some of the time. (Philosophers have this kind of problem with practitioners of other disciplines beside linguistics, as well, of course.)

In this book, a particular area of languages, their numeral systems, is dissected in the depth typical of a linguist's enterprise, but the motivation for the dissection is always the investigation of issues of wider significance. In this introduction, it is only necessary to give a brief foretaste of the areas in which one might expect numerals to relate to issues of more general concern.

In several publications (1980a, 1980b, 1982), Chomsky has suggested a more or less close affinity between the human language faculty and the number faculty. Indeed, in the 1982 conversations he comes close to identifying one with the other, suggesting that what underlies both is a kind of computational complexity that is equipped to deal with discrete infinities (Chomsky, 1982, pp. 20–2). A more cautious position is expressed in the following:

> To gain further understanding of the general nature of the human mind, we should ask in what domain humans seem to develop complex intellectual structures in a more or less uniform way on the basis of restricted data. Wherever this is the case, we can reasonably suppose that a highly structured genetic program is responsible for the achievement, and we can thus hope to learn something significant about human nature by studying the systems attained. Language is an obvious area ...
>
> Are there other systems, more distinctively human in character, more enlightening as regards deeper and more fundamental characteristics of the human species? Perhaps so. Thus, one curious property of the human mind is our ability to develop certain forms of mathematical understanding – specifically concerning the number system, abstract geometrical space, continuity, and related notions. ... It is certainly possible to enquire into the nature of these abilities and to try to discover the initial state of the mind that enables these abilities to develop as they do. (Chomsky, 1980a, pp. 248–9).

I will argue later in detail that the number faculty largely emerges through the interaction of central features of the language faculty with other cognitive capacities relating to the recognition and manipulation of concrete objects and collections. The relevant features of the language faculty include the pairing of words with concepts by the linguistic sign (à la Saussure) and highly recursive syntax. It is therefore not necessary to postulate an autonomous 'faculty of number' as a separate module of mind.

Prima facie, however, numeral systems lie in the intersection of the human language faculty and the number faculty. One might therefore expect numeral systems to be a focus of considerable interest, for the kinds of reasons given by Chomsky. There are several possible reasons why this interest has not in fact materialized. A couple of arguments that could be advanced why the location of numeral systems in the intersection of the language faculty and the number faculty should not persuade one to find numeral systems particularly interesting can be summarized thus:

1.1.1 Numbers are a special area of meaning, unlike the other kinds of meanings that are conveyed in natural language, and therefore the structure of numeral systems is unrepresentative of linguistic structure generally. Numeral systems are thus of only peripheral interest to those interested in what might be called the central cases of linguistic structure.

1.1.2 The ordinary resources of natural language are only usable to provide names for a small subset of the actual numbers one might wish to name. For very high natural numbers, zero, negative numbers, irrational numbers, and real numbers, one generally needs to go beyond the resources provided by ordinary language, and invent technical notations. Thus, natural numeral systems cannot be revealing of the nature of number generally. And in any case, the names given to things are arbitrary and do not reflect the nature of the things named.

These arguments seem to me the most likely ones to be put up by a linguist and a philosopher, respectively, if asked for a good reason for paying no attention to numeral systems. Admittedly, they are straw men, but plausible ones, and the reader might need convincing that such arguments do not clinch the case.

Both arguments allege peripherality. The linguist says numerals might interest a mathematician but they are only of marginal interest to linguistics. The mathematician says they might interest a linguist but they are only of marginal interest to mathematics. If such views reflect mere sectarian prejudice, they deserve no credence. It would be rather like saying that the duck-billed platypus is not an interesting animal because it is not a central case of a mammal, or of a bird, or of a reptile, but apparently something in between. Viruses are very interesting to biologists because of, rather than in spite of, the fact that they fall only marginally, if at all, within the domain of living things. Areas where intellectual domains border on each other are of great importance. They can provide windows through which the central doctrines of any one domain may be viewed from the perspective of another. For the scholar, work in such an area is something of a high-risk enterprise, because understanding of concepts from more than one discipline is required. Consequently, studies in such areas might tend to disappoint more often than work in 'core' areas. And, of course, in an area where intellectual domains interact, the reader too is likely to have his imagination stretched more than usual. The first stretch of the imagination that is required is to see that apparent peripherality to the

established core of some intellectual domain need in no way imply uninterestingness.

In fact, the positions represented by arguments (1.1.1) and (1.1.2) are not compatible with each other. The first says, roughly, that numerals are clearly wierd, atypical of language generally, because the things they denote, numbers, are entities unlike the kind of entities dealt with in the rest of language, say persons, places, things, actions, states, and qualities. And obviously arithmetic is involved in the interpretation of numeral expressions, though nowhere else in language, so numerals are, prima facie, odd. But note that to argue in this way is to concede that there is some kind of systematic relation between the form of a linguistic subsystem and the class of denotata it involves, which is what the second argument, broadly, denies. There has to be a trade-off between the two arguments. To the extent that numeral systems are peripheral to the core of language because they deal with mathematical concepts (a strange 'because'!), they are likely to be less peripheral to the study of number. And to the extent that numerals are peripheral to the study of number, because they are essentially linguistic, they are likely to be less peripheral to the study of language.

Numeral systems are in clear ways well integrated with the languages in which they are embedded. In the stream of speech, numerals receive no special attention, making use of the same phonological units (say phonemes) and processes (phonological rules) as the rest of the language. Phonologically, nothing distinguishes the numerals in the following transcription.

1.1.3 /ðərarθrihʌndrədnsIkstifaivdeizInəjir/

This is less true of written language, where the notation for numerals may often differ from that used for the rest of the language.

1.1.4 There are 365 days in a year.

But the use of an alternative notation for numerals is seldom, if ever, obligatory, and conventional orthographic forms exist. 365 can be written out as *three hundred and sixty five*. The alternative notation can be seen as an efficient shorthand for the longer forms, although it is no doubt significant that such shorthands are especially common for numeral expressions. But there are other shorthands, such as e.g., i.e., &, +, @, £, =, %, in quite common use.

The internal syntax of numerals is such that non-numeral elements are not usually interspersed with the constituents of numeral expressions. But there are cases where numeral and non-numeral constructions interpenetrate each other, as in the now archaic English *threescore years and ten*, where *threescore . . . and ten* is a discontinuous numeral constituent, interrupted by the noun *years*. Constructions such as this are common in some other languages, for example the dialect of Welsh described in Hurford (1975). Numeral expressions usually have a well-defined place in the larger constructions, for example noun phrases (NPs), in which they are embedded. In English they go between a determiner and an attributive adjective in an NP, as well as in certain predeterminer quantifying phrases. As has been pointed out by Corbett (1978a, 1978b), in many languages numeral expressions share distributional and morphological characteristics with nouns and/or adjectives, depending on their value. In Chapter 5, I shall argue that numerals are primarily adjectives, and secondarily nouns, and that the principal differences (for example in word-order) between numerals and (other) adjectives and nouns derives from their characteristic semantic denotations. Thus, numerals have a clear place in the syntactic organization of languages.

In one clear respect, numerals are unlike almost anything else in language. Numeral expressions are ordered, in the counting sequence. One cannot talk of the order of the NPs or of the sentences in a language, in the sense of saying what the first NP is, or the second, and so on. Of course, one can actually decide to ignore the ordering of numeral expressions and treat them as an unordered set, just like the set of NPs, or prepositional phrases, or whatever. This was done in Hurford (1975). But ignoring anything so salient is in principle unwise.

In descriptive and pedagogical grammars, numerals are usually given a chapter or section of their own, simultaneously indicating that numeral systems are naturally regarded as belonging to the language in question, and that they are to an extent self-contained and have distinct characteristics. Some languages have no numeral systems (Dixon, 1980, pp. 107–8), so clearly a numeral system is not an essential part of a language, but many other subsystems of languages of types which are not strictly universal have aroused much attention in linguists. Not all languages have a case system, some do not have a system for overtly marking the times of actions and events (although this can be achieved by circumlocution), many languages have no adverbs, some have no adjectives, and many languages make no use of lexical tone. But case systems, tense systems, adverbs, adjectives, and tone systems are all regarded as worthy of the linguist's attention, if their study illuminates the nature of language in some general way. Perhaps nobody has yet happened to see how a study of numerals does shed general light on questions of the nature of language and number.

The theoretical possibility that numerals may point to certain specific conclusions about human knowledge of number is argued in Section 2 immediately below; the conclusions themselves will be put out for inspection in Chapters 3, 4, and 5. A study of numerals also points, I believe, to a number of conclusions about the nature of human language. The third and final section of this chapter will outline a view of language which will show, I hope, the sense in which the study of this type of linguistic subsystem can contribute to the study of language as a whole. The outlines to be given in the next two sections will be fleshed out by example and precept in the rest of the book.

1.2 Language, Psychologism, and the Nature of Numbers

The status of numbers (as opposed to numerals) is a question at the heart of the foundations of mathematics and thus of the foundations of the physical sciences as well. Whatever numbers actually are, I will assume that numerals are used by people to name them (using 'name' in a non-technical sense, that is not implying that numerals are logically names, as opposed to, say, predicates or quantifiers). I also assume that for the vast majority of exact numbers, knowledge of them is only accessible via some numeral expression; most, if not all, numbers are not known without knowledge of some linguistic expression naming them. For the present, I give no detailed answer to the converse question of whether it is possible to know a numeral without knowing the number it names; the possibility is certainly not logically excluded. Given these initial assumptions, it is natural to claim, as I do, that a careful consideration of numeral systems and the manner of their evolution and acquisition sheds some light on the question of the manner in which numbers can be said to exist. Obviously, if one is to argue this, one must start from an assumption that numbers do indeed exist in some sense. But beyond this, we want to make no initial assumptions about just what kind of entities numbers are, although it is clear enough that they cannot be physical particulars. So, for the time being, a number is simply something that can be named by a numeral expression. (The fact that I have implicitly defined 'numeral' as any expression that denotes a number gives an unavoidable circularity. I assume, nevertheless, that you know what you are reading about.)

Languages do not treat all numbers even-handedly. In some languages, some numbers are not named, or are named only with difficulty or uncertainty. (And in some languages even, no numbers (except possibly 1) are named at all, that is these languages actually have no numeral system.) There are definite (rather obvious) constraints on the nameability of numbers across languages. Such constraints may be expressed in the general form:

1.2.1 If a language has a name for a number x, it also has a name for a number y.

Now if a person is able to name something, this reflects at least some minimal knowledge of that thing by the person. And it is plausible to suggest that the greater nameability of some numbers as opposed to others reflects differences in the relative accessibility to knowledge of various numbers. Furthermore the manner in which languages name particular numbers repeats itself significantly. Thus, we can state general tendencies of the following form:

1.2.2 A number x is named by an expression whose constituents are the names of the numbers y and z.

This also suggests that the numbers y and z are in some sense more accessible to knowledge than the number x. It is interesting to note that the number 2 is never (standardly) named by an expression like *one plus one*, although the number 11 is, not surprisingly, often expressed as something like *ten plus one*. Such facts could be applied to a constructivist theory of the nature of numbers, where the tools of construction are actually linguistic devices, i.e. words and grammatical constructions. I will in fact advance a view that as far as the evidence from language is concerned, the number 2 is not known by linguistic construction

from the number 1. Higher numbers, on the other hand, come to be known through the possibility of forming expressions for them by making use of certain linguistic devices, such as a rote-learnt counting sequence and the prior availability of certain grammatical constructions, such as conjunction. Thus, attention to numeral systems, that is to the systems of names which have been given to numbers in the various languages of the world, can contribute to the debate about the nature of numbers themselves.

Approaching numbers (whatever they are) from a linguistic study of numerals leads one through a consideration of the truth conditions of natural language sentences containing numerals (for example *I saw three men*, *Seven plus five equals twelve*, and French *Nous sommes quatre*). If a satisfactory account can be given of all usage of such sentences, and of the possibility of acquiring the mental rules and representations determining this usage, I assume that no further major questions about the nature of numbers remain to be answered.

The view that a study of natural language can illumine mathematics and logic is 'psychologistic'. Psychologism in this area has been attacked, most notably by Frege, on the grounds that it confuses objective facts about the real world (which may include abstract, but still real, objects) with subjective ideas. It has been stressed, often in critiques of Locke, that ideas are essentially private and that it is in principle not possible to identify the same idea in the minds of different people. Clearly there are some mental entities to which this criticism correctly applies. But it is absurd to claim that there are no shared psychological phenomena. How, otherwise, can communication between people, above a phatic or merely stimulus-bound level, take place? There is a public, conventional shared fixing of the objects about which people communicate, many of which are by no means simply given by the external world. And this fixing, both the process and the results, can be studied objectively.

The languages of the world can be seen as a vast laboratory in which billions of subjects have brought their native abilities to bear on the task of decoding the signals they receive from other subjects, composing signals to express their own messages, and contributing collaboratively to the development and elaboration of the codes they find. Languages are the accumulated products of millennia of subjects' responses to these challenges. Any significantly repeated pattern in the ways languages express particular content must tell us something about the interaction between native abilities, the nature of the content expressed, and the exigencies of social communication.

Without quarrelling with the view that numbers are in some sense abstract objects, one needs to ask the question how humans manage to get to know these objects, since they cannot be perceived by the senses

> How can I educate a child to make sure he gets reference to numbers right? ... Invoking mathematical intuition here would be like saying 'Human minds have access to a fifth dimension in which the cardinal numbers are strung out like perfect pearls, and our mental fingers can just point to them in order to fix the references of our number words' – a charming metaphor perhaps, but not even a start at an answer. (Hodes, 1984, p. 134)

The acquisition question has been approached surprisingly slowly by philosophers of number. Kitcher's recent large contribution (1984) stands out in the literature as 'fresh', 'original', 'debunking', 'alternative', according to reviews quoted on its cover. As Hodes shows, Frege became increasingly worried in his later years about how to reconcile his view of numbers as abstract self-subsistent objects with their acquirability, but such worries certainly cannot be said to have dominated Frege's major arguments on numbers. The acquisition problem is a crucial one to be solved in any account of the nature of numbers. Probably the acquisition problem is the source of the misgivings of many philosophers (for example Benacerraf, 1965; Field, 1980; Hodes, 1984) who have discussed the view of numbers as abstract objects. But I feel that in their efforts to construct an alternative view they are sometimes unable to shake themselves free of the spirit of the Frege/Russell tradition of enquiry in this area.

Philosophers who write about number are typically qualified to do so by an impressive command of mathematics, so that abstractions beyond the natural numbers, such as real and irrational numbers, are never far from their minds. But they are less interested in psychology, even of the purely speculative variety, and psychological considerations are indispensable to a solution of the acquisition problem. Many philosophers of number are by tradition occupied with the possibility of *defining* number(s), often in terms of classical set theory. But this focus on definition is

alien to an empirical enquiry into the psychological bases of knowledge of number and its acquisition. Definition (as opposed to desirable pretheoretical characterization) pre-empts empirical enquiry. Often the philosopher's definiens, classical set theory, is itself taken as given, its psychological validity unquestioned; but there is room to doubt the psychological basis of aspects of classical set theory, such as the distinction between individuals and one-member sets. By contrast, while trying to meet the work of philosophers, I will approach from a different direction, taking into account both psychological work on knowledge of number and the linguistic evidence of numeral systems for the ways humans (get to) know numbers.

Linguists also, especially those working closely with logic, tend to be so impressed with the apparent solidity of numbers that they assume a linguistic version of the 'fifth dimension ... perfect pearls' view derided by Hodes above. For example:

> let us assume that we have defined the set *K* of *finite cardinal numerals*. That is, *K* is the set {*zero, one, two*, ...}. (So we ignore whatever rules English has which forms say *two hundred forty six* from *two, four*, and *six*.) ... [In principle this] approach would treat *K* as a new primitive category whose type would be the natural numbers. (Keenan and Faltz, 1985, pp. 228–9)

Note also the many linguistically oriented logicians (to be mentioned in Chapter 4) who take as given an infinite series of subscripted existential quantifiers $\exists_1$, $\exists_2$, $\exists_3$, ..., corresponding to natural language numerals. Such approaches ignore the obvious fact that numerals are formed by rules out of a small finite vocabulary. They thus shed no light on a central problem for psychological studies of both the human language faculty and human knowledge of number, namely the ability to make infinite use of finite means.

It will be maintained throughout this study that language is a necessary instrument for the passing on of knowledge of number, and, furthermore, for the original invention of the abstract objects known. I prefer to say that numbers are invented rather than discovered. Invention (and failure to invent) always involves discovery, in the sense that one discovers that exemplars of one's invention will (or will not) perform in the ways foreseen. So to say that something is invented is not to deny that it, and facts about it, are discovered. But there can be discovery without invention, as in the prototypical cases of America or the source of the Nile. Invention typically involves a creative act of putting together existing elements (which may or may not be physical) in some novel way. Furthermore, what is invented is not a particular object. We say 'Bell invented the telephone', not ' ... a telephone', or ' ... telephones'. The definite article here is generic. What Bell did was to bring into existence a class of objects over and above any physical prototype telephones he actually built. Discovery is often more particular. I shall speak of numbers as a class of abstract objects invented by the first people to use numerals. I shall argue that the pre-existing elements put together in a novel way during the creative act of the invention of numbers include elements of language. Therefore there is something in the nature of language which fits, or opens onto, the nature of the invented objects.

The reason why systems of names, such as numeral systems, have not often been used as clues to the nature of the entities they name lies in a tendency, outside linguistics, to ignore the *system* of languages. In the case of simple names, it is obvious that, with the marginal exception of onomatopoeia, a name reveals nothing of the nature of the entity named. Close inspection of the sequence of letters S I X gives no insight into what, if anything, it refers to. Saussure is well known for his insistence on this arbitrariness of the sign, but his discussion of the 'motivation' of signs is less often noted.

> The fundamental principle of the arbitrariness of the sign does not prevent our singling out in each language what is radically arbitrary, i.e. unmotivated, and what is only relatively arbitrary. Some signs are absolutely arbitrary; in others we note not its complete absence, but the presence of degrees of arbitrariness: *the sign may be relatively motivated.* (Saussure, 1959, p. 131)

Interestingly, Saussure chooses to illustrate this with numerals:

> For instance, both *vingt* 'twenty' and *dix-neuf* 'nineteen' are unmotivated in French, but not in the same degree, for *dix-neuf* suggests its own terms and other terms associated with it (e.g. *dix* 'ten', *neuf* 'nine' ...). (p. 131)

The fact that *dix-neuf* means what it does and is put together out of *dix* and *neuf* is evidence of the perceived (or intuited) relationship between the respective entities named by the whole expression and the parts. More significantly, the fact that very many languages form expressions for 19 in an exactly parallel way is evidence for the community of these perceptions (or intuitions) across human populations. But lest it be thought that these relationships are *necessary* in some abstract sense, it should be pointed out that there exist a tiny minority of cases in which languages seem to have opted for a different way of arranging things. In Hurford (1975), instances are given of 'correct misinterpretations'. Here is one example:

> the Hawaiian word for 20 is *iwakalua*. *Iwa* is Hawaiian for 9, and *lua* is the word for 2. Humboldt noticed this discrepancy: 'Man kann das Zahlwort 9 (*iwa*) und 2 (*lua*) nicht verkennen, und müsste also annehmen dass hier eine Verwirrung der Begriffe stattgefunden und man 9 x 2 gesagt hätte'. (Humboldt, 1832–9, pp. 776–7).

Other examples include a case from a Bantu language, described as follows by Seidenberg:

> The almost universal word for 5 in Bantu is *tano*, or *tanu*, sometimes abbreviated to *tan*. ... The almost universal word for 3 in Bantu is *-tatu*, also frequently found in the form *-datu*. Then *tandatu* clearly derives from 5 + 3 = *tan* + *datu*. This etymology is quite clear and would no doubt readily be accepted but for the fact that *tandatu* means 'six' ... (Seidenberg, 1960, p. 255)

The point of citing such examples is to concede a millimeter or two to the Fregean argument that human ideas of number may apparently deviate from what are regarded as absolute mathematical truths; in dealing with numeral systems, we *are* dealing with human psychology. But the extreme atypicality of such cases, and the conformity of numeral systems the world over to 'standard' arithmetic, shows that the human psychological factors at work converge so significantly on certain patterns that these patterns can be taken as objects and studied objectively.

A research area ripe for development is the study of human knowledge of the world as revealed in the naming tendencies of natural languages. For example, to the logician, both common nouns and adjectives correspond to one-place predicates, with no logical distinction made between the types of meaning they convey. But obviously the traditional notional definitions of a noun as naming a person, place, or thing, and of an adjective as denoting a quality, have plenty of substance. The original barrier to the research programme suggested was the lack of any suitable non-notional definitions of syntactic categories, such as noun and adjective. But if syntactic classes are first defined language-internally in distributional terms, the core of their membership can then be compared semantically, and the vicious circularity can be avoided. More generally, despite the obvious lack of any iconic relationship between simple words and the entities they name, there are many instances where linguistic *structure* (in both syntagmatic and paradigmatic senses) seems clearly to be iconic in some broad, and non-naive sense. 'The order of elements in language parallels that in physical experience or the order of knowledge' (Greenberg, 1963a, p. 103). See also Kempson (1975, p. 56) where, following Grice, the order of conjoined sentences is related to the order of the events they describe. There are no doubt various subtypes of this broad iconicity; for recent detailed advocacy of iconicity, see Haiman (1980) and Hopper and Thompson (1984). The presence of such iconicity can be used to fuel arguments in two directions: universal patterns in linguistic structure can be explained if it can be shown that they are iconic to known universal patterns in human experience or culture (explanations of this kind are to be found, for instance, in Hyman, 1984 p. 78, and Comrie, 1984 p. 89); and, going in the other direction, linguistic universals can be included, because of iconicity, among the tools used to try to discover cultural and conceptual universals of an abstract nature.

A limited version of the idea of using a linguistic system to reveal the structure of speakers' knowledge of the world it describes is put into effect in anthropology, where the taxonomies imposed on the natural and social world are studied through the structure of, for instance, plant and kinship terminology. Some such studies draw conclusions about a single culture from a single language, whereas others are interested in general conclusions about human conceptions in the domains concerned. I hope in parts of this book similarly to draw several universal conclusions about human knowledge of number from a study of natural language numeral systems.

1.3 The Study of Language Systems

It will be useful to comment briefly on what is understood here by the phrase 'the study of language'. Chomsky makes the (perhaps deliberately) provocative remark that 'the notion "language" itself is derivative and relatively unimportant' (1980a, p. 127). Nevertheless the phrase 'the study of language' recurs as a theme in most of his linguistic writings, and in these contexts Chomsky consistently urges the pursuit of this study. Evidently, he is not advocating the study of something derivative and unimportant. He admits (1986, p. 28–9) to 'questionable terminological decisions' involving 'language' and 'grammar' in his early work on generative grammar. The 'derivative and unimportant' remark is unfortunate, and likely to be misinterpreted as indicating that Chomsky has no interest in empirical confrontation of his theories with language data. Any individual (part of a) language, French, for example, (and, *a fortiori*, the French numeral system) is, considered in isolation, relatively uninteresting. I have often given talks to university linguists on general organizational characteristics of numeral systems, and been struck by how frequently members of the audience note, in connection with no general point, that the French numeral system has the remains of a 20-based system in the expression *quatre vingts*. And then someone else will usually chip in with the information that in parts of French-speaking Belgium and Switzerland a purely decimal system with *septante*, *octante*, and *nonante* is found. Such facts, purely in and of themselves, are relatively uninteresting from the perspective of this study. What, on the other hand, is relatively interesting is the set of principles determining the structure of French and other languages. If some aspect of French (such as, perchance, its numeral system) can be shown to shed light on general organizational characteristics of languages, *that* begins to be relatively interesting.

I take the position that the study of language starts most naturally with languages. Languages are the objects partially described in traditional and pedagogic grammars. Although they are fuzzy at the grammatical and geographical edges, languages are in practice sufficiently clearly defined to be susceptible to solid factual description, which in turn can support theorizing. Discussions of how many languages are spoken in the world correctly hedge their statements by pointing out the indeterminacy involved in the language/dialect distinction, but are usually not thereby deterred from concluding that there are in the region of 4000–5000 of them. So languages are, at least roughly speaking, countable. Languages also loom large enough as individual entities in our consciousness to be nameable. Linguists drop the names 'Quechua', 'Walbiri', 'Hanunoo', 'Basque', and so on, as unselfconsciously as the names of people and places. Laymen know fewer languages by name, but it is clear that languages are very generally known as individual, describable, countable, and nameable objects. And it is with these pretheoretically available objects that the study of language can get to work.

Languages are clearly not to be defined in political terms. I know of no political unit larger than a parish of which all members speak the same language. To identify languages with political units generally is to associate oneself with such 'common-sense' but ignorant beliefs as 'The language of Yugoslavia is Yugoslavian', or 'In Switzerland they speak Swiss'. (The influence of such common-sense beliefs can be remarkable. My wife, a native American English speaker, now resident in Britain, was once hospitalized while on a visit to America and asked by a nurse whether she could understand English, because on the hospital registration form she had given her residence as 'England'. The nurse presumably worked on the common-sense belief that the language of the USA is English and the languages of other countries are different.) Both the USA and Britain have more than one language, and share a dominant language. The language-to-state relation is many-to-many. As politicians and historians know, the lack of fit between political units and language communities can be a potent and deep-seated source of conflict. Geographical labels, such as 'English', 'French', and so on, are generally a mere convenience, and sometimes an inconvenience, just as 'African' and 'Indian' help the zoologist to refer to different species of elephant. There is nothing intrinsically African in the African elephant. And there is nothing in the linguist's conception of a language that intrinsically connects it to a particular geographical or political entity. The linguist's pretheoretical conception of a language is an advance on the common-sense notion. The linguist will take into account such factors as as roughly common vocabulary, sound system, and surface grammatical structure. The pretheoretical notion is rough and ready and subject to revision. But often, I believe, the original identification of a

language as an object of study survives in a recognizable form the revisions brought about by theorizing.

Languages, as identified thus pretheoretically, are not spatio-temporal particulars like people, stones, or dewdrops. They are in some sense abstract objects. But there are clear, agreed-upon ways of discovering the facts about them. If I want to know whether modern Greek has resumptive pronouns, I know how to go about finding out. I go to the library and try to find the information in a grammar or handbook of modern Greek, or I try to find native speakers, or both. Using handbooks is no more than a short cut to consulting native speakers, since the handbook authors have usually gained their information from native speakers. If such sources of information disagree irreconcilably, there is a problem. Then one has to say that the sources are referring to different objects, using, misleadingly, the same name. Here the language/dialect indeterminacy appears, and one has to decide which of the several objects that now appear to exist the original question was about. Linguists are in general sufficiently concerned with replication of results and public criteria for the evaluation of their theories to restrict their studies to objects on which there is significant agreement (about the data!). Research questions which fan out into a maze of idiolects are not pursued by linguists whose primary aim is to investigate the structural principles of the objects known as languages.

Professional descriptive grammarians in a long tradition have, while acknowledging variability, found little difficulty in identifying the central objects of their descriptions. The following quotations make the traditional descriptive grammarian's attitude more explicit than many grammars do, but it seems that the attitude is quite general, and the approach quite satisfactory as a basis for reaching sometimes ambitious descriptive goals.

> we need to see a common core or nucleus that we call 'English' being realized only in the different actual varieties of the language that we hear or read. (Quirk et al., 1972, p. 13)

> The fact that in this figure the 'common core' dominates all the varieties means that, however esoteric or remote a variety may be, it has running through it a set of grammatical and other characteristics that are present in all others. It is presumably this fact that justifies the application of the name 'English' to all the varieties. (p. 14)

And even for a language as variable as Arabic:

> The term 'Arabic' is applied to a number of speech-forms which, in spite of many and sometimes substantial mutual differences, possess sufficient homogeneity to warrant their being reckoned as dialectal varieties of a single language' (Beeston, 1970, p. 11).

Ranged against such views, which emphasize notions such as 'central case' and 'core', one encounters the opposing view, which emphasizes notions such as 'demarcation' and 'boundary'. Where, precisely, does German stop and Dutch begin? The question cannot be answered in any way which relies on generally applicable linguistic (as opposed to political) principles. These familiar considerations lead Chomsky to assert, 'There is nothing in the real world corresponding to language' (Chomsky, 1982, p. 107). (The context makes it clear that Chomsky intends *language* the count noun, not *language* the mass noun.)

Clearly, there is a dilemma here. But one should resist facile conclusions. Nobody is gullible enough to follow Zeno to the apparently logical conclusion of his arrow paradox, namely that the arrow is stationary. And the problems of demarcation between languages and dialects should not impel us too hastily to the conclusion that languages do not exist. If a case could be made that languages do not exist, it would presumably be along the formal lines of an argument, which seems to be correct, that colours do not exist, unless one is prepared to accept a non-denumerable infinity of them. There is a continuously variable spectrum of colour, and one 'colour' merges imperceptibly with the next. Putting aside whatever in our perceptual apparatus makes some 'colours' (for example, red) salient to us, we may agree that in the reality beyond our everyday perceptions it appears that, though there is colour, there are no 'colours' in the sense of naturally occurring, in some sense finitely bounded entities.

But is the case of language(s) like that of colour, or is it more like the case of biological species? Here, in the main, it is agreed that natural classes exist, even though there are demarcation problems not unlike the usual Norwegian/Swedish case often cited in the linguistic argument. Are lions and tigers two species or one? They can be successfully crossbred. The zoologist will agree that it is difficult to define the notion of species precisely, and that in some sense different species sometimes seem to merge into one another. But you will not get the zoologist to give up the idea of species altogether.

The case of language(s) is, naturally, not exactly analogous to either that of colour(s) or that of species. But the case of species shows that the existence of boundary-drawing problems, as between lion and tiger, or Norwegian and Swedish, does not constitute a knock-down argument that the entities whose boundaries one is attempting to locate do not exist (or are 'epiphenomena', or 'not significant realities', or whatever banishing terminology one chooses). Certainly, the question of just what languages are is problematic. The answer is probably very complex, and probably different in at least some respects from the pretheoretical, 'everyday sense' of the word *language*, to which one need not adhere slavishly. It seems certain that social considerations will form part of the answer, and that languages will turn out to be, in some sense, abstract objects – although what I have in mind is quite far from Katz' (1981) view of languages as abstract objects.

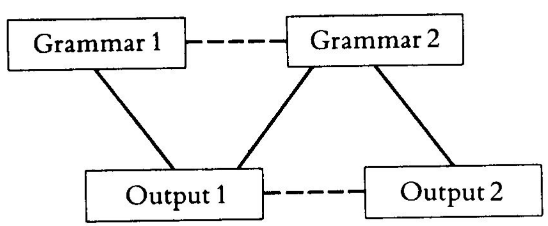
The view that a language is a social object of some sort is frequently met, but seldom articulated in a way which satisfactorily explains how such social objects interact with psychological representations in the minds of individuals. Saussure's *Cours* tantalizingly maintains both an emphasis on the social nature of *langue* and an emphasis on its psychological nature. 'Language exists in each individual, yet is common to all' (Saussure, 1959, p. 19). But Saussure did not present a model showing satisfactorily how a language could be seen simultaneously as belonging both to individuals and to the community. Indeed, Saussure's use of 'social', 'individual', and 'psychological' can be downright puzzling to a modern reader. 'The study of speech is then twofold: its basic part – having as its object language [langue] which is purely social and independent of the individual – is exclusively psychological, (Saussure, 1959, p. 18). Saussure was willing to speak of 'the collective mind of the community of speakers' (1959, p. 153). For most modern linguists, and certainly for Chomsky, psychology only concerns individuals, and he concentrates on the grammar of a language as possessed by an individual, with no attention paid to theorizing about the role of the community in maintaining the desirable closeness between the grammars of individuals. On the other hand, an unfortunate tendency simply to state that a language is a social entity, without enquiring into its relation to individual minds, can be heard from such scattered sources as the following:

> *The historical existence of a language is as much a part of it as anything an individual knows, at any particular time.* ... Languages have an existence in some sense independent of their speakers: that is, they have *traditions*; perhaps more accurately, they *are* traditions. (Lass, 1984, pp. 4–5)

> Languages are complex properties of human societies, not of individual brains. (Ladefoged, 1980, p. 502)

Fortunately, a more subtle view has begun to appear, in several publications by Pateman (1983, 1985, 1987) 'through time the content of mentally represented grammars, which are not in my view social objects, comes to contain a content which was in origin quite clearly social or cultural in character' (Pateman, 1985, p. 51). Pateman draws on work by Andersen (1973) which stresses that the mode of diachronic change in language is properly seen as represented by the solid diagonal lines in (1.3.1), and not by the horizontal broken lines.

1.3.1



In other words, grammars do not beget grammars: they are used in producing linguistic output, which in turn may constitute the primary data for the abduction of new grammars in the next generation. Lightfoot (1979, pp. 141–154), also drawing on Andersen's article, argues plausibly that for this reason there can be no formal theory of the possible differences between successive grammars. But this is not to say that we cannot formulate insightful and even explanatory statements about the ways in which languages change.

One could continue diagram (1.3.1) indefinitely across the page, with the zigzag causal line going alternately through

successive grammars and successive outputs. I believe that this view of linguistic change is similar in its essentials, *mutatis mutandis*, to that of Saussure and his editors (see especially Saussure, 1959, p. 143n). The middle diagonal line in diagram (1.3.1) is the line on which Chomsky's language-acquisition device (LAD) box sits in his well-known diagram.

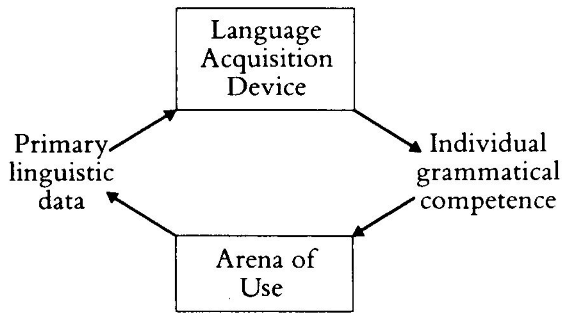
1.3.2

That is, the innate language faculty builds mental structures out of the raw ('degenerate' and so on) material of performance.

But I would claim that the input material is not so raw, having been processed by the communicative interaction of minds over many preceding generations. We need to consider a box on the line from grammar to output. The actual linguistic output of an individual is not just a random sample of the well-formed sentences generated by his internalized grammar, subject to degeneration by performance factors. Linguistic output both fails to reflect the full set of possibilities defined by grammatical competence and pushes beyond the limits set by competence. And this lack of fit between output and grammar can be attributed to factors in the arena of language use. Analytically true sentences, for example, semantically the most impeccable of sentences as defined by the internalized semantic rules, are (for that very reason) seldom uttered. Sentences whose topic-comment structure would only be appropriate in the most improbable situations of use, such as *America is round John*, are also rarely uttered. And we push over the limits of our grammatical competence when we need to, for rhetorical and humourous effect in word-play.

Chomsky's diagram (1.3.2) should be augmented by the addition of a connection back from an individual's internalized competence to his linguistic output, a connection mediated by the arena of use, as in (1.3.3). Diagram (1.3.3) shows a diachronic spiral and is, in fact, just Andersen's diagram (1.3.1) seen from another angle.

1.3.3



The arena is a rough, knockabout place. There can be winners and losers. A child may conceive of a meaning that is perfectly coherent and express it in a way that his internalized competence tells him is well formed, only to find that nobody understands him, either because his sentence is too hard to process, or because his meaning is so wildly original that no-one is prepared to take it on board. The child either learns to be more circumspect and find more acceptable ways to get his meaning across or he shuts up. In either case, his potential output is shaped by this experience. Less drastically, and more typically, a speaker learns the most successful ways of expressing his meanings, and the statistical shape of his output is thus influenced by his experience.

It is of course conceivable that the LAD is so rich that it makes full allowance for the effect of the arena of use on linguistic output. That is, the L.A.D. might be able to compensate fully for the 'distorting' factors affecting output and be able to retrieve a more or less perfect replica of the competence(s) involved in producing the output. But this strikes me as very implausible. Although the LAD may well be able to generalize in impressive ways beyond experience, it would be surprising if there were not other ways in which it simply interpreted its linguistic experience as determining evidence for particular quite specific rules.

If, in a particular culture, inanimate objects are very seldom referred to by grammatical subjects, for reasons having to do with presuppositions about the world common to the culture, it would seem plausible that a child could internalize rules whose effect was to prohibit the appearance of inanimate noun phrases in subject positions. 'Grammaticalization' of an originally non-grammatical fact would have taken place. I believe that such effects have been extremely prevalent in the shaping of the

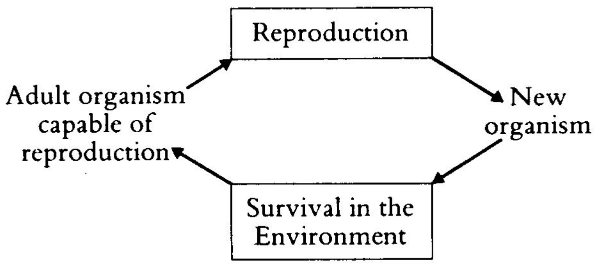
languages we see around us today. In Chapter 6, I shall present a model, supported by computer simulations of the interactions of individuals in a community, of how a particular very salient fact about numeral systems can arise through the elimination of pragmatically less-preferred expression types in the arena of use.

The arena of use is where linguistic inventions are tried and tested. If an individual conceives of some new way of expressing a meaning (which may itself in some sense be novel), he may manage to get others to adopt this new locution. It will spread according to its usefulness, and provide the basis for the internalization of rules by succeeding generations. A large part of the argument of this book is to the effect that numeral systems have evolved by successive small increments of linguistic invention. The successive inventions are built somewhat roughly on the pre-existing structures, so that growth marks can be seen in the resulting developed systems. And languages, like living organisms, can have vestigial characteristics.

Jakobson wrote 'two mutually opposed but simultaneous driving forces that control every linguistic event, which the great Genevan scholar characterizes as the 'particularist spirit', on the one hand, and the 'unifying force', on the other (1968, p. 16). As diagram (1.3.3) shows, primary linguistic data (that is 'every linguistic event') originates in individual competence (which with other personal traits provides the 'particularist spirit') and is filtered through the arena of use (which provides the 'unifying force'). Jakobson's work is widely regarded as a precursor of Chomsky's doctrine that the principal significant determinant of universal patterns in language is the innate apparatus of the language-acquiring child, which disposes it to prefer certain patterns (sounds, and so on) over others. But it is worth noting that Jakobson believed that sound changes could also originate by adults modifying the outputs of their own internalized grammars (to put it in modern terms) when interacting in the arena of use with a child. 'The phonological changes in language which arise from children are realized either by the adaptation of the older generation to the language of the child, or by the permanent reluctance of children, *i.e.*, the new generation, to accept a certain component of their linguistic inheritance' (1968, p. 18). We can reasonably see this as attributing sound change to the resolution of a tension in the arena of use. There is some kind of (presumably covert) negotiation as to whose form is adopted: the child's or the adult's. Jakobson devotes several pages to discussion of cases in which adults adopt childish forms, presumably with the purpose of being more clearly understood, or establishing solidarity with the child, or whatever. Interestingly, Jespersen (1964, pp. 178–180) also discusses similar cases in very similar terms. Evidently some kind of social force is exerting an influence on linguistic output, and in this way, both Jakobson and Jespersen hold, the form of the language concerned may be modified.

A close analogy between the evolution of languages and that of species can be shown by comparing the cycle in diagram (1.3.3) with that in (1.3.4).

1.3.4



(Thinking in terms of asexual reproduction might make this analogy easier to grasp, though this is not crucial.) Both reproduction and survival act as filters on the evolution of species. Chomsky has drawn attention to the LAD as a filter on possible languages. I wish to point out that the arena of use is also a filter. Thus, evolved languages occupy the intersection of sets of systems which can pass through both these filters. The notion of a language as an intersection is brought out in a further diagram (1.3.6) towards the end of this section.

At least some of Chomsky's view that languages are epiphenomena, and that the very notion of a language may be incoherent may well be attributable to an unwillingness to accept abstract realities of a certain sort. Chomsky may be able to conclude so readily that 'There is nothing in the real world corresponding to language' because of an ontological prejudice about the real world, reflected in a quotation such as 'If you are talking about language you are always talking about an epiphenomenon, you are talking about something at a further level of abstraction removed from actual physical mechanisms' (Chomsky, 1982, p. 108). An allusion

to 'the more significant reality of UG' (Chomsky, 1981, p. 8) seems to imply a problematic view that different types of reality may be classified, *a priori*, as more or less significant than others. Perhaps this difficulty can be circumvented by pointing out that the study of languages, as opposed to grammars, is merely a study at a higher level of abstraction. Just as it is methodologically very fruitful for the mentalist paradigm to abstract away from physical, neurological facts, it is possible to abstract away from considerations of individual competence and consider the (in some sense more abstract) entities known as languages, which are shaped by both mental and social forces.

Languages are clearly in a sense more abstract realities than individual competences; if one were to put it in Popperian terms, one would say that a language belongs to World 3, whereas an individual's knowledge of it belongs to World 2. Nevertheless, there is a sense in which languages are more amenable to collective investigation by a community of scholars than individual competences. Say one speaker has an internally consistent but quite idiosyncratic set of intuitions about grammaticality and the structural relations between sentences and phrases. In principle, it would be possible to study this speaker's internalized grammar. But to the extent that this speaker's grammar was idiosyncratic, no linguist interested in general (non-pathological) principles would pursue the case. Scientific results should be replicable. It is not practically possible, let alone desirable, for a community of researchers to focus on the grammatical competence of a single chosen individual. Researchers in fact pursue questions on which there is a fair degree of coincidence between the intuitions of different individuals, often the researchers themselves. It is reasonable to construe this research activity primarily as research into a language.

In Chomsky's most recent discussion of these issues (1986), he makes a useful distinction between an I(nternalized)-language, the language generated by the internalized grammar of an individual, and an E(xternalized)-language. There is some variability in Chomsky's characterization of an E-language. He introduces the notion, citing Lewis (1975), as 'a pairing of sentences and meanings ... over an infinite range, where the language is "used by a population" when certain regularities "in action or belief" hold among the population with reference to the language, sustained by an interest in communication' (1986, p. 19). Later, Chomsky glosses E-language as 'behavior and its products'

(p. 28); clearly, a pairing of sentences and meanings over an infinite range is not the same as behaviour and its products. Chomsky presents the distinction between E-language and I-language as an exclusive dichotomy; there has been a shift of focus, he says, from study of E-language to study of I-language. Jespersen is depicted as in some sense foreshadowing the shift of focus to I-language, a shift coinciding with the inception of generative grammar. Certainly, in Chomsky's work a mass of literature emphasizing the study of I-language has appeared. But I do not believe that the issue was at all clear-cut before Chomsky or that it is clear-cut in the practice of many working linguists, including generativists, since the beginnings of generative grammar. The distinction between I-language and E-language is useful in that it invites linguists to ask what is the actual object of their study. But a legitimate answer to this question can be 'either', or 'both'. This may account for the perplexing alternation between the 'psychological' and 'social' nature of *langue* in Saussure's *Cours*. Despite Chomsky's approving allusions to Jespersen as having I-language in mind, there is solid evidence that Jespersen also saw a language as something other than the internalized property of an individual; witness the following passage, with which Jespersen chose to introduce *The Philosophy of Grammar*.

> The essence of language is human activity – activity on the part of one individual to make himself understood by another, and activity on the part of that other to understand what was in the mind of the first. These two individuals, the producer and the recipient of language ... *and their relations to one another*, should never be lost sight of if we want to understand the nature of language and that part of language which is dealt with in grammar. (1965, p. 17, emphasis added)

Linguists in the past have not distinguished I-language from E-language. They have taken a language in a broad sense to include both kinds of matter in the spiral in diagram (1.3.3), that is both (mental representations of) a language system and language behaviour. This was, for example, how Sapir studied language(s). Naturally, specialisms and preferences arise, amd some concentrate on the system, others on the behaviour. But the studies are complementary, and mutually illuminate each other. A move to isolate one from the other is detrimental to the whole.

If one interprets an E-language not as 'behaviour and its products' but as some kind of abstract non-psychological Platonic object, such as Katz (1981) takes language to be, then I, like Chomsky, see very little point in a study devoted to it. Such a study, it seems to me, would quickly degenerate into insoluble dogmatic essentialist quibbles about whether such-and-such a property was to be considered, in the abstract and independently of empirical engagement, 'essential' to (a) language. The view of languages as abstract objects (for example, as sets of sentences) as it appeared in some early generative studies, arose, as Chomsky notes (1986, pp. 29–39), from the influence of the study of formal systems. Clearly there are important differences between natural languages and formal languages, captured well in a discussion by Moravcsik:

> (S1) A natural language is primarily a spoken language, to be used for person-to-person communication.
> (S2) A natural language is a biological phenomenon; its structure is constrained by biological mechanisms, e.g. the acquisition device.
> (S3) A natural language is a historical phenomenon; it is spread out in space and time, and is subject to change and development.
>
> (1983, p. 234)

Chomsky's exclusive focus on I-language neglects Moravcsik's aspects S1 and S3. I take as the object of study an entity which is both a biological and a historical phenomenon. Chomsky's unwillingness to take as the object of study something which is partly shaped by social interactions in the arena of use may stem from a view that the determinants of human actions present a 'mystery', as opposed to a 'problem'. In Chomsky's thought, mysteries are domains which, he believes (1976), are probably permanently beyond the reach of scientific analysis. Nevertheless, though we can rarely predict specific human actions, it is possible to characterize norms of, and regularities in, human behaviour, and to spell out a connection between these and the regularities characteristic of the arena of use.

When one concentrates on subparts of a language, say its tense system, its system for expressing anaphoric relations, its case system, or its numeral system, variability in the data decreases correspondingly. There probably are no two individuals out of several billion in the world whose entire linguistic system is the same in all details. But there are millions of English speakers of whom it can reasonably be said that they share the same numeral system (or case system, or anaphora system, or tense system, and so on). Linguistic subsystems such as these can be dissected out of the whole. The hallowed dictum that 'In a language everything is linked together with everything else' (Jespersen, 1909, p. v) does not mean that a language is a homogeneous mass in which subparts with clearly discernible functioning cannot be distinguished. And the simple fact of variability does not in itself undermine the claim that clearly discernible systems exist. Most French speakers use either *quatre vingts* or *octante* exclusively, but if there are some who use both forms, depending, no doubt, on social and contextual conditions, then one is not bound to conclude that for such speakers there is no clear-cut system. The most natural claim would be that there are in fact two clear systems, a wholly decimal one, and a partly vigesimal one, and these speakers make use, depending on the circumstances, of both systems.

None of the above should be taken to imply that the theory of the relation between languages, dialects, and idiolects, and their relation to society, is of no interest. Quite the contrary, in fact. Socio-linguists who concentrate on these relationships may well point out that the clearly describable, roughly countable, nameable objects I have characterized above as languages are in fact just the *standardized* languages. If allowed a quite broad definition of 'standardized', I accept this point. Standardization of languages is, however, a very widespread phenomenon, and appears to be something that happens naturally in linguistic communities. Clearly, there are social forces both favouring and opposed to standardization – labelled the forces of 'individualism' and 'conformity' by Hudson (1980, p. 14). But standardization occurs widely, resulting in the salient objects known as English, French, and so on, which the linguist can take as primary starting points for the study of the structure of language.

It happens that numeral systems in particular are subject to more drastic and rigid standardization than other subsystems of languages, for reasons having to do with the nature of numbers (numeral meanings). In Chapter 6 the kind of standardization found in numeral systems is explored in detail and a model of such standardization in terms of social linguistic exchanges between individuals in a community is proposed and investigated

with some exactness. Thus, in this case at least, it is not a matter of the theoretical postulation of well-defined (sub)systems made possible by abstracting and idealizing *away* from questions of the use of language in society. It is, rather, the (relatively) pretheoretical recognition of the obvious fact that well-defined linguistic systems exist, followed up by an attempt to explain the structure of such systems in terms of the most appropriate sorts of determinants, which may be, variously, social, psychological, or even mathematical. In short the study of language use in society, and linguistic variability (or the striking lack of it, as in cases of standardization) should not be seen as antithetical to any study which isolates and describes systems whose outlines and structural features are taken from the start to be fairly clear. Such structural studies can actually illuminate and clarify some of the socio-linguistic questions (much as the study of language acquisition can be illuminated by independent characterization of the object acquired, the language).

Languages, and especially some of their subsystems, are, then, objects with structure clearly definable up to certain ample limits, beyond which there is admitted fuzziness, and can be taken as independently given. Now, having emphasized what can reasonably be taken as given, I wish to draw attention to a type of judgement about aspects of linguistic systems which I believe never to be pretheoretically obvious, but always to be the outcome of a certain amount of theorizing. Within a language, linguists often claim that certain structures or subsystems are in some sense marked or odd. But such oddness can be of various types, perceived differently by different observers, and its source is often obscure. For example, the irregular morphology of frequently used forms, such as English *am*, *is*, *are*, *were*, *was*, *been* does not strike adult, linguistically naive speakers as in any way odd, whereas to the linguist they are marked by virtue of their deviation from otherwise regular patterns. On the other hand, multiply centre-embedded sentences, which the linguist might want to insist are in some sense perfectly well-formed, are very generally held by linguistically untrained native speakers to be 'not part of the language'.

Languages, the totalities described by descriptive grammarians, are as unyielding to linguistic theorists as the physical world is to physical theorists. Many aspects of them can be wrapped up in neat generalizations, projected from the interaction of parsimonious principles, and the like, but there always remain further puzzles, bits that won't fit nicely with any of the going theories. This is a truism. A reaction to this situation is the invocation of the distinction between core grammar and the periphery of grammar (Chomsky, 1981, pp. 7–8; 1982, pp. 108ff). This move is linked in an essential way to the psychological realist interpretation of linguistic theory as a theory of the innate human apparatus which makes language acquisition possible. Aspects of languages which are picked up despite their apparent complexity, and despite apparently insufficient exposure to relevant data on the part of the child, are to be handled under core grammar, while those aspects which a child requires some rehearsal to learn belong to the periphery. In a theory of language acquisition, it is reasonable to make such a distinction, but it is important to remember that this particular core/periphery distinction emerges from a view of the study of language as nothing but the study of the basis of language acquisition.

In the adult language of a community, which I take to be the type of primary object of linguistic study, one does not expect to find an obvious reflection of any specific core/periphery line of demarcation derived from considerations of language acquisition. Mature speakers of a language command the whole language fluently, marked or unmarked constructions alike. While different parts of the language may have been acquired at different rates, the fact is that they have all been acquired. Just as a language can, given time, assimilate a foreign vocabulary so thoroughly that native speakers have no intuitions about the different provenance of words, so an adult's language contains both marked and unmarked constructions without any visible seams to show the different bases of their acquisition.

The view that the only significant determinants of the structure found in languages are biological is a hypothesis. It could conceivably be falsified by demonstrating that some property or properties of languages can plausibly be attributed to some non-biological determinant. The issue is wide open. No convincing, closely reasoned, substantial case has ever been made for the attribution of any specific structural property of language to any demonstrably biological cause, nor has any such case been made for any demonstrably non-biological cause. Chomsky has eloquently and vociferously championed the idea that the study of linguistic structure is to be interpreted as the study of a part of Man's genetic endowment. It seems reasonable to expect that he is partially right. A belief in the pure *tabula rasa* is mystical.

Recent writers on the explanation of linguistic universals have tended, sensibly, to accept the principle that biological factors play a part, perhaps even a large part, in determining the structure of languages. But many writers have independently expressed the view that it must be reasonable to investigate the possibility of a range of determining factors other than 'innate peculiarities of the grammar representation centers of the human brain' (Fodor, 1984, p. 9). Contributors to this swell of opinion include J. D. Fodor (1984), Hyman (1984), Comrie (1984), Lindblom et al. (1984), and Aitchison (in press).

Chomsky's consistent playing down of the possibility of non-biological determinants of linguistic structure is quite remarkable. Although it cannot be taken as an argument if I say that I find this single-minded concentration on one source of linguistic structure to be quirky, that is indeed how this stance inevitably strikes me, in whichever of Chomsky's publications it reappears. The following passages are typical:

> By a 'true universal' we mean a principle that holds as a matter of biological necessity and therefore belongs to UG, as contrasted with a principle that holds generally as a matter of historical accident in attested languages. (Chomsky and Lasnik, 1977, p. 437n.)

> The theory of particular and universal grammar, so far as I can see, can be sensibly regarded only [*sic*] as that aspect of theoretical psychology that is primarily concerned with the genetically determined program that specifies the range of possible grammars and the particular realizations of this schematism that arise under given conditions. (Chomsky, 1980, p. 202)

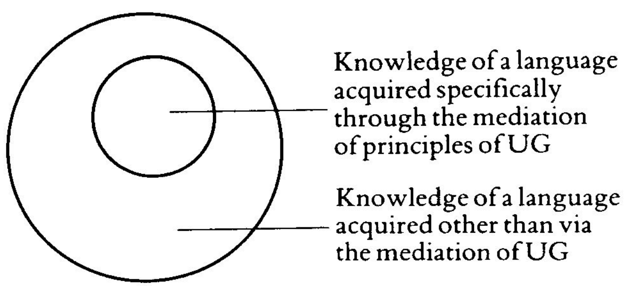
> what a particular person has inside his head [i.e. an idiolect] is an artifact resulting from the interplay of many idiosyncratic factors, as contrasted with the more significant reality of UG (an element of shared biological endowment) and core grammar (one of the systems derived by fixing the parameters of UG in one of the permitted ways). (Chomsky, 1981, p. 8)

> each actual 'language' will incorporate a periphery of borrowings, historical residues, inventions, and so on, which we can hardly expect to – and indeed would not want to – incorporate within a principled theory of UG. (Chomsky, 1981, p. 8)

The language of the first quotation above implies triviality in the non-biologically determined universals. The implied straightforward dichotomy between biological necessity and historical accident is overly simple. It would be counter to all good statistical practice if one assumed that some contingent property of a large number of attested systems was merely accidental, not significant. A sensible strategy would be to look for the necessity in such an impressive set of 'accidents'. The rhetorical deployment of 'accidents' (pejorative) and 'necessity' (ameliorative) in this quotation is a trick. Biological necessity itself results from accidental mutations in the evolution of the species. The argument in the last quotation is valid if one accepts the premise that a general theory of linguistic structure (UG) is to be identified *only* with a theory of the innate LAD. But this premise is not at all obvious. The dismissive relegation of 'historical residues, inventions' is not based on any reasoned argument that these factors are intrinsically unsusceptible to incorporation within a principled theory of some sort. I shall argue in later chapters that certain recurrent structural features of numeral systems are indeed to be explained as historical residues, much as geological structure is obviously to be explained as 'historical residue'. Geological theory is not unprincipled. Invention, also dismissed in the quotation above, will also play a part in the account I will give of certain recurrent aspects of the structure of numerals. Certainly, a language is an artefact resulting from the interplay of many factors, but it simply cannot be assumed that these factors are 'idiosyncratic'. A legitimate aim for linguistic theory is to characterize the several factors contributing to the form of language, showing how, because of their nature, the interplay between them must yield the structures of the type we find in languages.

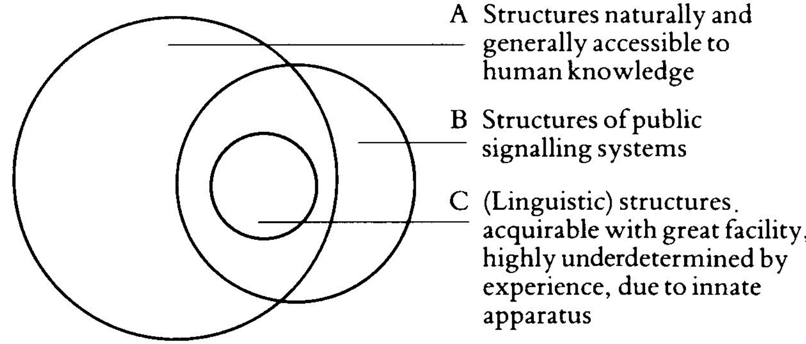
A couple of simple diagrams may make my general point clearer.

1.3.5



(1.3.5) represents what I take to be a view close, if not identical, to Chomsky's. The concern is with knowledge of a language and the innate psychological principles giving rise to this knowledge. My doubt about whether this picture is exactly what Chomsky has in mind lies in the fact that he may or may not wish to claim that knowledge of vocabulary, idioms, irregular forms, and so on, is mediated by UG. I would guess that the learning of vocabulary can be attributed to some kind of associative memory capacity not specific to the language faculty. After all, experimental chimpanzees, such as Washoe and Sarah, could manage to learn vocabulary. (1.3.5) can be compared with diagram (1.3.6), in which, for the sake of a plausible example, it is assumed that the fact that human languages are public signalling systems is also (that is in addition to the innate LAD) one of the factors determining the type of structure found in languages.

1.3.6



Here the set A minus B is the set of structures naturally and generally accessible to human knowledge other than structures of public signalling systems (for example, perhaps certain musical structures or the structures of various games). B minus A is the set of structures of public signalling systems not naturally or generally accessible to human knowledge (for example, the analog structures in the dance signalling systems of bees). The intersection of A and B (call it L) is the set of structures of public signalling systems naturally and generally accessible to human knowledge (that is, structures of languages, e.g. French, Hindi, Xhosa, Ameslan). L minus C is the set of linguistic structures acquirable by memorization, explicit instruction, and so on. Clearly, any such diagram drastically oversimplifies the real case, although I hope it will aid appreciation of an alternative view to Chomsky's. A number of comments need to be made.

There is an admitted vagueness in the idea of 'structure' made use of in diagram (1.3.6). Structure is an idea probably far better demonstrated by example than by any attempt at an encapsulating definition. People who talk of structures often seem to agree on what they are talking about. And 'structure' is hardly less vague than 'knowledge'.

The intersection of the two large circles, A and B in (1.3.6) represents human languages, in the sense described above. This area is the intersection of two larger, partially overlapping areas, to be defined by different theories, respectively a theory of human cognition, and a theory of signalling systems. For simplicity, only two possible domains involved in determining linguistic structure are mentioned in diagram (1.3.6), but it is to be expected that at least several other domains are involved, for example general factors affecting diachronic language drift, and general factors involved in the social pressures on language, such as pressures to conformity and standardization. This view of language is in no way new. In some ways, *mutatis mutandis* to allow for detailed insights and perspectives gained in the intervening seventy years of scientific attention to language, the view is that of Saussure's *Cours*.

This mode of accounting for linguistic structure is 'pluralistic' and (unashamedly) eclectic. Eclecticism deserves a bad name where it represents an unprincipled mingling of ideas from various domains as a means of placating adherents of as many sets of ideas as possible. An example of theologically bad eclecticism would be a religion which postulated a divine triumvirate of Jesus, Buddha, and Mao Tse Tung (an 'unholy alliance'). But the eclecticism proposed here in accounting for the nature of language is not proposed for the purpose of accommodating a maximum number of extant views on the nature of language in a single framework, regardless of the intrinsic merits of those views. Rather, it is an acknowledgement of a deep-seated eclecticism in language itself. A language is an artefact resulting from the interplay of many factors.

As soon as one begins to theorize in any domain, some items in the data tend to sort themselves out as being atypical of the general patterns found in the domain. Perhaps, from the point of view of the theory of signalling systems envisaged in diagram (1.3.6), human languages themselves are in some sense atypical of the domain of signalling systems as a whole. I do not know whether this is the case, but it is a possibility. If it is the case, then from the point of view of such a theory, a quite different demarcation between 'core' and 'periphery' would be appropriate. The point is that what emerges as 'core' or 'periphery' is determined by what one takes to be the domain of one's theory. 'The distinction [between core and periphery] is in part theory-internal' (Chomsky, 1982, p. 108).

The lack of interest shown to date by linguists in numeral systems may be due to a judgement that they do not belong to core grammar, in Chomsky's sense. I will argue that the 'interplay of many factors' account is appropriate, not only to whole languages, but also to their subparts. Although numeral systems present themselves, pretheoretically, as fairly clearly defined whole systems, different aspects of them are attributable to different factors – psychological, social, historical, and mathematical. At least one aspect of the structure of numeral systems is apparently attributable to innate properties of the LAD, so that it would fall within the Chomskyan 'core', but for other aspects, explanations which are neither directly psychological nor directly biological will be given. Mere promissory notes in an introductory chapter such as this do not in themselves make the case, but if the case is to be made, it has to be accepted at the outset that non-biological accounts of some universal aspects of linguistic structure are possible in principle. The goods will be delivered. But the scene-setting for the argument requires that one consider languages without any prior conceptions of what parts of them are of central interest ('core') and what parts to be relegated to some 'periphery'.

2

Explaining Linguistic Universals

2.1 Chomsky's Account of Universals and the Strategy of Generative Grammar

When a child acquires his native language, Chomsky's account goes, he internalizes a grammar of it. The predictive power of the internalized grammar extends far beyond the primary linguistic data to which the child is exposed during the critical period. Thus a child *generalizes* from observed data used by older speakers around him to a complete range of further examples which he intuits to be correct, even though he has never actually observed them. Many different kinds of generalization from observed data are logically conceivable; but only certain types of generalization, so the Chomskyan account goes, are in fact made by children acquiring their first language. That is, the child is disposed (innately) to extrapolating from observed data in only a subset of the logically possible ways. As an uncontroversial (and therefore trivial) example, take adult-to-child utterances such as the following:

2.1.1 If you're very good, I'll buy you an ice cream.
That was very very naughty.
Don't touch that – it's very very very hot!

Assume that the child has internalized a connection between the category 'Adjective' and the words *hot*, *naughty*, and *good*, and that she also knows that *very good*, *very very naughty*, and *very, very, very hot*, are all phrases of the same type, call it 'Adjp'. How does the child generalize from the three observed Adjps in (2.1.1) to a general rule for the formation of Adjps containing

very. Some of the possible rules compatible with the observed examples are given in (2.1.2).

2.1.2 (a) ADJP → (very) (very) (very) Adjective
(b) ADJP → (((very) very) very) Adjective
(c) $\text{ADJP} \rightarrow \left\{ \begin{array}{l} \text{Adjective} \\ \text{very ADJP} \end{array} \right\}$
(d) ADJP → (INTP) Adj
INTP → (INTP) very

We know that many children quickly learn that it is possible to repeat *very* an indefinite number of times – children often amuse themselves by keeping us waiting for the end of the sentence while they string out an enormous number of *verys*. The Chomskyan account assumes that all children learning the same language internalize the same rules, so children learning English cannot have internalized rule (2.1.2a) or (2.1.2b), which are equivalent in only allowing up to three *verys*. They appear to have a preference for a rule or rules with the recursive property of (2.1.2c) or (2.1.2d). ((2.1.2.d) is actually preferable on other grounds to (2.1.2c).) No child ever hears an infinite number of *verys* strung together – how could he? But children internalize a rule or rules allowing unlimited repetition of *very*. The assumed uniformity of such a preference in all English learners is attributed to innate mental properties of *Homo sapiens*.

Comparing the limited corpus of observed utterances in any language with the infinite set of intuitively well-formed sentences gives insight into the extent and type of generalization made by a person exposed to a limited corpus in acquiring his intuitive knowledge of what sentences are well formed in his language. In practice, generativists do not *compare* a native speaker's intuitions with the observed data which formed a basis for the acquisition of those intuitions. The methodological assumption is made that the observed data is so degenerate, and the acquired knowledge of the language so rich, that insight into the language learner's innate preferences for generalizations of a certain sort is overwhelmingly most likely to come from a study of the system of acquired intuitive knowledge alone. This amounts to a methodological judgement that the structure of the acquired system (or at least the theoretically interesting subset of it) is in large part due to the learner's innate language acquisition apparatus, and only in lesser part due to any organization or structure directly discernible in the data to which he is exposed. Thus the most theoretically interesting similarities between one speaker's knowledge of his language and that of his parent come mostly from the fact that both speakers brought the same, genetically determined, apparatus to bear on their experience and less from any similarities in their actual experiences. Indeed Chomskyans would presumably claim, possibly correctly, that two speakers could acquire the same internalized system on the basis of completely non-overlapping sets of observed sentences.

All humans are assumed to inherit an identical language-acquisition device (LAD). So all will acquire, according to the Chomskyan account, systems of intuitive knowledge with certain common structural properties, namely those due to the common mental inheritance, universal grammar (UG). The properties common to the systems of intuitive (linguistic) knowledge of all speakers of all languages are linguistic universals. Chomsky emphasizes a distinction between UG and linguistic universals. 'UG has never been thought of (within TGG) as a theory about what is universal to all languages, rather as the system that mediates between data and descriptively adequate grammars'. (Chomsky, personal communication, quoted in Winston, 1982, p. 85). But it is necessary to recognize, not only the difference which Chomsky emphasizes between UG and linguistic universals, but also the significant overlap, in view of the great and direct explanatory potential which UG has for linguistic universals. Chomsky acknowledges the importance of Greenberg's universals for his own study (Chomsky, 1982, p. 95, 111). The existence of linguistic universals can be explained by postulating in each new-born child a disposition to extrapolate in the same way from his first linguistic experiences.

I have tried above to make an objective and straightforward statement of the Chomskyan account of universals, distilled from such well-known publications as Chomsky (1965, Chapter 1, 1968, 1980a). Stated thus, the account, though making certain assumptions and claims that one may legitimately decline to accept, is internally coherent and not tautological. That is, one may find the case arguable and bring various kinds of evidence to bear upon it. But it is important to note the emphasis of Chomsky's work. The interest for Chomsky is not merely, in fact not even primarily, in finding explanations for properties of language which happen to be universal, whether such explanations are biological or not. Chomsky's driving motive is to find mental (ultimately physical) properties innate in humans, and linguistic universals are valid as evidence in this search. Chomsky's approach does not begin by taking some body of observed universals, such as those in Greenberg (1963a), and asking: 'What could have caused these?'. Rather, it is taken as self-evident that man is genetically endowed with rich mental structuring, and argued that these are reflected in subtle properties of language which are probably universal. Thus, to find linguistic universals for which there are plausible causes other than the innate structuring of the LAD is not necessarily to refute Chomsky's claim that such structuring is the cause of some (other) universals.

On the other hand, there is no need to accept Chomsky's implicit definition of the study of universals. Nor need we accept that the UG research programme has a monopoly over such terms as 'explanatory adequacy', 'linguistic theory', and 'psychological theory', as is implied in the following: 'While there may indeed be links between rules of grammar ... and perceptual strategies, and even functional explanations for these rules, the matter does not seem to bear on explanatory adequacy in the sense relevant for linguistic or psychological theory' (Chomsky and Lasnik, 1977, p. 438). To start with a body of universals and look for explanations, of whatever kind, mental, social, functional/evolutionary, historical (for example, monogenesis), is an intellectually legitimate exercise, appropriately labelled 'linguistic theory' and naturally concerned for the adequacy of its explanations. This is the approach of this book, which takes universal properties of numeral systems and seeks adequate answers to the question 'What could have caused these?'.

A point needs to be made about the relationship between the quest for innate properties of human mental organization and generative grammar. There are many substantive properties of language which were familiar before the development of generative grammar, for example phonological features, grammatical categories such as noun and verb, and components of meaning such as causation and concreteness. In some of these cases, at least, an innateness account is not controversial, especially in those areas where physical, as opposed to mental, properties are involved. So, for example, it is not contested that the universality of certain phonological features is due to the universal inheritance of physical organs of a certain shape. Similarly, there is a completely uncontroversial explanation, centrally involving an innate property of human beings, for a very striking universal tendency for natural language numeral systems to use a base of 10. To my knowledge, no contemporary linguist has ever thought it necessary to spell this explanation out, let alone argue against it. Facts and explanations as obvious as this have been beneath the concern of generative linguists.

Generative grammar has concentrated more on the study of form than on that of substance, and hence it is in the area of the form of language that generative grammar has made its largest contribution. To gloss a view expressed by Chomsky (1965, p. 30), the originality of generative greammar is in the construction of complex sets of interacting rules; formal properties of language are now talked of which could not have been imagined before the development of generative grammar, for example 'structure-preservingness', 'strict cyclicity', 'upward boundedness', and so on. The interest in such properties is in their possible universality, and generative grammar is seen as a new and powerful heuristic for probing such formal properties. Perhaps because of their very newness, no obvious explanations spring to mind for the existence of these universals, and the innateness theory fills the gap.

The link with the theory of innate ideas adds an incentive to the search for formal universals, and generative grammar provides a heuristic and a notation. The search for innate properties of the mind via a search for formal linguistic universals is not constrained in advance by any *a priori* considerations of what a mind may be like (apart from the requirement that it be finite). If a linguist (or psychologist) suggests a physical explanans for some universal of human language (or behaviour), it is expected to be possible in principle to demonstrate that humans do indeed possess the physical property appealed to. Thus if humans had the oral tracts of sparrows, it would be literally inexplicable (physically) why human universal phonological features are what they are. Searches for physical explanations must concern themselves with both ends of a question at once, with the existence of an explanans with the appropriate physical properties, and with a delineation of the relevant physical properties of the explanandum. In the absence of any *a priori* theory of what a mind (the assumed explanans of generative grammarians) may be like, generativists have perforce to concentrate their attentions on careful delineation of the formal properties of the explananda, namely the formal linguistic universals. A clear and characteristic research strategy for this task has evolved.

The strategy involves treating the sentences (clauses, phrases) of a language as an unordered set of strings of symbols. To each such string, one or more structural descriptions is deemed appropriate, associating with the string various properties at various linguistic levels. For example, at the surface syntactic level a string may have the property of being bracketed and having its constituents labelled in such and such a way, and at the semantic level it may have the property of bearing such and such an interpretation, and so on. The researcher attempts to devise sets of rules that generate sets of such structural descriptions. And he is enjoined, not simply to be satisfied with any set of rules that will generate the desired set, but to insist on discovering a set that captures all significant generalizations apparent in the data considered. I will not go into the question here of what constitutes a significant linguistic generalization (for some ideas on the subject, see Hurford, 1977, 1980 and references cited there). It is enough to point out that a generative grammarian believes that he knows a significant generalization when he sees one. (It should be clear that the generative research strategy that I describe in these terms is not a simple set of operational procedures, but relies on such undefined and probably undefinable elements as insight and intuition; this in no sense invalidates the strategy, of course.)

The pursuit of the significant generalization has led to the writing of grammars of considerable abstractness and complexity. The abstractness and complication are not goals in themselves, but neither are they necessarily any embarrassment. There is an assumption that if abstractness and complexity happen to lie at the end of the trail of the significant generalization, then the particular kinds of abstractness and complexity arrived at are real objects of discovery, in no sense artificial creations of the grammarian's procedures. If these objects are discovered in the grammars of a significant number of languages (or indeed even if not, for this step in the argument is sometimes passed over), it is concluded that they are formal universals, and the grammarian has defined his explananda. The explanans, the Mind with its genetically transmitted properties, he has had ready in his pocket all along.

The picture drawn above of the generative grammarian hunting the formal universal with his explanans for it already in his pocket is for many something of a caricature. Of Chomsky, it is no caricature, since he has pioneered the research strategy based on the capturing of significant generalizations and explicitly associated the formal universal results with innate properties of the mind. But it must be said that for most practising grammarians the real pleasure is in the hunt for the significant generalization, and they seldom bother at the end of their expositions to whip the Mind out of their pocket for the explanatory *coup de grâce.* Dropping the metaphor, few practising generative grammarians philosophize or psychologize about their work, but rather concentrate on the construction of grammars and the capturing of significant generalizations in them. 'I think a linguist can do perfectly good work in generative grammar without ever caring about questions of physical realism or what his work has to do with the structure of the mind' (Chomsky, 1982, p. 31). The question of the adequacy of an innateness explanation of formal universals, once discovered by the grammarian's research strategy, is usually left to the philosophers and psychologists. But there is a potential weakness in this division of labour. If a generative grammarian is not interested in the ontology of his constructs, he may develop types of constructs which are inappropriate to the mental interpretation which a psychological realist can put on them. Chomsky recognizes this possibility:

> Suppose we think of a linguistic theory, just like any theory, as a set of concepts and a set of theorems. Now, the set of concepts can be organized in all sorts of ways. The concepts have interconnections, and you want to express those interconnections as tightly as possible. The way to do that is through a constructional system in which you select a set of primitives, and a set of axioms which meet the condition that the concepts of the theory are defined in terms of the primitives and the theorems are derivable from the axioms. In principle you can pick your primitives any way you like, as long as they meet this condition. ... [But] now, if you think of linguistic theory within the framework of explanatory adequacy and language acquisition and so on, then there are other requirements. The set of primitives has to meet a condition of epistemological priority. If linguistic theory is supposed to be a model of how an idealized language acquisition system works, then the primitives have to have the property that they can be applied to the data pretheoretically. (1982, p. 118)

It is on grounds such as these that Chomsky doubts the correctness of generative theories (such as Relational Grammar) which make grammatical relations primitive.

Several of the formal universals to be discussed in this book were arrived at by the research strategy based on the construction of grammars and the capturing of linguistically significant generalizations. They are developed and explained in detail in Hurford (1975) (henceforth in this chapter *LTN*) on the generative grammar of numerals in a variety of languages. The assumption was made that the numeral constructions of any language, like the sentences, noun phrases, verb phrases, adjective phrases, and so on, could be treated as an unordered set of strings of symbols and associated with structural descriptions identifying their properties at various linguistic levels. Thus, where a grammar of English noun phrases would set itself the task of generating structural descriptions of all and only the well-formed expressions such as *men, ten tall men, ten tall men with walking sticks, a man, the man, the man at the bus stop*, and so on, the part of *LTN* dealing with English numerals set about generating structural descriptions of all and only the well-formed expressions of the sort *one, nine, eleven, sixty nine, three thousand and eighty eight*, and so on. The injunction to insist on the capturing of all the significant generalizations seen in the data was obsessively followed, with the usual increase in abstraction and complexity. Several reviewers (Griffiths, 1977, p. 222; Sigurd, 1977, p. 193) remarked explicitly on the emphasis on capturing significant generalizations, and another (Epstein, 1978, p. 123) complained about the resulting 'incredible complication of the theoretical framework'. As argued in a reply to Epstein (Hurford, 1979a) the work was a paradigm example of Kuhnian normal science; following the research strategy of seeking out significant generalizations was at all stages the central imperative. And the question of explanations for the formal universals arrived at, whether from innateness or otherwise, was not discussed, in keeping with the common practice of generative grammarians.

The present work assumes the correctness in some sense of the universals discussed, the fruits of the research strategy discussed above, but takes up the issue of what could count as plausible explanations for them (and others now dealt with for the first time).

2.2 Conflicting Generalizations, Complexity, and Irregularity

The linguist seeks generalizations. At the points where languages are irregular, the search is frustrated. But languages can also frustrate the neat capture of generalizations by presenting too many of them; there can be cases of conflicting generalizations. In *LTN* a range of such cases was discussed in connection with a device postulated within a generative framework and called the 'lexical extension component' (see passages indexed under 'lexical extension component' in *LTN* for full details). The cases which prompted the use of the lexical extension component were from languages as diverse as English, French, Danish, Welsh, Yoruba, and Ainu. The cases typically involve somewhat marginal and idiosyncratic phenomena in each language, but the recurring need for some mechanism to deal with them across languages and the demonstrated possibility of a formal mechanism embodying what they have in common is noteworthy. The typical case is where a single word clearly has some internal structure indicating that it is put together by rules incompatible with other rules called for in a grammar expressing all apparent generalizations in the data.

For example, Welsh *pymtheg*, 15, clearly reflects composition from the basic forms *pump*, 5, and *deg*, 10. But *pymtheg* is a single lexical item, on a level with *pump* and *deg* themselves, and furthermore there is no independently motivated rule putting together items from the class of *pump* with items from the class of *deg* to form items of the class of *pymtheg*. In fact, on other syntactic grounds, *pymtheg* needs to be assigned to the same class as *deg* itself.

Another example of the use of the lexical extension component involves English *million*. This is analysable in one way into *m* + *illion* by analogy with *billion*, *trillion* and such jocular coinages as *jillion*, *skillion*, *zillion*, and another way into *mill* + *ion*, by analogy with *millimetre*, *millisecond*, *millipede*, *milligram*, and so on. No compromise analysis, such as *m* + *ill* + *ion* is appropriate.

The existence of conflicting generalizations is a straightforward dilemma for a research strategy advocating the capture of all generalizations within a single consistent framework. The lexical

extension component of *LTN* got around the difficulty by postulating 'incomplete' lexical entries, that is entries with specified syntactic and semantic content, but empty of phonological content. Forms generated by the (rest of the) grammar as a whole were permitted to be inserted into the slots for phonological information in such incomplete lexical entries. In a sense, the lexical extension component was an input-output device which took partly incomplete grammars as input and gave fully complete grammars as output, the 'absent' information having been generated by the 'present' rules and lexicon of the input grammar. A statement of the basic lexical extension principle (modified slightly from *LTN*) is as follows:

> *2.2.1* Given an incomplete lexical entry associating a structure G with unspecified phonological content, where the semantic interpretation of G is S; and given also a derivation which associates S with a phonetic representation P; the incomplete lexical entry is filled out with the phonetic representation P.

In other words, languages may coin new words from the existing resources they possess. This is obvious and unobjectionable. The problem is that it is clearly a diachronic statement, and the lexical extension component was an attempt to capture a certain class of generalizations within an essentially synchronic framework. The lexical extension principle relates one grammar to another, extended grammar. It cannot therefore be seen as a component of UG, in Chomsky's sense, since UG is a characterization of the set of individual grammars. Put another way, UG is a function from language data to grammars; the lexical extension component is a function from grammars to other grammars.

The lexical extension component is a genuine linguistic universal, in the sense that it captures phenomena which recur strikingly often as one looks at language after language. But since it cannot be seen as a function from data to grammars, it cannot be part of an innate LAD.

The strategy of generative grammar reveals deep underlying regularities in languages, often masked by surface irregularities. Irregularities, no less than regularities, can occur with significant frequency across languages and thus merit the attention of the investigator of universals. Some of the universals of numeral systems to be discussed in this book are in fact universal irregularities. The relation between deep regularities and surface irregularities is often held to be an example of the complexity of languages. It is worth trying to sort out the relationship between 'complexity' and 'irregularity' in detail. In so far as the following discussion does not reflect common usage of these terms, it is an attempt to sharpen up a distinction commonly blurred, and to clarify the nativist argument based on the universal acquisition of complex systems.

'Regularity' (antonym: 'irregularity') is used here in a sense linked to that of 'productivity'. 'Regular' is a scalar predicate applied to parts of languages (typically constructions), and 'productive' is a predicate applied to the parts of grammars (typically individual rules) that generate them. Regular constructions are generated by productive rules and irregular constructions by unproductive rules. A construction is regular to the extent that its constituents can be replaced by large numbers of other members of the same grammatical category. By this criterion the sentence type represented by *the cat sat on the mat* is highly regular, and the idiomatic greeting *How do you do?* is rather irregular (cf. **Why does he do?*). The quintessential rule of grammar is traditionally conceived of as highly, if not completely, productive. But in fact languages contain large amounts of irregularity, so that many rules of grammar fall short of this ideal. Constructions involving closed word classes are by this definition less regular than those involving open classes.

The acquisition of very irregular constructions (for example, some idioms, stock phrases) can be accounted for by a fairly radical empiricist theory of learning by direct imitation of experience. To the extent that a construction is irregular it does not need postulation of rules in the traditional sense. Obviously languages contain large amounts of regularity, however, and the postulation of productive mental rules actually internalized by the child cannot be avoided.

A 'weak nativist hypothesis' goes no further than this. It is simply the negation of the strong radical empiricist hypothesis. All it claims is that the child is innately equipped with an ability and a disposition, in some quite unspecific sense, to invent productive rules which generate forms beyond his experience. But Chomsky's nativism is of a stronger form, and his case hinges not on the fact that children acquire impressively regular (or irregular) languages, but on the fact that they acquire control over impressively *complex* languages. Complexity (in the context of the innateness debate) is a concept largely independent of (ir)regularity.

'Complex' (antonym: 'simple') is a predicate applied to (parts of) grammars and by extension to the (parts of) languages they generate. The question of the meaning of simplicity is a notorious one in linguistic methodology and in the foundations of science generally. In the case of Chomskyan linguistics there exists the paradoxical situation in which the linguist as a theorist of particular languages is urged to discover the simplest grammar of a language, but in his role as proponent of the strong nativist cause he is called upon to present examples of extremely complex grammars. (For some refreshingly disarming comments in response to a question mentioning this paradox, see Chomsky, 1982, pp. 30–1.) Not surprisingly in this paradoxical situation linguists have tended to adopt interpretations of the terms 'simple' and 'complex' specialized for particular contexts. In the context of writing and comparing grammars, simplicity has for many become equated with 'capturing significant generalizations', and complexity must be inferred, in this context, to be a concomitant of the missing of such generalizations. Clearly, Chomsky does not argue his strong nativist case by asserting that we have evidence that children are disposed to internalize systems of rules which miss significant generalizations. So when a nativist argues that children internalize highly complex sets of rules, it is clear that we must understand 'complex' in a sense distinct from that assumed by linguists constructing, justifying, and comparing grammars. I sketch what appears to be the sense appropriate to the context of the innateness debate below.

Complexity here is a notion closely related to the depth, or interdependence, of rules in a set. To the extent that the import of a rule of grammar for the language can be seen without reference to other rules, that rule is simple. A maximally simple rule does not depend for its interpretation on other rules. (Obviously natural languages do not contain rules which are maximally simple in this sense: all grammars are to some extent interlocking.) For example, transformational rules, as in the standard model, are complex in that they operate on the output of other rules. The addition of a 'transformational level' to grammars increases their depth or complexity. One cannot tell what sentences are (not) generated with the aid of a particular transformation without also having access to other rules of the grammar. Similarly, the incorporation into a transformational grammar of the 'metarule' or 'traffic rule' of the transformational cycle makes the grammar more complex, since it increases the ways in which rules can interact to generate sentences.

The ability to internalize a deep, or complex, system of rules on the basis of limited data is obviously impressive. (Assuming, crucially and controversially, that no simpler system would be equally correct.) The Chomskyan claim is that we can get an idea, through the methodology of generative grammar, of the system of rules internalized for any given language, and that such systems are invariably complex. The more complex, or interlocking, a grammar is, the harder it is to induce its rules individually, from limited data. And it is also hard to induce the whole grammar, all at once. So some explanation of the apparent ease with which children learn languages is called for. The nativist's proposed explanation is that the child knows innately the principles by which rules of grammar interlock. He knows, for instance, whether transformational rules necessarily operate after all phrase structure and lexical rules. So this aspect of the complexity, the interlockingness, of grammars, at least, does not have to be induced, and the learning feat attributed to the child is plausibly diminished.

Under the interpretations of these terms suggested here, complexity and irregularity are clearly different. To the extent that there is a connection between them, it appears that it is an inverse connection. That is, a complex system depends on its component rules being productive to a certain extent, so that interaction between them can be exploited. If a given rule is of limited applicability, it produces fewer forms for other rules to interact with. For example, it was only because the individual cyclic transformations of the standard model were conceived of as highly productive that it was possible to take the further conceptual step of postulating a complex cyclic relationship between them.

Versions of the nativist thesis can be marshalled in order of their 'ambition', and this ordering can be related to successively more all-embracing sets of putative explananda, or evidence. Thus: (1) a quite weak version, which concerns itself with examples like *bringed* and *goed*, claims merely that children are innately equipped to acquire productive rules whose effect goes beyond their experience; (2) a stronger version argues that human grammars are complex and acquired impressively quickly, and that this indicates innate dispositions to acquire grammars of a quite specific form; finally (3) the stronger version is given extra support by the invocation of the universality across languages of such complex systems of productive rules. The crucial point to note here is that the argument from universality only goes through if the universal concerned is both complex and at least somewhat productive. The universality of some *simple* property of linguistic systems would not support the strong version of the nativist thesis. And no unproductive process, giving rise to irregular phenomena, can give support to linguistic nativism, either. Hence the mere universality of some formal characteristic is not in itself an argument for some specific innate linguistic structure.

We are all given to marvelling at the 'complexity' of languages, impressed by the burden they impose on the adult mind that tries to grapple with them. Of the burden there can be no doubt, but whether it mainly consists in complexity or in irregularity is a question that must be seriously considered. The two are not the same and only the former can be cited as evidence for innate, highly specific, principles of linguistic organization.

A great many characteristic properties of numeral systems across languages are in fact *ir*regularities. The same patterns of irregularity repeat themselves from langage to language. I will describe these informally in the next section, showing why, from the point of view of a generative grammar, they must be seen as irregularities. A historical/evolutionary explanation for the existence of these universal patterns of irregularity will be approached in the rest of this chapter and developed in later chapters. This explanation sees such irregularities as reflecting 'growth marks' left during the evolution of the developed systems.

2.3 Some Universal Irregularities

Irregularity is perceived in relation to an ideal of regularity. Perhaps the most ideally regular way of expressing numbers would use two primitive terms, 1 (or zero) and 'successor'. Abbreviating 'successor' to S, expressions for the first few numbers in such a system would be as in the middle column in (2.3.1); using a single polysemous symbol for both the basic number 1 and the successor function, the first few numbers might also be expressed as in the right-hand column.

2.3.1	1	1	1
	2	S1	1 1
	3	SS1	1 1 1
	4	SSS1	1 1 1 1

The disadvantages of such a system for spoken communication are obvious. Numbers of any size would require long cumbersome expressions placing an impossible burden on short-term memory. But one might expect to find some use of a system like this for the first few numbers. In fact, however, no (spoken) natural language numeral system even begins in this way: no language expresses 2 as anything suggesting 'successor of 1' or '1 1', and 3 as anything suggesting 'successor of successor of 1' or '1 1 1'. Some written systems work this way for the first few numbers, but, as I will argue in the next chapter, there is a sense in which such written systems are not fully linguistic.

Alternatively, a binary notation, as in (2.3.2) might be considered ideally regular.

2.3.2	1	1	
	2	1 0	
	3	1 1	
	4	1 0 0	
	5	1 0 1	
	6	1 1 0	and so on

Such a binary system relies on place-value correspondences, such as are never used in the syntactic and semantic organization of spoken languages. In fact a simple account can be given of a place-value system in terms of compositional semantics, but such an account makes crucial use of a hidden constant, the base number of the system (2 for a binary system, 10 for a decimal system, and so on). This constant is 'hidden' in the sense that no symbol in the expressions themselves has the value of this constant, which is nevertheless necessary for semantic interpretation. Languages do have expressions which are interpreted with the aid of hidden constants, but these are invariably deictic or supplied by the contexts or situations in which the expressions are used. The base numbers used in numeral systems cannot be seen as given by the context or situation in the same way as elements such as 'speaker', 'hearer', 'time of utterance', and so on.

A simple system sometimes called 'binary', but not a place-value system like (2.3.2) is used in some Australian aboriginal languages, as in a Queensland language cited in Tylor (1981, p. 243).

2.3.3	1	ganar
	2	burla
	3	burla-ganar
	4	burla-burla

This is like the system in (2.3.1), but with two basic terms instead of one. The problem for short-term memory is somewhat lessened. In this system it would be easier in a real life speaking situation to distinguish between the expressions for 3 and 4 than in the unary system of (2.3.1). But again, the cumbersomeness of expressions for higher numbers becomes a problem, and such systems do not go beyond about 5. To get further with a practically usable system, greater lexical resources are required. A typical numeral system has ten basic terms and uses syntactic combinations only to express numbers above 10.

The use of lexical resources up to a certain point, with subsequent resort to syntactic resources, creates a linguistic distinction which is arbitrary in terms of the number sequence itself. The actual numbers themselves are not inherently lexical or inherently syntactic (unless one were to claim that in some sense 1 is lexical and the rest are syntactic). The smoothness in the sequence of numbers is not matched in the sequence of numerals, in which there is a change of construction from single words to syntactic combinations of words at a numerically arbitrary point. This lack of smoothness in the linguistic sequence is not an 'irregularity' as discussed in the previous section, however. Questions of (ir)egularity only come into play once there are syntactic constructions. If it is taken as characteristic of numeral systems, as it is of natural languages generally, that they make infinite use of finite means, a distinction between simple lexical expressions and complex syntactic expressions is unavoidable. But I would wish to claim that the specific boundaries between meanings expressed by simple lexical means and those expressed by syntactic constructions do reflect historical stages in the growth of linguistic systems. The move into syntax is a leap. The Australian language illustrated above made this leap after 2; decimal systems made it after 10. The lexicon/syntax discontinuity in numeral systems, though not technically an irregularity, is nevertheless a growth mark in languages.

In typical decimal systems, there are often genuine irregularities involving both the series of lexical items from 1 to 10 and the syntactic constructions for numbers above 10. I will deal first with the irregularities in the lexical series, which involve the ways in which the simple numeral words relate syntagmatically and paradigmatically to other words.

The first irregularity comes with the difference between singular and plural. English *one* is a singular numeral, but *two*, ..., *nine* are plural. A construction is regular to the extent that its constituents can be replaced by large numbers of other members of the same grammatical category. In *one house*, the first word cannot be replaced by *two*, ..., *nine*; neither can the second word be replaced by any plural noun. In other constructions, where no interaction with a noun is involved, *one* has the same distribution as *two*, ..., *nine*, for example, *twenty-one*, *twenty-two*, ..., *twenty-nine*. But compared to *two*, ..., *nine*, *one* has a unique syntactic effect on a modified noun. Computer programmers providing user messages involving a number and a noun have to insert special *ad hoc* clauses to prevent ungrammatical sequences such as **1 DISK BLOCKS USED* or **YOU HAVE 1 MAIL MESSAGES*. Words for 1, which is semantically singular, have a different semantics from all the other, plural, numerals (a theme to be developed in Chapters 4 and 5).

In many languages, the words for 2, 3, and sometimes 4, behave differently from the words for 5–10. There is a growth mark at around 4. In many inflecting languages (for example, Latin, Russian, Welsh, Ancient Greek) the first few numeral words inflect, that is take various somewhat different forms, agreeing in gender or case as conditioned by their syntactic environment. This is true for words up to about 3 or 4, after which invariant (or in the Russian case, less variant) forms are used. The first two or three numbers are also linguistically marked by having suppletive (irregular) ordinal forms in many languages. Examples are English *first* and *second*, which are phonologically quite unrelated to the corresponding cardinals.

For syntactically complex numerals, I will discuss three formal characteristics which reappear significantly frequently in languages. These formal characteristics may conveniently be labelled 'discontinuity of additive constructions', '1-deletion', and 'base-suppletion'. I will discuss the sense in which these characteristics

are irregular and why such irregular characteristics cannot be accounted for by an innateness theory. Finally I will sketch an alternative explanation for the existence of these universal irregularities, based on the assumption that numeral systems evolve gradually over many generations with intermittent periods of stagnation during which no development takes place. The ensuing chapters of the book essentially flesh out this sketch.

Statement (2.3.4) below holds true for a striking number of languages:

> *2.3.4* Where a language signals addition by more than one method, the point (or one of the points) in the number sequence where the change from one method to another occurs is the point at which overtly multiplicative constructions are first used.

(By 'overtly multiplicative construction' I do not mean a construction necessarily containing a morpheme signifying multiplication, such as English *times*, but rather a construction clearly analysable into two constituents, the product of whose values yields the value of the whole construction.)

To illustrate, English expresses the numbers 13–19 as overtly additive constructions on the pattern *X-teen*. *-teen* is transparently (to the grammarian at least) a form of *ten*. Here a lower-valued numeral is prefixed to *-teen*. After 20, however, there is no affixation, and the lower-valued word follows the higher-valued one (for example, *twenty-one*, *thirty-nine*). *Twenty* is the first point in the system where there is any possibility of a multiplicative analysis of the surface form, as a phonologically modified form of *two-ty*. *Ten* is overtly monomorphemic.

Classical Welsh expresses addition by two connectives *ar* 'on', and *ac* 'and'. *Ar* is used for numbers up to 39 and *ac* is used for all higher numbers. In this system 40 is the first number expressed by an overtly multiplicative construction *deugain* (*dau ugain*, 2 × 20).

In Malagasy, the situation is like English, with *folo*, 10, being monomorphemic, and *roapolo*, 20, clearly analysable as 2 × 10. 'Remarquez qu'on emploie *amby* pour relier les unités aux dizaines (sauf de 11 à 19, ... dans lesquels on emploie *ambin' ny*)' (Rajaobelina, 1966, p. 35). Thus 12 is *roa ambin' ny folo* while 22 is *roa amby roapolo*.

This universal is only a tendency, as there exist clear counterexamples, for example, Italian, in which the method of signalling addition changes after 16 (..., *quattordici, quindici, seidici, diciasette, diciotto, ..., ventuno, ventidue, ...*). In Italian, the first construction for which a multiplicative analysis might be claimed to be transparent is *trenta*, 30, possibly *tre+anta*, where *-anta* is a form for 10. The method for signalling addition does not change around 30.

The claim that the tendency noted here is 'striking' is equivalent to saying that if one assumes *a priori*, as a null hypothesis, that languages are permitted to vary their method of signalling addition at any arbitrary point in the number sequence, then the number of languages that choose to do so at just the point where overt multiplicative constructions appear is statistically significant. Assuming some finite arbitrary cut-off point for a numeral system, say, for the sake of argument 10^6, after which use of the system becomes cumbersome, and given a base of 10, there is in principle a vast range of points at which a language could choose to change its method of signalling addition. Conceivably, for instance, there could be a language like English, except that after *fifty-seven* it reversed the order of summands and continued *eight and fifty, nine and fifty*, and so on. Such a language would be very unusual, to say the least.

A second universal irregularity can be labelled '1-deletion'. All developed systems use multiplication. In the basic case, a word expressing a 'single digit' value is combined with a word expressing some power (perhaps the first power) of the system's base (often 10). For example, we have in Mixtec (data from Merrifield, 1968):

2.3.5 ùù šiko — ùnì šiko — kṳ̀ṳ̀ siko

2 20 (= 40) — 3 20 (= 60) — 4 20 (= 80)

The single-digit expressions in a language form a natural class denoting a continuous sequence of numbers from 1 to the number just before the system's first base number (often 9). A general rule for forming the basic multiplicative constructions takes the form (2.3.6).

2.3.6 PHRASE → DIGIT M

This rule would generate an English structure as in (2.3.7):

2.3.7

PHRASE
DIGIT M
three hundred

(PHRASE and M are universal categories used in numeral systems.)

For many languages, however, it is necessary to state an exception to rule (2.3.6); this exception involves the word denoting the number 1. In *LTN*, the exception was handled by a deletion transformation, which could be optional or obligatory, depending on the construction in question. Examples from various language are given in (2.3.8).

2.3.8 Mixtec [see (2.3.5) above] òkò = 20 [*i̤i̤* šiko (lit. 1 20) is ill-formed.]

Classical Welsh 20 ugain [un ugain (1 20) is ill-formed]
40 deugain (2 20)
60 tri ugain (3 20)
...
180 naw ugain (9 20)

French un cent (1 100) or simply cent (100)
deux cents (2 100)
trois cents (3 100)
...
neuf cents (9 100)

Sar (Chad)

Unités	Dizaines /kùtə/ unité	Centaines / ɓú/unité	Milliers /dūbú/unité
1 kógəm	10 kùtə	100 ɓú	1000 dūbú
2 jō	20 kùtə jō	200 ɓú jō	2000 dūbú jō
3 mətá	30 kùtə mətá	300 ɓú mətá	3000 dūbú mətá
8 sɔsɔ́	80 kùtə sɔ́sɔ́	800 ɓú sɔ́sɔ́	8000 dūbú sɔ́sɔ́
	90 kùtə ndōhó	900 ɓu ndōhó	9000 dubú ndōhó

(Palayer, 1970, p. 55)

(For this language the categories DIGIT and M of rule (2.3.6) would have to be in the opposite order.)

Numeral systems would be more regular, and rule (2.3.6) more fully and transparently productive, if the word expressing 1 were treated exactly like all the other single-digit words, but very often it is not, making it necessary to introduce *ad hoc* machinery into the grammar of the system concerned. Of course, the number 1 is special in this respect; one would not expect a language to have an optional or obligatory deletion rule for words denoting other numbers, such as 2, 3, or 9. (J. D. McCawley (personal communication) tells me that 'Japanese *-bai*, as in *sanbai* "three times (as much)" can be used by itself to mean "twice (as much)"'. Such cases appear to be extremely rare, however.)

A third universal irregularity can be labelled 'base-suppletion'. In English, for example, one can reasonably analyze the suffix *-ty* as denoting 10, that is as synonymous with *ten* and *-teen*. There is a context-sensitive suppletion here, from a formal point of view the same kind of thing as happens with present tense forms of the English copula, *am*, *is*, and *are*. The base number of the system is realized in a variety of phonologically unrelated forms. It is quite common for the base number of a numeral system to be realized in a variety of ways, depending on syntactic context. Some more examples follow.

In French, one might reasonably analyse *-ante* as in *soixante*, *cinquante*, and so on as denoting 10, which in other contexts is expressed by *dix* (and sometimes, depending on one's analysis, by the suffix *-ze*). In the Mixtec examples of (2.3.5) note that 20 is expressed as *òkò*, but the form used in multiplicative constructions is *šiko*, which could be phonologically unrelated. In modern Hawaiian (Pukui et al., 1975, pp. 268–9) we have the forms given in (2.3.9):

2.3.9

		10	umi
3	kolu	30	kana-kolu
4	ha	40	kana-ha
5	lima	50	kana-lima
6	ono	60	kana-ono
7	hiku	70	kana-hiku
8	walu	80	kana-walu
9	iwa	90	kana-iwa

Here, clearly, *umi* is reasonably regarded, in a synchronic grammar, as a suppletive form of *kana*. (The form for 20 is irregular.)

In Telegu (Siromoney, 1968, p. 88) we have the forms given in (2.3.10).

2.3.10

		10	padhi
3	mūdu	30	muppai
4	nālagu	40	nalabhai
5	aidu	50	ēbhai
6	āru	60	aravai
7	ēdu	70	debbhai
8	enimidhi	80	enabhai
9	tommidhi	90	tombhai

Here again, in a synchronic grammar, it seems reasonable to regard the form for 10 as a suppletive variant of the form used in forms for multiples of 10. (The form for 20 is again irregular; clearly some phonological modification takes place for 30–90.)

The relationship between base-suppletion and 1-deletion is implicational. A language with base-suppletion invariably uses 1-deletion in the context of the (suppletive) form for the base number.

One might expect languages to prefer to change their methods of signalling addition at structurally natural points, and it might be suggested that the point at which overt multiplicative constructions tend to appear is just such a structurally natural point. But this suggestion is wrong, as an investigation of what might be meant by 'structurally natural' will show.

Given two linguistic rules, applying in complementary environments, if the difference between the two environments is structurally natural, it will be possible to specify at least one of the environments by means of a simple description in terms of classes of structural items which are motivated independently of the rules in question. The description must, furthermore, not be disjunctive, as this seems to run counter to what is understood by 'structurally natural', and postulating classes with just one member seems unnatural, too, unless they can be identified with the intersection of other, independently motivated, classes. As an example, in English, third person singular subject-verb agreement applies to all verbal items (counting auxiliaries as a subclass of verbal items) except the modals, and modals are identifiable as a class independently of the subject-verb agreement rule. The difference between modals and non-modals is structurally natural in this sense, as is the environment of the agreement rule.

In English numerals, one can perceive through the irregularities that the one-word expressions *one*, *two*, ..., *nine* form a natural class. They can stand alone as numerals uncombined with other words, and they alone may follow a *-ty* word in an additive construction (for example, *twenty-one*, *eighty-four*). Call this class DIGIT. Considering just the expressions from *twenty* to *ninety-nine*, the following formula describes all the well-formed expressions.

2.3.11 DIGIT -ty (DIGIT)

Taking the semantics of these expressions into account, it is clear that they can be interpreted by two operations, one of multiplication and one of addition. These same semantic operations would work perfectly well for such hypothetical expressions as *onety*, *onety-one*, *onety-nine*, (10, 11, 19) which would fit the structural formula in (2.3.11). But the surface syntax of the expressions from *ten* to *nineteen* (except the idiosyncratic *eleven* and *twelve*) requires a different formula, as in (2.3.12) (even, perhaps dubiously, counting *ten* and *-teen* as variants of the same morpheme).

2.3.12 (DIGIT) teen

The presence of the 'DIGIT *-teen*' forms in the system, to the exclusion of hypothetical '*onety* DIGIT' forms, necessitates an *ad hoc* and unnatural distinction between two (unnatural) subclasses of the expressions fitting the formula in (2.3.11). That is, the distinction between the hypothetical '*onety* DIGIT' and all the rest, from *two-ty* (= *twenty* after some phonological adjustment) to *ninety-nine*, is unnatural and can only be described by an ad hoc device in a synchronic grammar aiming to generate all and only the well-formed numeral expressions, regarded as an unordered set.

To make sense of the suggestion that changes in the method of signalling addition occur at structurally natural points in the numeral sequence, some characterization of 'structure' independent of the (surface) method of signalling addition must be given,

if the suggestion is not to become circular. The methodology of specifying all and only the members of a given set by means of generative rules, with a concomitant thoroughgoing insistence on formulating grammars that capture all the significant generalizations inherent in the material at all levels, provides the most impressively coherent way of characterizing linguistic structure yet devised. Applying this methodology to numerals, as is done in *LTN*, yields an analysis which depicts underlying regularity and homogeneity in additive constructions. The contrasting surface discontinuity in the method of signalling addition is accounted for by *ad hoc* machinery, such as suppletions and inversion rules with very limited contexts. Such machinery is countenanced, but only with reluctance, by the methodology of generative grammar.

In short, languages tend to change their methods for signalling addition at a point in the numeral sequence which is seen *not* to be structurally natural when a rigourous (generative) analysis of the structure of numerals is carried out.

Relating this now to the question of universals and language acquisition, it can be seen that the tendency toward discontinuity in the method of signalling addition is clearly a tendency toward *ir*regularity in particular grammars, a tendency for languages to require, in a particular subdomain, *ad hoc* statements in their descriptions. The most regular language would have a uniform way of signalling addition: it would not change its way of signalling addition at a structurally (relatively) unnatural point. Similarly, the tendencies of languages to use 1-deletion and base-suppletion in their numeral systems are tendencies toward irregularity, requiring the formulation of rules idiosyncratically affecting specific individual forms.

The general linguistic theory, claimed by nativists to be innate, specifies the complex and detailed regularities that a child, according to this theory, imposes on the linguistic data made available to him. A child is preprogrammed, according to the nativist account, to intuit certain complex regularities in the language used around him, regularities which are not necessarily transparently obvious from a small corpus. The universality of this preprogramming is held to account for linguistic universals. But nativists admit that no specifically linguistic faculty – that is nothing other than a general inductive problem-solving ability – equips the child to acquire *ir*regularities. Thus, though the invention of **bringed* by an English child is a small victory for the nativist, the nativist has nothing special to say about the eventual acquisition of *brought*. This irregular form is evidently acquired in spite of the innate *faculté de langage*, rather than with its aid.

The irregularity of *brought* is a language-particular fact: the question of how children acquire language-particular irregularities deserves attention, but such a study would not impinge on the hard core of the nativist research programme, because it does not involve universals. The existence of *universal irregularities*, however, begins to undermine the strong support that is often assumed to accrue to the nativist hypothesis from the simple existence of linguistic universals. Innate mental structuring could only explain universal regularities, however complex, but could not explain universal irregularities.

There could well be other examples of universal irregularities. Conceivably, it could be the case that in a significant number of languages the verbs for *come* and *go* are conjugated irregularly. If this were the case, I do not think that it would appeal to a nativist to explain the fact by postulating an innate disposition to expect irregularity in these particular forms. Even if such a hypothesis were advanced, it could soon be refuted by pointing to the absence of spontaneously created irregular forms for these particular verbs in child language. (A spontaneously created irregular form would be a form like *snuck*, a past tense form of *sneak*, apparently totally anomalous. Though *snuck* remains a problem, it is fortunately an isolated one.) Similarly, a nativist would not be attracted by the idea of an innate disposition to anticipate irregularity in the method of signalling addition up to the point in the numeral sequence where the first overtly multiplicative construction appears. This is because a blanket disposition merely to anticipate irregularity at a certain point implies an unmanageably large set of hypotheses confronting the language learner as to what the correct forms might be; the nativist's concern is always to postulate a minimum of hypotheses available to the child.

The discontinuity of additive constructions, 1-deletion, and base-suppletion are universals. But because they are tendencies to *irregularity*, they cannot be explained by appeal to innate linguistic preprogramming. It should not be (and perhaps is not) contentious to claim that not all linguistic universals are genetically determined. What follows from this claim? As much as follows from the converse claim that some linguistic universals *are* genetically determined (which I do not deny). What follows is that there is a potential research programme developing non-nativist theories of the determinants of language form.

2.4 Explaining Non-absolute Universals

If the sole concern of linguists was to identify properties shared by every single (speaker of every) language, they would have little to show for their labours. I know of no non-trivial statement which can indisputably be said to hold true for all known languages. On the other hand there are large numbers of interesting and quite possibly true exceptionless but conditional statements about languages. An example is Greenberg's 3rd universal: 'Languages with dominant VSO order are always prepositional' (1963a, p. 110), which is not a statement about all languages, but only about VSO languages. Such universals are conditional in the sense that they have a logical form like 'If x is VSO, then x is prepositional'. Furthermore, there are nontrivial unconditional statements which hold true of many but not all languages. Greenberg's 1st univeral is an example: 'In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object' (p. 110). Note the 'almost'. Many interesting universals both fail to apply to all languages and are conditional, for example, Greenberg's 4th universal, 'With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional' (p. 110).

Objections could conceivably be raised to the use of the term 'universal' for statements other than those applying unconditionally to all languages: but this issue is only terminological and I shall follow the usage adopted by linguists generally in applying 'universal' to conditional statements and statements to which there are exceptions. Understanding 'universal' in this way admits certain problems.

Firstly, there is the problem of defining significance. Many conditional statements or statements about some subset of the world's languages are trivially true – for example 'If a language has a word *Weltanschauung*, then it also has a word *Schadenfreude*' or 'x% of the world's languages have a word for snow'. Obviously one wants to exclude universals such as these from serious consideration. One wants to consider only those facts about natural languages that are in some sense surprising, given the conceivable possibilities. Most, and possibly all, of Greenberg's universals are surprising or striking in the required sense. The strikingness can in part be reduced to straightforward statistical significance, as suggested in Hurford (1977). In this sense Greenberg's 1st universal, quoted above, is significant (or would be, if given the appropriate precise numerical form). That is, the number of OS languages is startlingly low. The equation of significance with statistical improbability is implicit in Greenberg's statement of many of his universals, where phrases such as 'with overwhelmingly greater than chance frequency' occur. But the question of the significance of generalizations cannot be entirely reduced to statistics. For instance the two trivial universals given earlier in this paragraph could be shown to be statistically highly unlikely to be true (and therefore significant). It might be correct, as suggested in Hurford (1977), to solve this problem by judging that these trivial statements, regardless of what a statistical theory of significance may have to say about them, are *not generalizations* and for that reason not worthy of our attention. But this move may be no more than using the intuitive notion of what constitutes a generalization (as opposed to the statistically defined notion of significance) as a convenient hidey-hole for problem cases. I offer no solution to these problems here, but take the view that there are a large number of non-trivial universals (however they may be identified) which apply only conditionally to languages or else express striking tendencies to be observed in languages. The point has to be made because almost all of the properties of numeral systems that I shall discuss are universal tendencies. That is, they are intuitively non-trivial statements true of a strikingly large number of languages, although not completely exception-free.

Secondly, having made the methodological decision that conditional universals and universal tendencies are not to be dismissed, there is the problem of accounting for them within any theory (such as a version of Chomskyan nativism made suitably precise) whose machinery makes absolute predictions. With absolute universals, there can be a direct logical inference from the uniformity of the genetic inheritance to the universality of the acquired structures, but if the structures in question are not universally acquired, such a direct inference is not possible.

In some cases of conditional universals, the logical mechanism of the Chomskyan account can easily be altered in an appropriate way: these are cases where the stipulated condition must *logically* be met in order for the statement to hold. An example is Greenberg's 38th universal: 'Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings the subject of the intransitive verb' (1963a, p. 112). Clearly, logically the second clause in this statement could not hold true where there is no case system. To account for an instance like this, a natural move for a nativist account would be to posit an innate disposition to acquire a case system of the sort described, and to stipulate further that this disposition is only activated on exposure to primary linguistic data in which the existence of a case system is discernible. This is a slight and apparently natural extension of the Chomskyan account as sketched above.

All of the properties of numeral systems which I shall discuss are of this logically conditional type. That is, they take the form, 'If a language has a numeral system at all, then that system has property *X*'. In keeping with the comments above, I make the innocuous assumption that any child acquiring a numeral system, with whatever properties, must normally be exposed to data indicating the clear presence of a numeral system. That is, a normal child does not produce a numeral system conforming to the patterns generally observed in languages, *ex nihilo*, like a rabbit from a conjuror's hat. But I shall propose an alternative to the kind of nativist explanation which attributes a large innate contribution to the structure of the system from each new language acquirer.

The alternative explanation attributes to the *normal* acquirer of a numeral system no particularly spectacular disposition to make generalizing leaps beyond the data to which he is exposed. Rather, he receives fairly clear and explicit exemplars, and even a degree of deliberate instruction. The focus in explaining universal characteristic properties of numeral systems now falls on the abnormal case. I shall argue that *some* people, children probably included, are indeed capable of producing little bits of numeral systems *ex nihilo*, as if from a conjuror's hat. In short, there are (rare) linguistic inventors, and I presume that the process of linguistic invention is constrained and shaped by the mental tools with which the inventors are endowed. So the account of a logically conditional universal like 'If a language has a numeral (or case, tense, gender, pronominal, and so on) system at all, this system has property *X*' now runs 'If (part of) a numeral (case, tense, and so on) system is invented at all, the circumstances of a human inventor (including his mental structure) will lead him to invent it with property *X*.'

There are two important differences between this account, which still relies on (presumably innate) mental attributes of humans, and the Chomskyan account. Firstly, this account locates the source of universal properties not in all, but only in some, language users, namely the inventors. And secondly, the time-scale is expanded, in that the universal properties of a system, rather than being rapidly induced afresh by each new acquirer, may conceivably be introduced piecemeal over a long historical period: once present in the system they are acquired by new speakers by means requiring less attribution of rich and specific innate mental structuring. Chomsky claims, in effect, that we all in a sense rapidly invent our own language on the basis of skimpy examples, rather than learn it. For numeral systems, at least, I claim that relatively few people invent and most simply learn, and that several non-trivial universal characteristics of numeral systems are plausibly accounted for under this view.

For conditional universals where the connection between the condition and its following clause is not logical, but contingent, an innateness account would have to attribute to the newborn child knowledge of some ruly connection between the statement and its condition; for example between being a prepositional language and being a VSO language. The relationship between the two linguistic properties in this case is qualitative, perhaps capturable by postulating some innately known category to which verbs and adpositions (that is pre- or postpositions) both belong.

One can envisage a more quantitative relation between properties connected by a conditional universal such that some structure X is (contingently) more easily acquirable by someone who already possesses structure Y, but not logically beyond acquisition by someone without structure Y. An analogy (not completely apt) is the case of Spanish being more easily acquirable by an Italian speaker than by an Eskimo speaker; this analogy lacks the required unidirectionality of the connection, however (as Italian and Spanish can be interchanged). An analogy having the required unidirectionality would be the case of the ability to speak, which is easier to acquire for hearing than for deaf children, but not absolutely (and certainly not logically) beyond acquisition by the latter. I shall argue in Chapter 3 that for numbers up to about 10, knowledge of particular numeral word/concept pairs (for example, *seven*/7) is more easily acquirable by someone who already has the preceding word/concept pair (for example, *six*/6), because of the ways in which basic numeral word/concept pairs are learnt, although the possibility of acquiring *seven*/7 before *six*/6 is certainly logically conceivable. This style of explanation clearly requires an innate ability to form mental representations of the (in some sense abstract) objects concerned, yet does not seem to stipulate pre-existing knowledge of the connection between the objects.

The logical mechanism of an innateness theory of universals is more problematic in the case of universal tendencies than in those of absolute universals (conditional or otherwise). Where there are exceptions to any universal statement, however strikingly few, one faces the problem of accounting for the non-application of the relevant innate disposition in these cases. The type of mechanism that is advanced by innateness theorists for these cases involves notions of markedness. The exceptions to universal tendencies – for example, the few OS word-order languages that have been observed – are said to be marked. In exceptional circumstances, a child may internalize knowledge of a marked system, but normally, *ceteris paribus*, he will be more strongly disposed to internalize an unmarked system. If a marked system is not transparently discernible in the observed data, then the child 'assumes' (unconsciously of course) that the system to be built is the one he is innately equipped to prefer, and accordingly internalizes an unmarked system. Circumstances external to the postulated innate LAD are responsible for the exceptions to universal statements. No detailed theory of the interaction between internal innate factors and external factors in the primary linguistic data is in fact articulated, but it is not impossible in principle that some detailed acceptable account along these lines will ultimately be formulated. For the present the question remains: why, exactly, are there exceptions to striking universal tendencies? And why, indeed, do individual languages sometimes prefer to shift from an unmarked system to a marked one? The problem of exceptions to otherwise striking universal tendencies is a major challenge to any kind of explanation for universals, including explanations from innateness.

The problem in accounting for universal tendencies, as opposed to exceptionless universals, arises from the typical monolithic deductive form of the explanations offered. Work on explaining linguistic universals is at a relatively primitive stage, where often a single type of explanation – for example, innateness, functional, monogenesis – is offered, without due consideration of the possibility of interference from other factors. That is, the logical structure of an explanation for a universal is often of the form:

2.4.1 Factor (e.g. innate structure) → Universal

Perhaps a charitable reading would always interpolate a '*ceteris paribus*' clause into such explanations. In fact, as language is a tool of human interaction, other things are seldom, if ever, equal. A satisfactory style of explanation needs to provide an explicit place for the interference of other factors, ideally not merely saying that if other things are not equal then the proposed explanation will not work, but specifying as far as possible the ways in which different kinds of interfering factors will affect the phenomena concerned. The structure of explanations will thus look more like:

$$\textit{2.4.2}\quad \left.\begin{array}{l}\text{Factor 1} \rightarrow\\ \text{Factor 2} \rightarrow\\ \quad\vdots\\ \text{Factor } n \rightarrow\end{array}\right\} \rightarrow \text{Universal tendency}$$

I call this style of explanation 'ecological'. Explanations of why biological species have developed particular physical characteristics suffer generally from appearing unpredictive and *post hoc*. The gut of the giant panda is something of a mystery, given its diet. But there can, nevertheless, be better or more convincing explanations of this sort, in terms of the features of the particular ecological niches which species occupy. The more one can point to different aspects of the external environment, to which aspects of the structure of the creature or plant concerned are plausible adaptations, the more convincing is one's overall explanation. In cases such as these, where one is explaining something initially mysterious in terms of a coincidence of known and possibly even familiar phenomena, Ockham's razor does not apply. Explanations in the form of (2.4.2) are not to be discounted in favour of those of the form of (2.4.1) because they are less 'economical'. The entities concerned are not abstractions postulated to account for observed patterns; since they exist already, the theorist is not multiplying entities in appealing to them. One can see languages as occupying 'ecological' niches in a space bounded by individual human psychology and the exigencies and limitations of human

social interaction. This ecological mode of explanation will also be adopted in Chapter 3 to explain the continuity of the number sequence itself. Thus *both* a command of order and a command of one-to-one mapping of collections are taken to pre-exist the human command of number and *jointly* to create the conditions in which number and numerals can arise. No embarrassment will be felt at not reducing the notion of number to just one or the other of these factors.

In the absence of a model of human interaction, the linguist can build the effect of human interaction into his explanatory models in an idealized way by postulating random interaction between members of a speech community. That is, a null hypothesis about the nature of interaction can be assumed. So long as a claim is made about the effect of the sum of social encounters between speakers, it is sufficient at the present stage of knowledge to work with this null hypothesis. But note that this is not the null hypothesis about the effect of interaction: that is it is not the hypothesis that human interaction has no effect on linguistic form. It is, rather, assumed that human social interaction does have an effect on linguistic form, and that such interaction is essentially random.

The introduction of a random component into an explanation does not necessarily weaken it; in fact it opens up the possibility of predicting statistical distributions. Explanations in classical genetics have a crucial random factor. When half the chromosomes from a diploid cell are selected at meiosis to form a haploid cell, this selection can be regarded as essentially random. This is not to deny that there can be weightings, so that different chromosomes have a different chance of selection, but given a sample space defined by these weightings, the selection of items from the space can be taken as random. Thus an elegant explanation is provided for the perpetuation of populations with mixtures of inherited characteristics in enduring, or regularly changing, proportions. And of course, randomness also plays a crucial role in explanations in quantum physics.

Postulating random interactions between speakers for the purpose of explaining linguistic universals need not bring any special philosophical difficulties concerning the undermining of causality. At the present stage of knowledge, postulating random interactions can simply be taken as a profitable methodological step, regardless of whether human interactions are in the last analysis really random or really caused.

Admitting the interfering effect of random interaction between speakers in principle allows the possibility of exceptions. Almost all the universal characteristics of numeral systems to be discussed in this work are striking tendencies to which there are exceptions. This study will not appeal in any specific way to the interfering effect of social interaction, except in the last chapter (Chapter 6) where a computational model is given of the social process by which numeral systems arrive at standardized base numbers. It will be seen that in this case postulating random interaction between speakers actually obviates the need to attribute implausibly powerful arithmetical capacities to ordinary individuals. The cumulative effect of random interactions is to build a pattern followed by individuals, but not in any direct sense attributable to individuals' original mental structure. Numeral systems tend very strongly to conform to such patterns, but do not invariably do so.

The processes by which new linguistic structures are invented and propagated within a language community are subject to the same kinds of constraints as other linguistic interactions. It is to some extent a matter of chance what gets invented first (although clearly some inventions are necessary before others can be made) and which inventions survive the test of communicative usage. Since, as I shall argue, invention is one of the main sources from which structures get into a language, the inventories of structures possessed by languages are due in part to haphazard social processes.

2.5 Invention and the Non-universality of Numeral Systems

Not all languages have numeral systems (Dixon, 1980, pp. 107–8). While I will concentrate in most of this study on the existence of certain universal properties of numeral systems, the actual non-universality of numeral systems themselves is highly relevant to basic theoretical arguments relating linguistic universals to innate properties of the human mind. I will begin by discussing this basic argument, using the non-universality of numeral systems to bring out a necessary distinction, that between the *ordinary acquisition* of a system and the *invention* of a system.

If monkeys could talk, they would. This, unsubtly, paraphrases Chomsky's view of the relation between the universality of

language in human communities and certain species-specific mental characteristics. In Chomsky's words: 'it is just extraordinarily unlikely that a biological capacity that is highly useful and very valuable for the perpetuation of the species and so on, a capacity that has obvious selectional value, should be latent and not used. That would be amazing if it were true' (1982, pp. 18–19). For Chomsky, then, it is practically inconceivable that a highly useful and valuable capacity should be merely latent, without being actually used. Obviously, a numeral system, being only a part of a language, is not so highly useful and valuable as a whole language, but, equally obviously, numeral systems are useful and valuable. Imagine two marginally different humanoid groups competing for resources in the same habitat. One species can count, the other cannot. *Ceteris paribus*, the counting group would have an advantage over the non-counters. Members of communities without numeral systems (for example, Australian aborigines before contact with the colonizing culture) clearly have no trouble in learning a numeral system, so the capacity for a numeral system can be said to be latent in these individuals, though not used. Boas (1939, pp. 218–19) emphasizes that lack of a rich numeral system in no way demonstrates lack of ability to form higher numeral concepts, when exposed to them. It must be an advantage of some kind to have a counting system, yet Australian aborigines in their native languages find no compelling need for one, even though they have the innate mental apparatus to master an elaborate numeral system. So it is in fact possible for a valuable and useful capacity to remain latent, without being used. The extent to which the existence of latent but unused capacities undermines Chomsky's argument is uncertain, due to the practical impossibility of comparing the degree of usefulness of a numeral system with that of a language as a whole. Just how much of an advantage would possession of a numeral system be? But the existence of unused latent capacities should lead us to examine the Chomskyan argument in more detail. In what follows, I am not in fact concerned to show that Chomsky's conclusions about apes are wrong, but rather that the usual conclusion about humans and their innate mental characteristics needs to be developed in a more complicated way.

Chomsky's nativism is, above all, a hypothesis about language acquisition by children. The existence of innate mental capacity enables children to acquire complicated systems on the basis of (allegedly) degenerate data. But the innate mental ability, on the part of individuals, to acquire a system, given suitable triggering experiences does not guarantee that a community of such individuals could, spontaneously and without such triggering experience, invent such a system. Imagine a group of wolf-children, humans brought up together but with no experience of human language. Would they develop (invent) some rudimentary human language? It is not obvious that they would, despite the advantage that would clearly result. The 'if monkeys could talk' argument neglects the difference between learning an established system and developing, or inventing, a new system.

> Actually the circumstances in which language must have begun represent a combination for which we can provide no instances. We have animals among themselves, animals in linguistic communities, and humans among animals, and in none of these cases does language develop. We have humans raised in linguistic communities and, in these circumstances, language does develop. What about a human born into a human society that has no language? We don't know of any such societies and so we don't know of any such individuals. But these must have been the circumstances of language origination. (Brown, 1958, p. 192)

Jespersen (1964, pp. 180–188) devotes a whole section to discussion of pairs of people in isolated situations, including some twins he managed to study himself. Clearly, 'private' languages, largely unintelligible to outsiders, do develop in such situations. But what is not clear is how much of an example of language in use needs to be set by the society of origin of these couples, in order for the cosy code truly to take off. The elaborate secret language reported by Diehl and Kolodzey (1981) is clearly based on English. Jespersen seemed convinced that new languages could spring up quite spontaneously, even giving some credence to the idea that this accounted for the large number of indigenous languages of California. But there is no convincing evidence of this.

Chomsky postulates that in the prehistory of Man a (series of) biological change(s) took place such that homo-plus-faculté-de-langage is a descendant of homo-minus-faculté-de-langage. (One need not go into the terminological question of whether 'plus-faculté-de-langage' *equals* 'sapiens'.) Merely postulating these changes is not sufficient to account for the fact that Man is a

language-using species. The implicit definition of faculté de langage, for Chomsky, is an innate ability to acquire a grammar, given suitable triggering data, that is pre-existing language use. So the evolved ability to acquire a grammar is not in itself an explanation of how Man got language (use) in the first place. One must postulate a capacity in at least some humans to invent (subparts of) language systems over and above the capacity to acquire them given triggering data. The lack of numeral systems in Australian aboriginal languages is attributable, not to any incapacity in the aborigines to acquire such systems, but to the fact that no aborigine ancestor ever invented a numeral system (or if one did, the invention was later lost, which is less plausible). Putnam (1980, p. 297) mentions the possibility of primitive language being invented 'by some extraordinary member of the species', but does not develop the idea. (Vico is sometimes mentioned in discussions of linguistic invention. Beyond agreeing with Vico that language is in some sense invented, I would not wish to link my own argument with his rather fantastic speculations on the origins of language.)

To clarify the notion of invention I have in mind, I agree with Chomsky in the following: 'Have we, as individuals, "made" our language? That is, have you or I "made" English? That seems either senseless or wrong' (1980a, p. 11) But in what follows, an unnecessary reference to choice is brought in. 'We had no choice as to the language we acquired; it simply developed in our minds by virtue of our internal constitution and environment.' It is not relevant to whether something has been invented that the inventor 'chose' to invent it. Clearly invention involves deliberate action, but an inventor does not necessarily 'choose' among alternative inventions somehow simultaneously available to him. Chomsky continues: 'Was the language "made" by our remote ancestors? It is difficult to make any sense of such a view. In fact, there is no more reason to think of language as "made" than there is to think of the human visual system and the various forms that it assumes as "made by us"' (p. 11). Certainly no individual remote ancestor made the whole of our language. But individual remote ancestors may well have contributed little bits. No-one would deny that new words and phrases are coined by individuals and that these are passed down to future generations. But for Chomsky it seems as if words and phrases are not so much part of a language as *vehicles for* language, which for him is a matter of deep abstract principles. A language is undoubtedly organized according to deep, abstract, general and far-reaching principles. I would not want to claim that these principles could have been invented in any reasonable sense of the word. But a language also possesses specific, relatively superficial properties, including, but not restricted to, its vocabulary. There is a cline of depth or abstractness between a language's superficial characteristics and its deepest underlying principles, which it presumably shares with other languages. We cannot prejudge the issue of exactly how far along this cline properties of languages may be invented by individuals. And we cannot assume that properties shared by all (or many) languages were not invented, in some reasonable sense of the term. There may well be innate constraints on what is inventable, which determine a narrow range of actual inventions (see the argument near the end of this section).

The idea of invention, in any area, but especially in the area of language, must be approached with due caution. It is often merely a useful simplifying assumption to speak of the invention of something as if this were a definite event taking place at a particular time and place, brought about by a single agent. In rare cases, perhaps, a single genius makes a great conceptual leap in isolation and carries out its practical realization alone, or in the case of a less practical idea, single-handedly produces a monolithic statement of the conceptual advance. But more often, all one can do is identify some transitional period before which the invention was not in evidence and after which it definitely was. During the transitional period the new idea simultaneously develops and gains currency, by increments both large and small, and possibly even with some backward steps. During this period, individual people play different roles. Some play a completely passive role, as mere 'consumers' or receivers of the new idea; some both receive the new idea and actively transmit it to others, without contributing to the idea themselves; and some develop and add to the idea in original ways. These roles may not be sharply differentiated: there may be a continuum from the most passive receiver of an idea to its most creative developer. And individuals who contribute creatively to the development of one idea may have little to contribute in other fields. (Schon (1963) investigates in detail some of the philosophical problems with the notion of invention, emphasizing the role of metaphor and analogy in the displacement of old concepts by newly invented ones.)

In the area of language, all humans play the role of consumer, or acquirer, with ease, in fact with such ease that the need for individuals to play a deliberate role as transmitters of language is reduced (although not eliminated): for much of language, if there is enough of it around, children just pick it up. But the role of inventor, or developer, is not played with such ease, or so frequently. Few individuals add to the stock of syntactic constructions in their language, although the nonce-formation of individual words and phrases is more common.

No idea is utterly new, and there are degrees of novelty, as well as of inventiveness. Chomsky's theory of language acquisition is that each new language acquirer in a sense reinvents the language of his community. The particular kind of inventiveness (granting for the moment the aptness of the term) involved in ordinary language acquisition is common to all normal children. Pateman (1987, Chapter 3) gives three instances of more genuine inventiveness, in the sense that rules or structures are produced which the prior language-users in the communal environment did not possess. These instances are: children inventing aspects of a new creole (citing Bickerton, 1981); second language learners inventing an 'interlanguage' (citing Corder, 1981); and deaf children of hearing parents inventing a system of gestures (citing Feldin et al., 1978). Apparently, then, the invention of such structures in such situations comes fairly readily to typical humans. But it seems that the invention of (part of) a numeral system does not come so readily. The development of a numeral system is a significant addition to the riches of a language, a step which has not been taken in the histories of all peoples. Though all humans appear to have the capacity to acquire a numeral system, only some humans have the attributes or the opportunities which give rise to the development of a numeral system *de novo*.

A language invention capacity could be innate in all humans, just as the ordinary language acquisition capacity, but might require a kind of triggering environment which happens in practice to be rarer than what is required for ordinary acquisition alone. Conceivably, all humans could have the language invention faculty, but this might only flourish in certain cultural circumstances. Presumably a cultural environment in which innovation and deviation from convention was strictly discouraged would not be congenial to the invention of genuinely new forms going beyond the existing competences of members of the community.

Perhaps effectively very small linguistic communities living nomadically at subsistence level do not offer much opportunity for language invention to flourish, either. Perhaps cultures in which there is no amassing or exchanging of wealth, however meagre, tend not to give rise to the invention of numeral systems. But several Pacific cultures – for example, Hawaiian, Tongan, and Pukapukan – testify against this, having words for precise values as high as 100,000 without any obvious strictly practical use for them. (See *LTN*, pp. 202–3; Tylor, 1891, p. 241, Seidenberg, 1960, p. 278 and references cited there.) On the other hand, it is quite conceivable that individuals differ in their innate capacity to invent novel features which may become part of their own, and others', competence. Some may be richly endowed with the relevant inventive capacity; others, possibly, not at all.

It seems reasonable, and in keeping with the common meanings of these terms, to regard invention as a special case of acquisition. Restricting the discussion to the case of the invention or acquisition of elements of a language, in both cases the individual's competence is changed by the addition of some extra piece of (linguistic) information. In the prototypical case of acquisition, the acquirer is catching up with more advanced members of his community, in acquiring a rule, or item, already possessed, in more or less the same form, by them. In the case of invention, the speaker adds to his own inventory a rule or item not already possessed by the rest of his linguistic community. It seems reasonable to say of a speaker who has invented an item that he has acquired it, but the converse is not necessarily true.

That invention is a special case of acquisition is further reinforced by the following argument. Imagine two linguistic communities, absolutely identical except that in one an item X is part of the language, whereas in the other X is absent. (That is the two communities would become identical if someone in the X-less community were to invent X and it were to become current.) Now any particular individual with the capacity to invent X in the X-less environment would presumably also have the capacity to acquire X if placed in the other (X-ful) environment. But not all those individuals capable of acquiring X in the X-ful environment can be credited with the capacity to invent X had they been born into the X-less environment. To illustrate with a non-linguistic example, it took an extremely rare kind of person (an Einstein) to invent the general theory of relativity, but it takes a less rare kind of person to assimilate it once it exists. The

first Einstein was not exposed to the theory, but 'acquired' it by invention. Presumably a second Einstein would have no difficulty at all in acquiring the theory if exposed to it. Wilder, quoting an anthropologist Ralph Linton, implies that invention by abnormal individuals played a role in the development of elaborate numeral systems: 'if Einstein had been born into a primitive tribe which was unable to count beyond three, lifelong application to mathematics would probably not have carried him beyond the development of a decimal system based on fingers and toes. (1968, p. vii). To summarize so far, the actual *non*-universality of a particular, easily acquirable, subsystem of language indicates that an innate capacity to acquire a system on exposure to relevant data does not of itself guarantee the existence in use of that system. To account for the original genesis of linguistic (sub)systems, there must be a capacity for language invention, as yet only vaguely specifiable, possessed, presumably innately, by at least some humans. Invention is a special case of acquisition, in the sense that instances of the invention of items are a proper subset of instances of the acquisition of items.

The inclusion relationship between acquisition and invention has implications for arguments relating language universals to the ordinary (non-inventive) acquisition of language. Two kinds of hypothesis might be advanced to explain the universality in languages of some property X:

2.5.1 Acquisition Hypothesis
Given data (possibly null) compatible with X (but not necessarily entailing X), an individual can only acquire a system with property X.

2.5.2 Invention Hypothesis
Given zero data, an individual can only invent a system with property X.

An acquisition hypothesis entails the corresponding invention hypothesis, since invention is a subcase of acquisition. The case of zero data is simply one of many cases of data compatible with X: zero data is compatible with any structural property, therefore with X, whatever it may be. If something is beyond acquisition by any means, it is beyond (acquisition by) invention. Consequently, one cannot consistently maintain an acquisition hypothesis *against* the corresponding invention hypothesis. This can perhaps best be shown quite formally.

For this purpose, let X be the predicate 'has the property X', I the predicate 'is inventable *de novo*', and A the predicate 'is acquirable given (possibly null) data'. Then, where x ranges over systems:

2.5.3 $$\forall x\ [I(x) \rightarrow A(x)]$$

that is, if a system is inventable it must be acquirable. This states the logical or semantic relationship between acquirability and inventability. An acquisition hypothesis accounting for the universality of X would have the following content:

2.5.4 $$\forall x\ [A(x) \rightarrow X(x)]$$

that is, this hypothesis claims it is (contingently) the case that any acquirable system must have property X. A corresponding invention hypothesis would have the content:

2.5.5 $$\forall x\ [I(x) \rightarrow X(x)]$$

that is, any system inventable *de novo* must have property X. Now one cannot hold (2.5.3) and (2.5.4) true, but (2.5.5) false without becoming involved in a contradiction. The negation of (2.5.5) would be equivalent to (2.5.6).

2.5.6 $$\exists x\ [I(x)\ \&\ {\sim}X(x)]$$

that is, there is at least one system which is inventable *de novo* but does not have property X. By existential instantiation:

2.5.7 $$I(s)\ \&\ {\sim}X(s)$$

that is, some arbitrary system s is inventable *de novo* but lacks property X. Now, via *modus tollens* from (2.5.4) and the second conjunct of (2.5.7):

2.5.8 $${\sim}A(s)$$

that is, s is not acquirable (since it lacks property X). And via *modus tollens* from (2.5.3) and (2.5.8):

2.5.9 $${\sim}I(s)$$

that is, s is not inventable (since it is not acquirable). But (2.5.9) contradicts the first conjunct of (2.5.7), which was also derived from the same premises. That is, trying to maintain an acquisition hypothesis while rejecting an invention hypothesis leads to the contradiction that some arbitrary system is both inventable and not inventable.

Conversely, however, an invention hypothesis of the type (2.5.2) can, logically at least, be maintained while denying the corresponding acquisition hypothesis. In principle, evidence for the position in which an invention hypothesis is true but the corresponding acquisition hypothesis is false would come from a situation where one could actually watch the processes of both linguistic invention and ordinary language acquisition in a community. In practice, however, no observer of the linguistic scene has satisfactory access to such evidence.

If an invention hypothesis were true and the corresponding acquisition hypothesis false, the universal concerned would still demand explanation, since there could conceivably be a community somewhere whose ancestors had invented property X, but in which this property had later fallen from use and been lost. An invention hypothesis requires an auxiliary hypothesis explaining how X, once invented, is faithfully transmitted to subsequent generations. For 'abstract' formal universals, not transparent in data to which a child is likely to be exposed, this may be problematic. For more superficially obvious properties of language systems, one can claim that the data to which children are exposed are rich enough to permit learning of the relevant property with a relatively unspectacular innate contribution from the learner. The universal properties found in numeral systems are generally of the latter, rather 'surfacy' type.

In fact, neither kind of hypothesis, if taken alone, has much explanatory potential. At best they are generalizations about ordinary acquisition and invention, respectively, and they say nothing about the bases of the generalizations concerned, whether psychological, social, or whatever. Determinants of the processes of ordinary acquisition and invention can conceivably be either external (for example, connected with the communicative purposes to which language is put) or internal to the individual (that is psychological or mental). Explanations for universals in terms of constraints on acquisition or invention only begin to get interesting when they discriminate in a clear way between the possible bases or determinants of such constraints.

I assume that there are strong innate psychological constraints on what is inventable, and in what circumstances. There is an obvious quantitative constraint, which causes development typically to proceed in small increments. Whole systems are not invented at a stroke. A typical invention only supersedes its predecessors by a small amount. This quantitative restriction is presumably a function of both the mental limitations of individual inventors and the (part social, part psychological) limitations on how much innovation a community can take up at a time. There are possibly also qualitative constraints, of course.

Numeral systems universally show marks of successive phases of invention in the building up of the whole. It is one of the main contentions of this book that such universal growth marks should be accounted for in terms of historically sporadic invention and that the only role played by ordinary acquisition is in the relatively faithful transmission to later generations of the increments made by invention. This argument will be taken up next in the last section of this chapter.

2.6 Explaining the Universals Evolutionarily

The universal irregularities that I have mentioned can be explained by appealing to the way numeral systems evolve historically. The existence of these universals is due to the fact that numeral systems evolve slowly from humble beginnings over many generations, with periods of stasis punctuated by occasional innovative extensions. Note that, to the extent that human psychology figures in the argument, an indirect contribution from innate human characteristics is not, and could hardly be, ruled out.

At this stage it is helpful to digress somewhat to sketch an outline of what often seems to be the succession of historical stages by which counting and numeral systems emerge. What I give here is no more than a brief summary of the conclusions drawn from various surveys of the anthropological literature, much of it nineteenth century, (for example, in Tylor, 1891; De Villiers, 1923; Conant, 1923; Dantzig, 1940; Menninger, 1969) . A lot of this literature strikes a modern reader as condescending in tone, with frequent use of expressions such as 'savage', 'primitive', 'barbarous', 'less civilized races', and so on, and perhaps the facts reported lack some insight and understanding, but there seems to be no reason to doubt the factual essentials of the sketch I give below. The contrast between the voluminous reporting of primitive counting practices by nineteenth century scholars and the relative lack of work on the subject in the twentieth century is striking. But several recent works on Papua New Guinea, which are clearly well-disciplined and objective, confirm the general picture emerging from the nineteenth century collectors: Lean (1985–6) is a survey work combining modern systematic comprehensiveness with an orientation toward data very reminiscent of nineteenth century collecting; and articles by Saxe (1981, 1982a, 1982b) and Saxe and Moylan (1982) give an in-depth picture of the bodypart counting practices of the remote Oksapmin people.

An early and simple way of representing the number of objects in a collection, or the number of days in some period, is by means of a corresponding collection of sticks, or pebbles, or whatever. If a man wants to remind himself how many pigs he has, he makes a pile of pebbles, one for each pig. The pebble pile serves as his record of the size of his holdings, even when the pigs themselves are scattered and out of sight. Far from being superseded by more modern methods, this method still has some advantages; to this day, cricket umpires in Britain count the six balls in an over by shifting one of six buttons, or little stones, from one hand to the other with each successive ball, and when a hand is empty, they announce the end of the over. This method has some of the advantages of permanence that come with writing. But the sticks, or buttons, or whatever, have no names, and therefore this system provides no names for the numbers themselves. It may not always be easy to find convenient sticks or pebbles. Fingers (and toes) are always available, so, instead of representing the number of his pigs by a pile of pebbles, a man may represent them by holding up the appropriate number of digits (if his collection is small enough). But then it is not easy to use his hands for other things without breaking up the digital representation. To overcome this problem partially, a single bodypart (finger, knuckle, part of the palm) can be made to stand conventionally for the row of other parts (fingers, knuckles) it takes to reach that part. For example, the index finger of the left hand can stand for the four fingers of the left hand; the left thumb can stand for all the digits of the left hand. But this presupposes the development of a conventional order in which the relevant bodyparts are touched. The examples given would work only if everyone started counting on the little finger of the left hand, moving in the obvious way towards the thumb. This is the type of method that Dixon reports: 'Aboriginal Australians did have ways of measuring and indicating, say, the number of days until some planned social event, through pointing at different points on the palm of the hand; this is the only example of a type of "sign language" being employed in place of the lexical resources that languages from other areas would use' (1980, p. 108). With this method, then, there are still no verbal names for numbers, but the various bodyparts, fingers, and so on could themselves be taken as signs for the numbers. And there is, at this stage, a conventional ordering of these bodypart signs.

Typically, the fingers and other bodyparts involved have verbal names. While pointing at a bodypart sign for a number, the verbal name of that bodypart can be uttered. Conant, citing Haddon (1889) describes the following counting method from the Western Torres Straits:

> Beginning with the little finger of the left hand, the natives counted up to 5 in the usual manner, and then, instead of passing to the other hand, or repeating the count on the same fingers, they expressed the numbers from 6 to 10 by touching and naming successively the left wrist, left elbow, left shoulder, left breast, and sternum. Then the numbers 11 to 19 were indicated by the use, in inverse order, of the corresponding portions of the right side, arm, and hand, the little finger of the right hand signifying 19. The words used were in each case the actual names of the parts touched; the same word, for example, for 6 and 14 [i.e. 'wrist']; but they were never used in the numerical sense unless accompanied by the proper gesture, and bear no resemblance to the common numerals, which are but few in number. (1923, pp. 17–18)

What is of interest here is not the particular sequence of bodyparts used (which is what catches Conant's attention), but the extension of the use of a word for a bodypart to signify a number. But the fact of this extension is apparently signalled overtly by a pointing gesture. At this stage, the 'sign language' of the previous stage is still present, but now a conventional sequence of words is established.

Also of interest is the fact that these bodypart/number terms 'bear no resemblance to the common numerals, which are but few in number'. One must assume that these 'common numerals' were numeral words, in my terms, integrated into the rest of the language, for example, by being usable as noun phrase modifiers, or perhaps predicates. Although there are other cases of the expressions used in counting differing slightly from the numeral expressions integrated into the rest of the language (for example, in Arabic), it is the case much more usually that the expressions used in counting and in numeral systems coincide.

Saxe gives the following modern description of a bodypart counting system: 'To count as the Oksapmins do, one begins with the thumb on one hand and enumerates 27 places around the upper periphery of the body, ending on the little finger of the opposite hand, (1982a, pp. 159–60). Saxe's diagram (reproduced below) illustrates this procedure and associates each body part with a distinct word. But it is not clear from his description quite what function the words themselves play in the counting activity. He gives a photograph of Oksapmin children doing arithmetic by pointing at parts of their body.

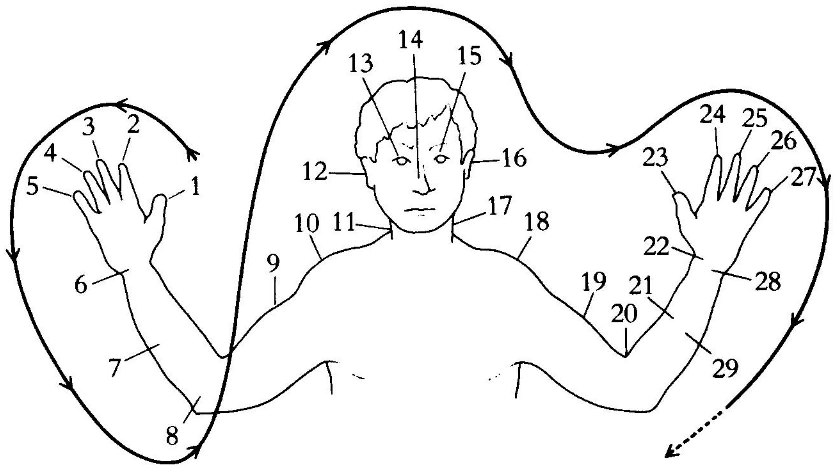


Figure 2.6.1 (1) tip^na, (2) tipnarip, (3) bumrip, (4) h^tdip, (5) h^th^ta, (6) dopa, (7) besa, (8) kir, (9) tow^t, (10) kata, (11) gwer, (12) nata, (13) kina, (14) aruma, (15) tan-kina, (16) tan-nata, (17) tan-gwer, (18) tan-kata, (19) tan-tow^t, (20) tan-kir, (21) tan-besa, (22) tan-dopa, (23) tan-tip^na, (24) tan-tipnarip, (25) tan-bumrip, (26) tan-h^tdip, (27) tan-h^th^ta.

The normal processes of language-change over a long period would lead to words which originally had bodypart associations losing these associations and becoming pure numeral, or at least counting, words. There would tend to be a phonological split, reflecting the clear semantic difference between number and bodypart concepts.

I do not claim that all fully developed numeral systems evolve through all of the stages mentioned above, although it seems plausible that such stages are typically involved in the evolution of numeral systems. It cannot be concluded for certain from the anthropological literature that numeral systems always develop out of recited counting sequences, although it seems very likely that this is the most usual course of development. In the Torres Straits example cited, the counting expressions bear no resemblance to the 'common numerals'. Similarly, the evolutionary relationship between numeral systems and systems of grammatical number is quite unclear from any indications of change in progress that one might hope to glean from the anthropological literature. The remoteness in time and space of the origins of numeral systems, together with the possible effects of cultural and linguistic mixing and borrowing, make the evolutionary picture hard, if not impossible, to discern by any method resembling direct observation. Nevertheless, in the next chapters, psychological and synchronic linguistic considerations will be invoked in support of certain speculations about the evolution of numeral systems.

I assume that there can be an early stage in the development of any numeral system when it has a small lexicon but no syntax internal to the numeral system – no way of putting number words together to form expressions for further numbers. Obviously the development of syntactically complex numerals cannot precede the development of a basic numeral lexicon, and it may lag by some appreciable time. The numeral words in these primitive lexicons would have syntactic properties relating to their distribution in noun phrases and sentences, however.

The linguistic irregularity manifested in agreement patterns and suppletive ordinals which affect the first three or four numeral words in the sequence, but not subsequent ones, can be explained by postulating a primitive stage in languages when they had only three or four numerals. The syntax of these numerals reflects a very ancient characteristic of the languages concerned. When subsequent numerals were invented, they were annexed to the old system, but not constructed to harmonize in style. One can 'read' the history of a system, just like the history of an old building, from the contrasting styles of its pieces, from the foundations up.

The invention of a full range of complex syntactic numerals presumably did not take place at a single stroke. The irregularities involving the the method of signalling addition, 1-deletion, and base-suppletion can all be explained by postulating that systems typically rest for a period in their development at a stage where there is a single syntactic rule for forming (relatively simple) complex numerals, expressing the addition of a 'digit' to the chosen base number. (Just how a base number gets chosen will be dealt with in Chapter 6.) This simple rule would be something like (2.6.2).

2.6.2 Numeral → Base Digit

The constituents need not be in this order, and there may or may not be a morpheme explicitly marking the addition operation. At this stage a (proto-)decimal system will only have expressions for numbers up to 19 or 20. The English *-teen* words would be generated by such a rule. During the static period before the next development in the system, the complex forms produced by the rule become lexicalized to some extent. The recitation of a sequence of numerals in counting would tend to reinforce the impression that these complex numerals are single items. The original psycholinguistic productivity which the rule may be assumed to have had, at least for its inventors, is to some extent lost, in the same way as rules of complex word-formation often lose their productivity. The perceived productivity of the rule is so diminished that it is not taken as a basis for the next round of invention, when the system is extended to higher numbers. Thus, when a protodecimal system is extended beyond 20, the new rules can be genuinely new. Rules for forming new expressions above 20, interpreted by addition, will not necessarily be the same as the old simple additive rule (2.6.2). This accounts for the tendency to irregularity in the method of signalling addition just at the point where a multiplicative rule appears for the first time.

When a multiplicative rule appears for the first time, the usefulness of the new rule is in expressing numbers higher than the base number. A single-word expression already exists for the base number; there is no need to adopt a new expression for it. Thus, although in principle the new rule may make available a new expression, (1 × base), for the base number, a simpler expression already exists, and one would not expect the new form to oust the preexisting simpler form. A generative grammarian looks at a whole system from a synchronic vantage point after its historical development, and all that can be done from this perspective is to describe, by a bit of *ad hoc* machinery, the irregularity captured by a rule of 1-deletion. But looking at a system as an accumulation of augmentations added sporadically in stages over history, the irregularity can be explained.

When multiplicative constructions are first invented, there is no requirement that they name the base number by the word already used to express it in isolation, although obviously the new expressions are more transparent if this is done. The presence of 1-deletion would tend psychologically to distance the base-word used in isolation from the form used in the new multiplicative constructions. Paraphrases, more or less elliptical, could be used. Thus, hypothetically, one might find forms like 'bundles', 'heaps', 'groups' appearing for the base number in the new multiplicative expressions. This is the base-suppletion irregularity.

The implication of this argument is that 'unnatural' systems, which are indeed invented whole at one stroke, will not show the irregularities seen typically in systems which have evolved naturally over centuries, unless their inventors deliberately disguise them to look like natural systems by introducing 'life-like' irregular features. There is an interesting case of a whole numeral system invented all in one piece. Modern Welsh has abandoned the vigesimal system and adopted a wholly decimal system. I have not been able to discover the exact details, but from personal communications with Welsh speakers (mainly Gwen Awbery, curator of dialects at the Welsh Folk Museum, Cardiff, and John Phillips) it seems likely that the new decimal system was fairly deliberately devised specifically to facilitate arithmetic teaching in Welsh-language schools, as an alternative to the old vigesimal system, which did not match up with the Arabic place-value notation. I have heard the Welsh League of Youth (and even the BBC!) named as responsible for devising the modern system used in arithmetic teaching. This new Welsh system shows none of the irregularities discussed above. It uses the existing Welsh words for 1–10, and looks as shown in (2.6.3).

2.6.3

	10 un deg	20 dau deg	90 naw deg
1 un	11 un deg un	21 dau deg un	91 naw deg un
2 dau	12 un deg dau	22 dau deg dau	92 naw deg dau
⋮ ⋮	⋮ ⋮	⋮ ⋮	⋮ ⋮
8 wyth	18 un deg wyth	28 dau deg wyth	98 naw deg wyth
9 naw	19 un deg naw	29 dau deg naw	99 naw deg naw

It is most unusual for any system not to use 1-deletion with the base number itself. This system is perfectly regular and appears to have been put together all at one time, like Esperanto.

Other irregularities in numeral systems can equally plausibly be seen as growth marks. The sporadic incidence of 1-deletion with higher bases (for example, French *cent*, *mille*) can be explained by postulating static periods before the introduction of multiplicative constructions using these higher bases, in the same way as 1-deletion with the original base, typically 10. In many languages there are several places in the numeral sequence where the method of signalling addition changes. For instance, the conjunction *and* in English is only used in numerals involving the addition of a number below 100 to a number above 100. Changes above the point where multiplication first appears are almost exclusively associated with points where a new numeral word, such as *hundred*, is introduced into the system. The Arabic peculiarity of using plural nouns with numerals from 2 to 10, but singular nouns with complex higher-valued numerals is also presumably a growth mark.

Attending to the fact that languages grow diachronically allows one an explanation of irregularities. Common types of irregularity are accounted for by common patterns of growth. This historical view and the synchronic/nativist view complement each other well in explaining complementary aspects of languages, roughly the irregularities and the regularities. This is a somewhat anti-Saussurean point, since Saussure emphasized the primacy of synchrony over diachrony. Some aspects of synchrony can only be explained diachronically.

3

A Continuous Sequence of Counting Words

3.1 Continuity of the Basic Lexical Sequence

Numeral systems express the lowest numbers by purely lexical means and may then resort to syntax to express higher numbers. In the typical case there are single words for all the numbers up to and including 10, after which morphosyntactic combinations are used, although, of course, bases other than 10 do occur. A language may have only a primitive numeral lexicon, giving its speakers the facility to express small numbers, but may lack syntactic resources in the numeral system, so that higher numbers are not expressible. Of such primitive numeral lexicons, and of the sets of lexical items denoting numbers equal to or less than the base number in more developed systems, the following generalization applies almost without exception.

> *3.1.1* For every number up to the first base number in a system with syntactic numeral constructions, or for every number up to the limit of the system in the case of a primitive numeral lexicon, there exists a lexical item expressing that number.

In other words, down at this level, among the relatively small numbers, there are no accidental gaps in the lexicon. This is not an assertion merely about some unexploited 'potential' of a numeral system, apparent to an external analyst, but not realized by speakers of the language concerned; that is it is not an assertion like 'every language has the potential to express modern biological or physical theories – the necessary technical terms can easily be coined or borrowed'.

It is conceivable, although I believe very unlikely, that this continuity is an artefact of methods of data collection. A natural way of asking an informant about his numeral system is to get him to count, that is to list expressions in order of their value. If a gap appears in the sequence, the analyst from a Western European language may be surprised and tempted to urge the informant to fill the gap in some way. In that case, if there were really no expression for the number concerned, the informant could insist on this fact, and the fieldworker would have to record it. In all the major taxonomic and typological surveys of the world's numeral systems (for example Pott, Kluge, Seidenberg, Conant, Lean), I have not come across a single instance of a gap in a numeral sequence being explicitly attested in a synchronic description of a language's numeral system. Naturally, there are sometimes gaps in a reported sequence due to lack of information, but nowhere have I seen it asserted that a word for a number is actually lacking from a lexicon which has an expression for some higher number (other than a multiple of the base number).

In some languages we can find what might be construed as evidence for the existence of a discontinuity in numeral expressions at a historically previous stage in the language. The following Ainu data are from Pott (1847, p. 86):

3.1.2
1 schnepf
2 tup
.
.
.
8 tubischambi (glossed as '2 from 10')
9 schnebischambi (glossed as '1 from 10')
10 wambi

Lean (1985–6, vol. 2) reports, of 25 out of 26 languages of the Manus province (Admiralty Islands) of Papua New Guinea, that 'The words for the numerals 7, 8, and 9 contain, respectively, the words for 3, 2, and 1 thus implying the use of subtraction from 10' (p. 62). In cases like this, where subtraction is used, one may speculate that the expression for 10 existed before the formation of the expressions for 8 and 9, which are bimorphemic and use the word for 10 as a point of reference. If this indeed was the case, that particular stage in the history of these numeral systems was an exception to an otherwise very general pattern. On the other hand, it is possible that the original base number of these systems was 5, and that the expression for 10 was originally formed as a multiple of the expression for 5, in which case these languages provide no evidence against generalization (3.1.1). The remaining Manus province language, Nauna, does indeed form 6, 7, 8, 9 as 5 + 1, 5 + 2, 5 + 3, 5 + 4, as do nine of the 18 languages from the neighbouring New Ireland province (Lean, 1985–6, vol. 1). The other nine New Ireland languages have monomorphemic forms up to 10, in line with generalization (3.1.1).

De Villiers postulates a historical counter-example to generalization (3.1.1). 'It would appear from a comparison of the various Bantu languages that words expressing the first five numerals and 10 must have existed at the time when the Sechuana stock diverged from the others, before any words existed expressing the intermediate numerals 6, 7, 8 and 9' (1923, p. 56). Perhaps a few counter-examples to the continuity of initial lexical numeral sequences do exist. If so, generalization (3.1.1) must be redrawn as a very strong tendency, rather than as an absolute universal. Such a strong tendency clearly still needs explaining, counter-instances notwithstanding. Strong tendencies pose a more complex problem for would-be explainers, as there is the additional onus of attempting to explain how counter-instances to the general trend can come about. But we can be grateful to the counter-instances for evidence that the generalization is in no sense merely a tautology, a logical consequence of our conception of 'numeral word'.

Note that all the apparent counter-instances to (3.1.1) involve a language having a word for 10 without having words for 8 and 9; the salience of the number of digits on the two hands doubtless plays a role here. The denotation of a short continuous sequence of numbers starting with 1 seems very natural for an embryonic numeral system, but it is not the only conceivable possibility. One may at least imagine, in a science-fiction sort of way, a small numeral lexicon with only forms for 7 (the numbers of stars in the Pleiades) and 12 (the number, say, of gods in the pantheon of the tribe concerned). Somewhat less exotically, one might imagine a primitive numeral lexicon denoting just 1, 2 (the number of a person's hands), 4 (the number of a person's limbs), 5 (the digits of one hand), and 10 (the digits of two hands). Tylor gives a hypothetical example of the same kind:

> If we allow ourselves to mix for a moment what is with what might be, we can see how unlimited is the field of possible growth of numerals by mere adoption of the names of familiar things. Following the example of the Sleswigers we might make *shilling* a numeral for 12, and go on to express 4 by *groat* [there were 12 pence in a shilling, and 4 in a groat]; *week* would provide us with a name for 7, and *clover* [normally a 3-leafed plant] for 3. (1891, vol. I, p. 258)

Tylor's examples are not well chosen, as three of them are culturally defined objects, rather than naturally occurring, and it is hard to imagine the cultural objects being defined without prior numeral words. But the fact that such possibilities are theoretically conceivable shows that the continuity is a contingent, rather than a necessary definitional truth about numeral systems.

It is not impossible to know sequences of expressions which have gaps in them in more or less the sense of this discussion. For example, the street I grew up on did not have numbered houses, but rather each house had a name, such as *Rosegarth*, *Allwynds*, and so on. As a child, I made an effort to learn the names of all the houses in sequence (about 30 in all). There were times when I knew subparts of the sequence of names, such as those at the beginning and the names of houses near my own house, but did not know the names of some intermediate houses. I certainly knew each member of the sequence of referents (houses), but for some of them I did not have a name. Why are not some numeral systems, at least, like this, with a continuous sequence of potential referents (the numbers), but gaps in the sequence of expressions where particular middle-valued numbers have no names?

One can even, in fact, easily imagine a second language learner knowing only a discontinuous sequence of low-valued numeral words in the second language, due to imperfect learning. For example 'eins, zwei, drei, er umm I've forgotten the word for *four*, fünf, sechs, er umm ... acht, neun, zehn'. Gelman and Gallistel (1978, p. 91) report cases of 2½-year-old children counting with a stable, but incorrect, sequence of number words, such as *one*, *two*, *six*. Obviously, one must not jump to the conclusion that such children acquire the word *six*, with the meaning 6, before they have acquired *three*/3, *four*/4, and *five*/5, but such evidence deserves to be considered further.

I discuss below three hypotheses which might be entertained to explain the fact that a word for a number is not acquired before the acquisition of a word for its predecessor. These hypotheses, respectively the 'Referential/Pragmatic', the 'Conceptual/Verbal', and the 'Ritual' hypotheses are all deliberately drawn rather starkly; they do not correspond exactly to extant philosophical '-isms'. Each hypothesis, however, elevates to primacy one of the three points of the 'triangle of reference' of Ogden and Richards (1949, p. 11), given in (3.1.3).

3.1.3 thought or reference

/ \

symbol referent

The Referential/Pragmatic hypothesis essentially makes the existence of things in the world (referents) the factor determining the continuity of numbers; the Conceptual/Verbal hypothesis makes the availability of concepts (roughly thoughts or ideas) of numbers the primary determining factor; and the Ritual hypothesis derives the continuity from the prior existence of a series of symbols. The three hypotheses are discussed specifically in relation to the continuity of simple numerals. Where there are significant similarities between these hypotheses and familiar philosophical positions, these will be pointed out. All three hypotheses can be thought of either in the context of ordinary language acquisition or of genuine linguistic invention, although the first, the Referential/Pragmatic hypothesis, relates somewhat more naturally to invention than to ordinary acquisition.

Progress in understanding a problem can often be facilitated by drawing up sharp dichotomies (or trichotomies), such as empiricist/nativist, or nominalist/conceptualist/realist. Such simple polarized positions have a methodological usefulness, but it is not to be expected that any one of them is simply right, while its rivals are simply wrong. Further progress is made by articulating more subtle and complex positions, oriented in various ways around and between the simple polar extremes. Some synthesis of the three hypotheses presented in this chapter is probably correct, as will be argued in the last sections (3.5, 3.6).

3.2 The Referential/Pragmatic Hypothesis

The Referential/Pragmatic hypothesis, a version of traditional empiricism, is that numeral words originate out of a communicative need for expressions applicable to collections of things. Just as the predicate *red* describes a property, 'redness', of an object, so, for instance, the word *two* describes the cardinality 'twoness' possessed by a pair of things. And similarly for 'threeness', 'fourness', and so on. This hypothesis assumes that such cardinalities of collections of things are evident fairly directly to humans and that real-world factors determine the selection of numbers denoted by a primitive numeral lexicon. Such factors would be of the kind involved in standard examples such as the large range of Arabic words denoting camels, or Eskimo words for different kinds of snow. An account in such pragmatic terms would claim that lower-valued number-words are invented and occur in more of these rudimentary numeral lexicons precisely because they are the ones humans need to use most. This would amount to a universal claim that in practical human affairs the need to refer to some specific low number n is likely to arise more often than a need to refer to its successor $n + 1$. There is actually some evidence for this.

Columns L and S in (3.2.1) give the frequencies with which English numeral words from *two* to *nine* occurred in two separate word-counts reported in Thorndike and Lorge (1944); column AHD gives the rank order of frequency reported for these words in the *American Heritage Dictionary*.

3.2.1

	L	S	AHD
two	5958	6195	64
three	2673	2972	126
four	1637	?	219
five	1462	?	295
six	806	978	393
seven	615	570	722
eight	657	402	755
nine	468	294	1116

Figures in column L are from a count of over 4.5 million words from 'recent [pre-1944] and popular magazines'. Figures in column S are from a count of over 4.5 million words from 'a miscellany of juvenile and adult reading'. The two queries in column S represent unrecorded information. The rank orders in column AHD are from a sample of 6 million words. All three columns show a steady decrease in frequency from *two* to *nine*: the exception to this trend between *seven* and *eight* in column L is presumably an insignificant anomaly. I take these figures to indicate that in these extensive and diverse corpora the need to mention lower cardinalities occurred more often than the need to mention higher cardinalities. No other explanation suggests itself.

Furthermore, one can think of many everyday situations in which the difference between a low number and its successor is more important than the difference between a high number and its successor. When baking a small cake, the difference between 2 eggs and 3 eggs is more crucial to the outcome than is the difference between 8 eggs and 9 eggs when baking a larger cake. Thus, presumably, words denoting exact low numbers are more useful than words denoting exact high numbers. Beyond quite low numbers, such as 9, however, the hypothesized pragmatic effect loses its plausibility. The pragmatic/referential consideration, the idea that worldly necessity is the mother of linguistic invention, may make some contribution to an explanation of the fact that primitive numeral lexicons denote short continuous sequences of numbers starting at 1. It is certainly not a countervailing factor.

The Referential/Pragmatic hypothesis, as it is intended here, is mildly teleological. It sees the goal of referring to a particular cardinality as pre-existing the means to refer to it, and in some sense causing the means to be invented. I say 'mildly' teleological, because the process can presumably be relatively informal and haphazard, and the success of the newly coined term in referring to the intended cardinality lies partially in the chance of whether the hearers take it as so referring.

The general idea of a pragmatic/referential determinant can be given a somewhat different emphasis, concentrating less on the notion of usefulness and correspondingly more on that of salience. It might be argued that 'oneness' in the world is more salient to humans than 'twoness', and 'twoness' more salient than 'threeness', and so on. Certainly, the relative perceptual salience of entities, properties, and so on must play a part in the structuring of other areas of the lexicon. Colour terms provide the best known example. But, although the idea does not seem so promising in the area of numeral words, it might be plausible for the extremely low numbers, up to about 3. The property of being a physical object is clearly salient for humans, and the notion of an object might be linked with (or actually subsume) the notion of 'oneness' or singularity. In this sense, oneness is more salient than plurality. In some similar way 'twoness' might be more salient than 'threeness', but it is not very plausible to claim greater salience for, say, 'eightness' over 'nineness'. And it is very likely that 'tenness', the property associated with the set of digits on both hands, is actually more salient than, say 'eightness' or 'nineness'.

It seems that J. S. Mill may have believed in something like the Referential/Pragmatic hypothesis:

> And although a hundred and two horses are not so easily distinguished from a hundred and three, as two horses are from three – though in most positions the senses do not perceive any difference – yet they may be so placed that a difference will be perceptible, or *else we should never have distinguished them, and given them different names.* (1906, p. 400, emphasis added)

Any version of the Referential/Pragmatic hypothesis has a clear weakness in that it assumes humans can recognize the cardinality of small collections of things in some sense directly, without reference to other (lower) cardinalities. That is, this hypothesis, as stated, treats the recognition of the cardinalities of sets rather like the recognition of individual cats, dogs, or types of camels as such. Gelman and Gallistel discuss in detail the common belief that for small numbers of items in a collection there is recognition 'by means of a direct perceptual-apprehension mechanism, sometimes referred to as *subitizing*', (1978, p. 64). They write: 'The process is generally thought to be simple: each numerosity is grasped, apprehended, taken in as a whole, seen as a pattern. The idea is that there exist pattern recognizers that detect oneness, twoness, threeness, and so forth. ... *Twoness* and *threeness* are considered percepts like *cowness* and *treeness*' (p. 65). After discussion and consideration of the literature of psychological experiments, Gelman and Gallistel's own conclusion is that subitizing is unlikely to be a basic human ability, independent of counting. Although adults do subitize for low numbers, Gelman and Gallistel conclude that this is an advanced skill acquired after learning to count. My own interpretation of the literature, including some very suggestive studies post-dating Gelman and Gallistel's work, would not dismiss a pre-counting ability to subitize, but would retain it for extremely low cardinalities, such as 2 and 3.

To start with a study in the postnatal ward itself, Antell and Keating (1983) showed that babies in the first week of life were able to discriminate between an array of 2 dots and an array of 3 dots. The baby subjects responded to a difference in stimulus by staring longer at a stimulus card. Reasonable controls were used to eliminate the possibility of response to length of line and spacing between the dots. Although the babies discriminated between 2 and 3 dots, they did not discriminate between 4 and 6 dots. Antell and Keating's work replicated for 1-week-olds the results achieved for 22-week-olds within an identical experimental paradigm by Starkey and Cooper (1980). These investigators also found a discrimination between 2 and 3 dots, but none between 4 and 6. The significance of these studies is that they were carried out on indisputably prelinguistic subjects. Of course, the internal mechanism responsible for the infants' discriminations could in some sense be *analogous* to externalized verbal counting, but clearly it is not the same thing.

Russac (1983) experimented with 2-year-olds and found that most were capable of learning to discriminate betwen numbers of dots in the range 1–4, although their verbal counting abilities were not well enough developed to help them in their responses. The subjects presumably subitized, but it was noticeable that when the experimenter tried to elicit specifically verbal counting behaviour, subitization was not apparent. Schaeffer et al. (1974) propose a model (based on experimental data) of the development of number skills; this model 'posits the hierarchic integration of six number skills', one of which is 'pattern recognition of small numbers'. Mandler and Shebo (1982) conclude from their experiments with adults that there is a 'response to arrays of 1 to 3 that is fast and accurate and is based on acquired patterns' (for example line for 2, triangle for 3).

These studies of subitization clearly only relate to very low cardinalities, up to about 3. The Referential/Pragmatic Hypothesis, which presupposes an ability to perceive cardinality directly, loses plausibility for numbers above this range. One is thus led to consider that conceptual/intellectual factors must also play a part, and probably the major part, in explaining the continuity of basic numeral lexicons. Even if human communities generally are likely to *need* to refer to the number 6 more urgently than to the number 7, it is also plainly true that humans are likely to *be able* to refer to the number 6 before they can refer to the number 7, and for reasons untouched by the Referential/Pragmatic Hypothesis. The Conceptual/Verbal and Ritual Hypotheses, described next, take up this point.

3.3 The Conceptual/Verbal Hypothesis

The Conceptual/Verbal Hypothesis is that humans innately possess (or are innately equipped to acquire on exposure to a normal human environment) the concept of the number 1 and enough further conceptual apparatus to construct (if somehow prompted, and subject to the limitation described below, involving the availability of words) a mental representation, or concept, of the successor number to any number for which they already possess a concept.

A complete account of all that is essential in the mathematical concept of natural number requires Peano's five axioms, given below:

3.3.1

1 Zero is a natural number.
2 The immediate successor of any natural number is a natural number.
3 Distinct natural numbers never have the same immediate successor.
4 Zero is not the immediate successor of any natural number.
5 If something holds true of zero, and if, whenever it holds true of a natural number, it also holds true of the immediate successor of that natural number, then it holds true of all natural numbers.

(from Barker, 1964, p. 58)

To account for ordinary knowledge of number, however, something slightly different from Peano's first axiom will be adopted. The ordinary conception of number, as reflected in natural language numeral systems, does not include zero. Asked to give the smallest number, a person without mathematical schooling will answer 'one'. In the great majority of languages (Chinese being the conspicuous exception) no word for zero figures in the formation of higher numeral expressions. The English words *nought* and *zero* are by no means as felicitously used to modify nouns as *two*, *three*, and so on. 'We have nought bananas' is odd, as is 'We have zero bananas'. And, interestingly, one does not meet the suggestion that the word *no*, which is acceptable in this context, is a numeral. So a natural version of Peano's first axiom would state that 1 is a natural number.

Although in an abstract, theoretical sense (that is given infinite computing time), the number 1 and the successor function will yield all positive integers, no finite organism can literally possess separate mental representations of each individual number. The Conceptual/Verbal Hypothesis is that the limitations on what numbers are actually mentally represented are severe; let us say for present purposes that (except perhaps in rare cases of genuine invention) number concepts (mental representations) can only be reached via verbal descriptions, that is through descriptions couched in words already possessed. So, for example, given the words *number*, *after*, and *one*, together with representations of their meanings, and some suitable syntax, an individual can construct the phrase *the number after one*, and the interpretation of this phrase will be new in the sense of not having been previously constructed in the mind of that individual. According to this hypothesis, then, a syntactically complex mental representation (concept) of a number can be generated without there being a pre-existing word for that number. I assume a further psycholinguistic constraint to the effect that complex phrases with iterated instances of the same operator, such as *the number after the number after one*, are especially hard to process.

Given such assumptions about how concepts for numbers can become present to the mind, and given an initial state in which the concepts of number, 1, and successor are possessed, along with the words for them, it follows that concepts for particular numbers can only become available to the mind in strict succession, and must be accompanied 'one step behind', as it were, by the invention of number words. In other words, using English words to make the example concrete, the concept of 2 is available through the phrase *the number after one*, but, because of the difficulty of processing the more complex *the number after the number after one*, the concept of the number 3 is not available until the concept 2 has been assigned a single word. At this stage, the phrase *the number after two* can be constructed, giving access to the concept 3, which can then be assigned a word, and so on. The Conceptual/Verbal Hypothesis is a story of a continuous sequence of meanings being constructed in a person's mind, provoking a search for a corresponding sequence of words.

The Conceptual/Verbal Hypothesis is a case, specific to

numerals, of a general proposal discussed in detail by Fodor (1976). Fodor postulates a 'language of thought', an internal language, and asks how this language might differ from an 'external' language (English, for example). He suggests that the internal language might in fact be hardly less economical in its vocabulary than the external language, there being often a counterpart in the internal language to a word in the external language. One may think of this as a word/concept pairing, or a Saussurean sign, with its two halves, the *signifiant* and the *signifié* (not that Fodor makes such an allusion). The nub of Fodor's proposal is:

> I think the following is a serious possibility: *bachelor* [the concept] gets into the internal language *as an abbreviation for a complex expression of the internal language*: viz., as an abbreviation for *unmarried man*. The abbreviatory convention is stored as a principle of logic (i.e. as *bachelor* ⇔ *unmarried man*) [a meaning postulate]. Since, in the course of learning English, 'bachelor' gets hooked onto *bachelor* [i.e. the word gets hooked into the concept] and 'unmarried man' gets hooked onto *unmarried man*, *bachelor* ⇔ *unmarried man* can be used to mediate such inferential relations as the one between '*x* is a bachelor' and '*x* is an unmarried man'.
>
> ... On the present model we would expect (a) that there *won't* be a correlation between the definitional complexity of a term and the difficulty of understanding a sentence which contains the term (see above); but (b) in certain cases there *will* be a correspondence between the relative definitional complexity of a pair of terms and the order in which they are learned. Since we are now supposing that the process of definition is, as it were, ontogenetically real, we would expect that the child should master terms corresponding to the *definiens* before he masters terms corresponding to the *definiendum*. If, e.g. *only* is defined in terms of *all*, we would expect 'all' to be learned before 'only'. Which, in fact, it is. (1976, pp. 152–3, italics in original)

In the case of numerals, and in Fodor's terms and notation, the Conceptual/Verbal Hypothesis is that the child is born with the concepts *one*, *number*, and *successor* (or *after*). At an early stage the words 'one', 'number', and 'after' are hooked onto these concepts, thus licensing the hooking of the complex (internal) expression *the number after one* onto the English expression 'the number after one'. *Two* gets into the internal language as an abbreviation for *the number after one*, and *two* gets hooked onto the English word 'two'. The advantage to the organism of having *two* in the internal language, in addition to *the number after one*, is that comprehension of (that is accessing of the internal expression corresponding to) 'two' is simple and fast. Generating correct inferences from an expression containing 'two' via the meaning postulate for the internal language *two* ⇔ *the number after one* may take longer, and may take place 'off-line', that is not during the necessarily rapid act of intake and comprehension of an utterance.

In the case of numerals, this approach seems especially plausible. A hearer can, in the appropriate rapid and superficial sense, comprehend the utterance 'I have six brothers and sisters' before, and even without, going on to make relevant inferences or ponder relevant questions, such as 'The speaker has more than five brothers and sisters' or 'How many of each?'. For numerals, Fodor's 'language of thought' seems to make a more plausible suggestion about language comprehension than an alternative 'mental models' approach, such as Johnson-Laird's (1983). And the language of thought idea provides an articulated theoretical environment for the Conceptual/Verbal Hypothesis, which attributes the universal continuity of basic numeral lexicons to constraints on the sequence in which word/concept pairs can be acquired.

Some empirical support exists for ideas like Fodor's on the similarity between the internal representation of a concept and the external language in which it is expressed. Goodglass, in a survey article on aphasia, mentions 'a concurrence of findings that arousal of a conceptual representation is not an all or none process (contrary to subjective experience) and that its status is related to success in naming' (1983, p. 135)

Locke clearly believed in the necessity of verbal names complementing the abstract ideas for a grasp of number. The quotation below is from one of two sections of his *Essay* subtitled 'Names necessary to Numbers'.

> For he that will count Twenty, or have any *Idea* of that Number, must know that *Nineteen* went before, with the distinct Name or Sign of every one of them, as they stand marked in their order; for where-ever this fails, a gap is made, the Chain breaks, and the Progress in numbering can go no farther. So that *to reckon right, it is required*, 1. That the Mind distinguish carefully two *Ideas*, which are different one from another only by the addition or subtraction of one Unite. 2. That it retain in Memory the Names, or Marks, of the several Combinations from an Unite to that Number; and that not confusedly, and at random, but in that exact order that the numbers follow one another: in either of which, if it trips, the business of numbering will be disturbed, and there will remain only the confused *Idea* of multitude, but the *Ideas* necessary to distinct numeration, will not be attained to. (1975, p. 208)

Beside the emphasis on the joint importance of '*Ideas*' and 'Names', Locke also attaches importance to the notion of an ordered sequence of words, which must also be retained in memory. Rote memorization of a sequence of words is not intrinsic to the Conceptual/Verbal Hypothesis discussed here, but will play a part in the next hypothesis to be discussed, the 'Ritual Hypothesis'.

The intuitionist school of thought on the foundations of mathematics – associated chiefly with Brouwer – emphasizes that mathematical entities are constructed. In the Conceptual/Verbal Hypothesis developed here, there is a sense in which the concepts of numbers are also said to be constructed, but there are in fact several important differences between intuitionism and the hypothesis sketched here. Firstly, for the intuitionist, the natural numbers are not themselves constructed, but are given in intuition; other mathematical entities are constructed by the mind using the intuitively given numbers as building blocks. Secondly, in the Conceptual/Verbal Hypothesis, language plays a crucial part in the mind's construction of new concepts, whereas for the intuitionists 'neither the ordinary language nor any symbolic language can have any other role than that of serving as a nonmathematical auxiliary, to assist the mathematical memory or to enable different individuals to build up the same set' (Brouwer, 1913-14, quoted by Wilder, 1968, p. 247). The absence of a role for language in the Intuitionists' account of the construction of mathematical objects is extensively and, I believe, persuasively argued against by Popper (1972, pp. 128–40).

With the origin of words and concepts, there is clearly a chicken-and-egg problem. Which comes first, the concept of 2, or the linguistic expression *the number after one*? Given the idea of the invention of bits of a linguistic system (including its meanings) by some individuals, together with the, more passive, adoption of inventions by other individuals, we can answer 'Both!'. Starting with a community where all individuals had the word/concept pair *one*/1, and no further numerical word/concept pairs, and signs corresponding to number and successor, an inventive individual might somehow conceive for the first time in the history of the community of the number 2, and then realize that this can be expressed as *the number after one*. He might then utter this expression, and a hearer would experience the expression before decoding it, according to the syntactico-semantic rules of the language, as the number 2. In this case, the concept comes before the words in the inventor's mind, whereas the words precede the concept in the non-inventor's mind. But the concept need not always precede the words in the inventor's mind, and perhaps normally does not. He might be playing with words, constructing expressions in an experimental way, perhaps as a way of discovering new concepts. One can see the logical paradoxes and imaginary numbers as examples of the formation of syntactically well-formed expressions such as *the set of all sets which are not members of themselves* or *the square root of minus one*, whose interpretations in a model would be impossible to conceptualize. (To return to the chicken-and-egg metaphor, these might be like eggs which are laid but do not hatch.)

Now, if anything like the above, 'Conceptual/Verbal' account were true, one would expect some traces in natural language expressions for 2 of the syntactically complex expression *the number after one*. Obviously, in languages which have had a numeral system for a long time, it would be reasonable to allow that any original syntactic complexity in the expression for 2 is likely to have been obliterated by language change. But if there were truth in the idea, one would expect there to be *some* trace of syntactic complexity in an expression for 2 somewhere in the world, for example in one of the very rudimentary numeral systems. But, to my knowledge, there is no such trace. There are binary systems, like *one*, *two*, *two-one*, *two-two*, *two-two-one*, in which the expressions for 3 and above are syntactically complex, but there is no hint in any spoken language of the number 2 being conceived of as some function of the number 1. In Fodor's example, the simple expression *bachelor* is introduced as an abbreviation for the pre-existing complex expression

unmarried man. If, similarly, *two* is introduced as an abbreviation for *the number after one*, we must ask why we do not find evidence of any system in the state just preceding the introduction of *two*, that is with *the number after one* or some clear equivalent used systematically to denote 2, but without (yet) a single word equivalent of *two*.

In written language, one does find ways of representing numbers which might possibly prompt one to argue that an expression for 2 is syntactically complex, or at least reflects the fact that 2 is a function of 1. Roman numerals would be the best known example. 1 = I, 2 = II, 3 = III. Some types of Egyptian hieroglyphics actually took this principle much further, with sometimes ten upright bars inscribed to represent 10 (Menninger, 1969, p. 42, and Seidenberg, 1960, p. 233, citing Lepsius, 1865, table following p. 64). In such a system, one might argue that '//', the expression for 2, was syntactically complex, the written equivalent of *one-and-one*, thus reflecting the intuition that 2 is the successor number to 1. If interpreted loosely enough, one can certainly find some truth in this proposal, but I wish to argue that such written representations do not support the specific Conceptual/Verbal Hypothesis under discussion here, namely that the continuity of basic numeral lexicons in spoken language arises from the number sequence being constructed, through language, from principles encapsulating the essentials of Peano's axioms.

It is noteworthy that iconic representations such as the Roman and Egyptian hieroglyphics occur only in written language, and never, apparently, in spoken language. We know that the first ten written numerals in the Roman system in no way mirrored the corresponding words in spoken Latin. Thus 2 is not **unus unus* and 3 is not **unus unus unus*. It seems reasonable to regard these Roman and Egyptian expressions as incorporating into writing a marginally prelinguistic method of representing the cardinality of collections of things. Thus, where a speaker of a language with no numeral system might hold up a bunch of sticks and say, 'I've got this many pigs', the Egyptian hieroglyphic artist would carve as many sticks as there were objects he wished to record the number of. Interestingly, however, this method of representation is already somewhat linguistic (ideographic as opposed to purely pictographic) in that the carver carves a *number* of sticks accompanied by just *one* symbol for the object enumerated. The collection of objects being recorded is thus analysed into two concepts, the number (represented by the sticks) and the type of the object making up the collection, in a kind of modifier-noun construction. But, for low numbers for which a single word exists, the internal structure of the modifier, the written numeral – for example Roman III – is not a reflection of any structure in the spoken word.

Representing the successor function as *s*, the essence of the Conceptual/Verbal Hypothesis is that an expression for 2 arises as *s*(1), for 3 as *s*(2), for 4 as *s*(3), and so on. In spoken natural language numeral systems, there is no trace of any sequence with this syntactic pattern, and in fact hardly any trace of the number 2, the crucial first step, ever being syntactically complex in any way at all. Indeed the very fact that numeral systems always express the first few numerals, sometimes as far as 10, though sometimes only as far as 2, by lexical means, can be taken to indicate that these first number concepts are not psychologically complex, that is not formed by syntactically assembling more basic concepts, such as '1' and 'successor'. Obviously this conclusion only stands for a sense of psychological complexity closely parallel to linguistic complexity. In many ways, individual lexical items are clearly semantically complex, even though morphologically simple, a fact which formed the basis for componential analysis. But the fact that 2 is never expressed as *one-and-one* and 3 is is only rarely expressed as *two-and-one*, whereas 11 is very commonly expressed as *ten-and-one*, suggests that at the very bottom of the number sequence, number concepts are not constructed syntactically (that is by syntactic operations on the 'language of thought'). Basic numeral lexicons resemble in some ways other restricted mutual antonymy sets, such as the set of names for days of the week, (*Monday*, *Tuesday*, ...), months (*January*, ...), points of the compass (*North*, *South*, *East*, *West*). Many such lexical sets are neatly structured in ways that make componential analysis possible, but usually such structure is not particularly transparent in the morphology of the terms.

3.4 The Ritual Hypothesis

The Ritual (or 'Eeny, meeny, miny, mo') Hypothesis is that at a stage before the development of proper numeral words, rituals exist in which sequences of words which have no referential, propositional, or conceptual meaning are recited while the human actor simultaneously points (in some way) to objects in a collection, often pointing to all members of the collection, and never pointing to the same object twice. Children, when they have finished a bowl of plums and put the plumstones on the side of their dish, sometimes recite in this way, pointing to each stone in turn, 'Tinker, tailor, soldier, sailor, rich man, poor man, beggar man, thief, (doctor, lawyer, Indian Chief)'. In a similar, though not identical, ritual [known in Britain as 'dipping' – see Opie and Opie (1959) for many examples] children select a player to play some special role in a game, such as the role of seeker in hide-and-seek. Here there is the same component of reciting a sequence of words (often nonsense words such as *eeny, meeny, miny, mo*) in a conventional order, each word accompanied by pointing to a distinct player. Only when all players have been pointed to are any players pointed to a second time. The difference between this and the *tinker, tailor* ritual is that in *eeny, meeny, miny, mo* the ritual ends when the word sequence is exhausted, whereas in *tinker, tailor* the ritual ends when the collection of objects is exhausted.

Both *tinker, tailor* and *eeny, meeny* rituals serve some practical purpose of the reciter. The *tinker, tailor* ritual may be said to serve the purpose, in the child's make-believe world, of divining the future. Obviously, it is not a successful means to this end, but that is how it is interpreted. Coming to the end of the plumstones on *poor man*, a child will say, half seriously, 'Oh, I'm going to be a poor man', and perhaps take another plum if he does not fancy that particular prognostication. In modern societies such rituals for divining the future are not taken seriously by most adults, but the naturalness and degree of seriousness with which they are taken by children is remarkable.

The purpose of *eeny, meeny, miny, mo* rituals is to select a member of a group at random. Such rituals are generally effective and robust enough to produce a result regarded as satisfactorily random by the users. (With small numbers of players, an experienced dipper can manipulate an *eeny, meeny, miny, mo* ritual to produce a desired result by selecting the right place to start.) An *eeny, meeny* ritual actually models exactly a type of simple procedure commonly used for generating random numbers by computers (for example as mentioned in Clocksin and Mellish (1981, pp. 148–9)). In such a procedure, given the task of generating a 'random' number between 1 and *n*, a 'seed' number (perhaps taken from the computer's clock at the time of the computation) is divided by *n* and the remainder is taken to be the random number generated. In a dipping ritual, *n* is the number of players, and the seed number is the number of words in the ritual sequence. Going round the circle of players, reciting one word per player, in effect divides the number of words in the sequence by the number of players. Any remainder of the division is represented by the number of places between where the ritual started and where it ends. The parallel is exact.

Thus the *eeny, meeny, miny, mo* dipping ritual can be given an arithmetical interpretation, even though the people who use it (children) are generally quite unaware of any arithmetical connections. The ritual serves a practical purpose simply and well. Since recitation of the sequence must always go the full length, none of the intermediate words have any significance in terms of the ritual. That is it doesn't matter, for instance, who is pointed to on *miny*. In the nature of this ritual, only the last word is in a position to be assigned any special significance.

The Ritual Hypothesis being put forward for examination here is that numeral systems arose out of counting, developed as a method of achieving a practical purpose simply and reliably, using a conventional sequence of recited words. The purpose involved is the comparison of collections of items in terms of the numbers of their members. Perhaps it is possible to conceive of this ritual being used to ascertain parity of different collections without the involvement of concepts of particular numbers, just as a linguist may be prepared to say that two expressions 'have the same meaning', without in any way being prepared, or able, to say what that meaning *is*. I may 'measure' the width of an alcove to see if a piece of furniture will fit into it by holding a string across it (and then across the piece of furniture) without any value on an absolute scale of inches ever actually entering into the calculation. Similarly, one could tell whether two collections are equally numerous by reciting a conventional sequence of words against the members of each in turn, and seeing whether both recitations finish at the same word, without necessarily interpreting that word as expressing or representing the (concept of the) number involved. (Wittgenstein, 1974, pp. 351–358, argues that something of this kind is possible.)

The sequences of words used in such rituals would *become* interpreted numerically, and it is clear from the nature of the ritual that the most natural development would be for each successive word to become associated with each successive

number, without gaps. There would be no gaps, because there is no sense in which the original uninterpreted sequence of expressions *could have* gaps; each word in the sequence, except the last, would be followed by the next word in the sequence, of course. There might be pauses as the sequence is recited, but not gaps. Before assigning interpretations to words in a ritual sequence like this, the only information a speaker would have about any particular word would be of its position in relation to its neighbours in the sequence. 'The moment any series of names is arranged in regular order in our minds, it becomes a counting-machine' (Tylor, 1891, vol. I, p. 258) This is, then, the story of an established sequence of expressions in search of a sequence of meanings, and given the nature of the ritual, the number sequence is the best, possibly the only, candidate.

A picture similar to the Ritual Hypothesis is given by Benacerraf.

> There are two kinds of counting, corresponding to transitive and intransitive uses of the verb 'to count'. In one, 'counting' admits of a direct object, as in 'counting the marbles'; in the other it does not. The case I have in mind is that of the preoperative patient being prepared for the operating room. The ether mask is placed over his face and he is told to count, as far as he can. He has not been instructed to count anything at all. He has merely been told to count. ... It seems, therefore, that it is possible for someone to learn to count intransitively, without learning to count transitively. But not vice versa. This is, I think, a mildly significant point. (1965, pp. 49–50)

Benacerraf's position is linked with a restrained form of anti-number-realism.

> On this view the sequence of number words is just that – a sequence of words or expressions with certain properties. There are not two kinds of things, numbers and number words, but just one, the words themselves. Most languages contain such a sequence and any such sequence (of words or terms) will serve the purposes for which we have ours, provided it is recursive in the relevant respect. In counting, we do not correlate sets with initial segments of the numbers as extralinguistic entities, but correlate sets with initial

> segments of the sequence of number *words*. The central idea is that this recursive sequence is a sort of yardstick which we use to measure sets. Questions of the identification of the referents of number words should be dismissed as misguided in just the way that a question about the referents of the parts of a ruler would be seen as misguided. (p. 71)

The Ritual Hypothesis discussed here is, however, compatible with number-realism; all it claims is that the acquisition of knowledge of numbers proceeds via recitation of a sequence of initially non-denoting words.

The Ritual Hypothesis is that a particular property of basic numeral lexicons, the continuity of their sets of referents, is to be attributed to the development of such lexicons from an *activity*. That is, according to this hypothesis, a piece of conventional behaviour, counting activity, gave rise to an abstract system in which the words of the original sequence could be conceived of independently of the sequence. This would be the beginning of the route leading to the full integration of a numeral system into the rest of a language. The Ritual Hypothesis can be interpreted as a hypothesis either about the original invention of counting and numeral systems or about the ordinary acquisition by any child of, first, the ability to count, and then a numeral system. But in either case, whether invention or ordinary acquisition, although tracing numerals back to a practical activity, the Ritual Hypothesis makes a claim about the way in which human psychology comes to grips with numbers and numerals. It is clear, from both anthropological and psychological sources, that instances are common in which children or whole communities have mastered a counting activity without a full command of a numeral system integrated into everyday language.

The Ritual Hypothesis emphasizes the non-conceptual side of verbal activity with numeral words. Evidence from aphasics shows that the ability to recite sequences of words, such as the counting sequence, or the alphabet, or the days of the week, can remain unimpaired in patients who otherwise have severe difficulties in finding words for concepts. One such case is mentioned in Miller (1951, p. 244) citing Weisenberg and McBride (1935, p. 302).

In a book summarizing a large body of experimental research into children's understanding of number, Gelman and Gallistel (1978) conclude that children possess a mental 'scheme' which

enables them to learn to count, and against which they (unconsciously, of course) judge their own developing performance. Very interestingly, some children adopt stable, non-standard sequences of words for counting with. 'Some of these children produce lists that they are unlikely ever to have heard used for enumeration. An occasional child uses the alphabet, which is a series that they have indeed heard but not in the context of counting' (1978, p. 204). Tylor anecdotally testifies to the same phenomenon of children's use of nonstandard word sequences for counting: 'I have read of a little girl who was set to count cards, and she counted them accordingly, January, February, March, April. She might, of course, have reckoned them as Monday, Tuesday, Wednesday' (1891, vol. I, p. 258). Gelman and Gallistel break down the activity of counting into a number of 'principles', such as (1) using stable sequences of words, (2) associating each word in the sequence with one, and only one, of the items in the set being counted, and (3) attaching some special significance to the last word reached when all the items have been 'tagged'.

> The significant fact about these [idiosyncratic, nonstandard] lists is that they are used in a way that is prescribed by the counting principles. It seems reasonable to conclude that the availability of the principles governs such behavior. Any other conclusion would require postulating the existence of systematic behavior that resembles counting and occurs by chance. Granted, when we first encountered such behaviors we thought them random uses of number words and the alphabet. But when we subjected them to analyses suggested by the counting principles, we discovered that such children were telling us, in their own way, what they knew about counting. (1978, p. 204)

Gelman and Gallistel's cognitive scheme, with which, they propose, young children are equipped in advance of their learning to count, avoids the use of such expressions as 'the concept of the number 1' and 'the concept of the successor function'. It may thus seem closer to the Ritual Hypothesis than to the Conceptual/ Verbal Hypothesis, but, in the light of Gelman and Gallistel's work, I do not wish to drive too thick a wedge between these two hypotheses. What they have in common is an appeal to specific cognitive structure possessed by the child in advance of his learning a numeral system, a cognitive structure which guarantees, in either case, that the sequence of numeral expressions learned will correspond to a continuous sequence of numbers, without any gaps.

Gelman and Gallistel do not, unfortunately, report exactly how many children in their sample used unconventional sequences of words (such as the alphabet) for counting, although there are many implications throughout the book that a significant proportion of children used such idiosyncratic (but stable) sequences of their own. Obviously, often these sequences were quite short, but the point is that they were used *consistently* by their users. Gelman and Gallistel do not entertain the possibility that there may be differences in the underlying cognitive structures which children possess before they start to learn to count. As is usual with such studies, they assume homogeneity of the species and propose general conclusions, compatible with all their data. For present purposes, however, I want to continue to explore an alternative possibility, namely that there are two sorts of language acquirers (that is human beings): the inventors and the non-inventors.

Gelman and Gallistel write:

> we find two sorts of children – especially in the youngest age group. One sort of child counts in the conventional way. A two-item array yields a 'one, two; two' answer; a three-item array yields a 'one, two, three; three' answer. We cannot and do not rest our case on such evidence. One could account for such data by simply assuming that the child who counts 'one, two' or 'one, two, three' has committed these words to rote memory. One could assume that the memorizing of the conventional list in the conventional order precedes the induction of counting principles. In other words, one could argue that skill in reciting count-word sequences precedes and forms a basis for the induction of counting principles. We, however, advance the opposite thesis: A knowledge of counting principles forms the basis for the acquisition of counting skill. (1978, p. 204)

The first kind of child, in my terms the non-inventor, acquires a simple counting sequence in a way compatible with an empiricist theory of language acquisition. For these children, during the early stages, their acquisition of the numeral system could be like

the acquisition of a piece of an oral tradition, like reciting a nursery rhyme, or the Lord's Prayer. Reciting the numeral sequence, perhaps with the appropriate pointing gestures, could be just something that older speakers get them to do by rote. It comes naturally only to the extent that they can actually manage to do it; it does not come spontaneously, without some coaching by elders.

If Gelman and Gallistel's proposed cognitive scheme were present in every human being in a strong enough form to make counting behaviour just bubble up spontaneously, using whatever (conventional or unconventional) sequence of words happened to be available, then every human community would possess a numeral system (barring wierd circumstances, like the natural urge to count being repressed by some religious taboo, or other extraneous factors). But not every human community does have a numeral system. Members of communities without numeral systems have no trouble learning a numeral system, but, apparently, no member of such a community has actually invented a numeral system, or got one adopted by the community at large. So the cognitive scheme is not sufficiently strong in all individuals to guarantee that everybody counts. It could be that Gelman and Gallistel's second kind of child, the kind who spontaneously uses an unconventional sequence of words to count with – in my terms, the inventor – has been absent or very rare in the history of communities without numeral systems. This, plus cultural isolation, would explain the lack of a numeral system.

If the distinction between individuals who are natively inventors and those who are not can be sustained, this suggests a modified nativist account of the universal continuity of basic lexical numeral sequences. According to this account, some humans, the inventors, are innately pre-programmed to invent numeral systems with this property; they possess something like Gelman and Gallistel's 'counting scheme', which guarantees that any system they invent will map a sequence of expressions onto an unbroken sequence of numeral values with unit increments. The fact that the non-inventors also have systems with this property results, in this account, from their being simply the *receivers* of previously invented systems. Unlike many other parts of language, numeral systems are in fact passed on from one generation to the next with a certain amount of explicit teaching, with sessions of practice, exercise, and error-correction by adults. Since invention is a special case of acquisition, some constraints on invention also apply to ordinary acquisition. Thus ordinary, noninventive acquirers of basic numeral lexicons require, just as much as the inventors, the crutch of a recited sequence. But, unlike the inventors, they rely on being taught such a sequence, rather than making up, or adapting, one of their own.

3.5 Synthesis – A Pluralist Account

The Referential/Pragmatic, Conceptual/Verbal, and Ritual Hypotheses have been presented above as stark opposing answers to the problem of the acquisition, by invention or otherwise, of primitive numeral lexicons. Each hypothesis has at least some small drawback making it implausible as sufficient alone to account for the continuity of basic numeral lexicons.

The main problem with the Referential Hypothesis is its reliance on an assumed human ability to distinguish the numerosity of collections of things directly, without counting. Such an ability, based on no foundation of counting, is only firmly established for very low numbers, such as 2 and perhaps 3, so it is not plausible to claim that, say, 6 and 7 are differentially accessible to a significant degree. Nevertheless, the Referential Hypothesis has two points in its favour: firstly, it is plausible for the very low numbers – 1, 2, and 3; and secondly, the apparent greater usefulness in human affairs, as evidenced by the frequency data cited, of exact number words for lower values, can clearly be a reinforcing factor to whatever mechanisms do in fact account for the continuity of basic numeral lexicons.

The main problem with the Conceptual/Verbal Hypothesis exactly complements the problem with the Referential Hypothesis. Whereas the Referential Hypothesis, with its reliance on direct recognition of numerosity without counting, is satisfactory for the very low numbers, in particular 2 and possibly 3, the Conceptual/Verbal Hypothesis lacks support at just this crucial early stage, in that there is no evidence that expressions for 2 in spoken languages are formed from anything like *the number after one*, indicating construction of numerals on the basis of the successor function. Above 2, one can find examples of languages expressing, say, 3 as 2 + 1, or 4 as 3 + 1, etc., giving some plausibility to this hypothesis for numbers above 2. The Conceptual/Verbal Hypothesis as it has been presented is also unduly abstract in its emphasis. It depicts the child as a little Peano, capable of grasping and putting to use abstract concepts such as 'number' and 'successor', but gives no attention to the factor of the child's systematic exposure to concrete exemplars from life.

The Ritual Hypothesis, on the other hand, in its basic form, is unduly concrete. It roots the properties of an abstract system (a natural language numeral system) in an apparently instinctual activity which has no essential connection with the concept of number. For the Ritual Hypothesis, the problem is how the transition from an uninterpreted activity to a system of word/concept pairings comes about. And, because it starts from conventional sequences of simple words, which (somehow) become interpreted numerically, the Ritual Hypothesis cannot account for the occurrence of syntactically complex expressions such as *three-and-one* for 4.

In these hypotheses, then, there are three complementarities: very low numbers versus the rest; simple words versus syntactically complex expressions; and abstract versus concrete. I will suggest that in fact all three hypotheses identify contributing causes to the linguistic facts that we wish to account for, and that a satisfactory explanation requires all three sets of contributing causes to be present jointly. Each hypothesis taken on its own does indeed account for some observable phenomenon, but not the totality we are interested in here, namely the continuity of basic numeral lexicons. I shall start with the Referential/Pragmatic Hypothesis and the very low numbers 1, 2, and 3.

Taken on its own, the Referential Hypothesis could well account for the simple implicational hierarchy noticed in systems of grammatical number. To cite Greenberg again: 'No language has a trial number unless it has a dual. No language has a dual unless it has a plural' (1963a, p. 94). As far as I can ascertain, items in systems of grammatical number are not used in conventional reciting sequences – they are not used to count. And systems of grammatical number rarely, if ever, distinguish a number higher than 3. The domain of grammatical number systems thus corresponds closely to the very low numerosities which are recognizable by subitizing (see the studies on subitizing mentioned in Section 3.2 above). And the preponderance of the dual over the trial is, I think, plausibly accounted for by a joint appeal to the greater salience of 'twoness', compared to 'threeness', and to the greater usefulness in practical affairs of a method for referring to exactly two things, compared with a method for referring to exactly three things.

These factors would also account for the widespread existence in languages of words such as English *pair*, *couple*, *brace*. These words indicate specific (low) numbers, but are not part of, or derived from, the numeral system proper. Typically, languages have fewer such words for referring to larger collections of things. This reflects an obvious and general truth that languages tend to have words for that which is easy to distinguish and useful to name. In the same way, numeral words for 2 arise without underlying conceptual construction by the successor function. This is not to say that the etymology of words for 2 will not reveal some conceptual complexity. Menninger (1969, pp. 12–16, 172–6) reviews (somewhat picturesquely, but no doubt with sufficient veracity) some of the etymological connections of words for 2, including connections with concepts such as 2nd person (the first 'other' person), male-female pairs or couples, division and sundering, joining and connecting (for example by twisting strands), folding, equivocation and doubting, forked shapes, and so on. Bagge (1906) argues for an etymology of the words for 1 and 2 in Indo-European deriving them from the near/far demonstrative pair meaning roughly *this* and *that*, or *here* and *there*.

In summary, twoness can be perceived without counting. (See the baby and child evidence from Starkey and Cooper, Antell and Keating, and Russac, cited in Section 3.2.) Words for 2 do not reflect any conceptual analysis of 2 as the first step in the infinite march to the drum of the successor function. The same can be said, to a lesser extent of 3 and 4. Bagge's etymological suggestion for Indo-European forms for 3 is that it has the same origin as forms such as *through* and *trans*, meaning roughly *beyond* or *over*, indicating that 3 was once the limit of Indo-European numerals. But the direct perception of numerosity without counting or analysis in terms of a potentially infinite series only seems capable of taking human linguistic/numerical abilities to around 3. Beyond this, counting activity and some awareness of an abstract ordered sequence play their part. 'Sameness of number, when it is a matter of lines "that one can take in at a glance", is a different sameness from that which can only be established by counting the lines' (Wittgenstein, 1974, p. 354).

The Ritual Hypothesis, taken on its own, can account for instances of ritual recitations accompanied by pointing to objects in a collection, such as the *Eeny, meeny, miny, mo* ritual. Such rituals come naturally to children, and in contexts other than counting. According to Gelman and Gallistel, children spontaneously indulge in such behaviour, sometimes inventing their own peculiar sequences of words to be recited.

Given two or three number words, say, *one*, *two*, and *three*, which have arisen by the route sketched in connection with the Referential Hypothesis, that is by direct recognition of numerosities without counting, and because of the usefulness of being able to refer exactly to collections of two and three things, the establishment of a conventional recited sequence among these words adds to them a dimension which was previously absent or at least only latent. Whereas *three* was previously just the word used to refer to collections of three things, it is now also the word immediately after *two* in the conventional sequence. Before the establishment of the conventional recitation sequence, one could not say of *three* that it was the word immediately after *two* in the conventional sequence. Thus *three* acquires a new significance as holding a place in a conventional sequence, a significance not at odds with its previous significance, but new nevertheless.

The relation 'immediately after' between words uttered in a sequence is a temporal relation between discrete, physically experienceable events, and is thus more easily graspable from experience than the abstract mathematical notion of 'successor', which is not a temporal relation, and does not relate physical events. It would not be controversial to claim that the human organism is innately equipped to grasp the notion of temporal relation between discrete physical events, such as 'immediately after'. Thus a ritual counting sequence provides a physical exemplar with the same formal properties as we attribute to the abstract number sequence. The strong claim made by the Conceptual/Verbal Hypothesis that we have an innate conception of the successor function can be adapted in a way which integrates with the Ritual Hypothesis. Human beings, we can say, have the innate capacity to grasp the notion 'immediately after', a relation between experienceable events in a sequence, and they also have the ability to abstract away from the physical and temporal to acquire an abstract relational concept with the same formal properties (irreflexivity, asymmetry, intransitivity) as the original temporal notion. On this account, human beings are innately equipped (as perhaps no other organisms are) to acquire one of the abstract concepts involved in a full appreciation of number, namely the concept 'successor', but, except perhaps in the rare cases of inventor geniuses, can in fact only acquire these concepts via the triggering experience of being exposed to a conventional counting sequence, which provides a physical/temporal examplar of a system with the necessary formal properties. 'Klahr and Wallace (1973, 1976) proposed that subitizing predates counting as a means of determining numerosity and that counting initially takes on quantitative meaning by being used in the subitizing range' (Fuson and Hall, 1983, p. 59). This is also the position I adopt and it seems to be well supported by the psychological studies I have cited.

I suggest, then, that there is discontinuity at about the number 3, between the way the very lowest word/concept pairings are acquired (from experience of very small collections), and the way higher number word/concept pairs become known (through the counting sequence). There is further evidence for this discontinuity in two correlated formal linguistic discontinuities (first mentioned earlier in Chapter 2, Section 3), whose origins are presumably phylo- or glossogenetic rather than ontogenetic.

In many inflecting languages (for example Latin, Russian, Welsh, Ancient Greek) the first few numeral words inflect, that is take various somewhat different forms, agreeing in gender or case as conditioned by their syntactic environment. This is true for words up to about 3 or 4, after which invariant forms are used. It is of the essence of a rote-learnt sequence of words that each word have a single form. Rituals demand exact conformity from one performance to the next. If the sequence of words is learnt *before* its application to determining and expressing the cardinalities of collections, grammatical notions such as gender and case cannot be involved. So numeral words which originate in the recited rote-learnt sequence would be expected to be uninflected. The occurrence of variant inflected forms of words for 1, 2, 3 (and 4) suggests that these words originate in ways more closely integrated with their eventual use as modifiers of nouns indicating collections of things. Bagge (1906) also argues for the greater antiquity of the Indo-European forms for 1–4 on the basis of their declinability.

The first two or three numbers are also linguistically marked by having suppletive (irregular) ordinal forms in many languages. Examples are English *first* and *second*, which are phonologically quite unrelated to the corresponding cardinals. (In many languages, the word for *second* is cognate with the word for *other*.) The distinction between cardinal and ordinal only makes sense if the numerals can be considered as embedded in a linguistic context

and conveying information about the number of members in a collection (for cardinals) or the position of some individual in a sequence (for ordinals). For merely reciting in a rote-learnt sequence, the distinction between cardinal and ordinal is not relevant. The same rote-learnt sequence can be used for arriving at either a cardinal or an ordinal conclusion. But for the very low numbers 1, 2 and possibly 3, the frequent morphological unrelatedness of cardinal and ordinal numerals suggests that, down at this level only, it is possible to conceive of the ordinal meanings as unrelated to the cardinal meanings.

The existence of special forms such as *both* relating specifically to the number 2 also testifies to the particular salience of collections of just two items. A propos of *both*, I have neat, though anecdotal, evidence of the possibility of acquiring a word for 2 independently of the counting sequence. Simon Fairclough, aged 2.4, says *both* for 2 (and perhaps other plural numbers) in contexts where there is clear reference to a collection of objects. So he utters examples such as *There are both bickies* (for *There are two biscuits*), and *Give me both* (for *Give me two*, not necessarily in the context of two already mentioned objects, as would be the case in the adult usage of *both*). But Simon uses *two* in the ritual counting sequence *one, two, three, four*. This child has a word for 2, which he has not yet conflated with the second word in the conventional counting sequence.

This isolated example from the acquisition of English numerals is echoed in the Chinese numeral system, where there are two quite different words for 2, depending on whether one is reciting the counting sequence or expressing a proposition about some collection of two objects. Thus the second word in the standard counting sequence is *erh*, whereas the word meaning 2 used with nouns (and their accompanying classifiers) is *liang*. In the case of a particular English child, a combination of innate and environmental factors led him (temporarily at least) to maintain a distinction between the second item in the counting sequence and a word denoting pairs of things. In the Chinese case, presumably, these factors have happened to persist with enough individuals to cause this distinction to become established as part of the language.

Benacerraf, as a philosopher, regards the suggested discontinuity as 'likely'.

> A likely story is that we normally learn the first few numbers in connection with sets having that number of members –

> that is, in terms of *transitive* counting (thereby learning the use of numbers) and then learn how to generate 'the rest' of the numbers. ... Learning these words, and how to repeat them in the right order, is learning *intransitive* counting. (1965, p. 50)

(I would prefer to substitute 'numerals' for 'numbers' in this quotation; Benacerraf conflates numbers with numerals for reasons having to do with his anti-number-realist argument.)

The suggested movement from perceptually based acquisition of quantity concepts, including number, to language-based acquisition, including the involvement of counting, also emerges emphatically from Siegel's psychological work. The following quotation reiterates conclusions drawn several times, in connection with different experiments, conducted mainly on 4-to 6-year-olds.

> In general, there is an increase with age in the degree to which language plays a role in the child's understanding of quantity. Perceptual, nonquantitative factors play a significant role early in development and appear to precede the use of language. As the child develops, there is movement away from a perceptual matching strategy to a conceptual, numerically based one. ...
>
> In summary, we have demonstrated the predominance of perceptual nonlinguistic operations in early quantity concepts and the increasing role of language in the solution of tasks involving elementary notions of quantity. (1982, pp. 152–3)

Acquisition of the meanings of higher-valued numerals and of truths involving them must come, in the view taken here, via language and the conventional counting sequence Certainly, children can easily learn stable sequences of words and even make up their own. The step from ritual recitation of a word sequence to the practical use of this sequence in assigning cardinalities to collections may possibly not come easily to all children, but all normal children are capable of making this step. In this area, many children are given a lot of deliberate help by adults, with a certain amount of drilling in the counting sequence, and the explicit going over of simple conclusions to be drawn from its application. You've got three toys here. Now, what comes after three? So if we put one more toy in, how many have we got?'

And so on. Obviously, this kind of coaching happens much more in some households than in others, and, as one would expect, there are great differences in the very elementary arithmetical abilities of 4 and 5-year-olds, related to socio-economic class.

Hughes (1984, 1986) compared working- and middle-class children on a range of simple arithmetical tasks, ranging from perfectly concrete visible operations with collections of two or three objects to quite abstract exercises involving verbal sums as difficult as 6 + 2 = ?, presented orally.

> there was a substantial difference between the middle-class and working-class children in their overall performance. ... This difference was equivalent to about a year's difference in age: the working-class 4-year-olds were performing at about the same level as the middle-class 3-year-olds, while the working-class 5-year-olds were performing at the same level as the middle-class 4-year-olds. (1986, p. 32)

Similarly, also probably indicating the role of some coaching and explicit instruction in the applicability of the rote-learnt counting sequence to determining the cardinality of collections, Fuson and Hall write:

> The main body of evidence indicates that middle-class children are able to apply the cardinality rule by the age of four and that inner city children may be somewhat delayed in this task. ... Ginsburg and Russell (1981) reported ... that most of their middle-class pre-schoolers (mean age 4–3 years) displayed the cardinality rule for sets of three, five, eight, and eleven, whereas less than half of their inner-city age-mates (mean age 4–5 years) did so. (1983, pp. 64–5)

The 'cardinality rule' mentioned here is the rule by which an inference regarding the cardinality of a set is made from an instance of the counting activity ending with a particular numeral word. To acquire this rule is to learn the connection between what Benacerraf calls 'intransitive' counting and 'transitive' counting. If learning this connection seems not to be the kind of step that could be taken with any ease or assurance by a child, it must be remembered that the term 'step' reflects a possibly somewhat misleading idealization. As with the term 'invention' earlier, the most one can hope to observe is some difference in the states of affairs obtaining at separate points in time. A so-called 'step' is a theoretical notion postulated to account for this difference. What actually goes on in the mind of the child during this 'step' may be very gradual, tentative, and erratic. But, I claim, something does happen, involving the verbal activity of counting, and enabling the child to progress beyond the limit of the number 3, which seems to be the upper bound for purely non-verbal concepts of particular numbers.

Hughes also has evidence for this distinction between non-verbal reasoning and (covert) verbal reasoning involving counting for higher numbers. In the following quotations 'small numbers' are 1–3 and 'large numbers' are 5-8.

> The interesting question is whether we in fact deal with small numbers *in a different way* from large numbers. It has often been suggested, for example, that we can judge the number of items in a small visible group by a direct process of visual apprehension sometimes called 'subitising' (e.g. Klahr and Wallace, 1973). For numbers larger than four this process becomes less reliable, and we usually have to count instead.
>
> The children's spoken comments as they worked on the problems suggested that they were in fact using different strategies for small-number and large-number versions of the Box task. For problems involving small numbers, they would either simply name the final quantity of bricks, or count up to that number as if they had constructed some sort of image or representation of the bricks in the box. ... In contrast, most children who succeeded on large-number problems appeared to be using a different strategy based on *counting on* from the *initial* quantity. ... For subtraction, they would have to work down the scale. ...
>
> Further evidence of these two different strategies came from children who *used their fingers* to represent the contents of the box. Usually this only occurred with small-number problems, and involved the representation of the *final* number of bricks in the box. ...
>
> Other children used an intriguing strategy which seemed to rely on a direct *visual image* of the bricks. These children tapped at different places on the closed lid of the box while answering, as if the lid was transparent and they were counting the bricks inside. As with the finger strategy, this was almost entirely restricted to small-number problems. ...
>
> This strategy was less likely to be used when the numbers involved were five or more.
>
> [A] larger study also gave further support to the idea that the children use different strategies for the small-number and large-number problems. As before, some children used their fingers to represent the overall number of bricks for small-number problems, while others explicitly used the counting-on strategy for large-number problems. There was also some indirect evidence for these strategies in the relative success rate for addition and subtraction problems. If the children were performing the small-number problems by constructing some representation of the final amount, then we would not expect any difference in the relative difficulty of problems involving addition and those involving subtraction. This indeed was what I found for the small-number problems: children were just as successful on addition problems as on subtraction problems. However, a different pattern might be expected for the large-number problems. If children's strategy on these problems is to count up or down the number scale, starting from the initial contents of the box, then one might expect that addition would be easier than subtraction. After all, children have more experience counting up the number scale (five, six, seven) than counting down (seven, six, five). Again, this was confirmed: I found that for the large-number problems, children were more successful with addition than with subtraction. (1986, pp. 28–31)

Hughes tested children on five tasks of increasing abstractness, or decreasing embeddedness in a concrete situation. These tasks were labelled Box Open, Box Closed, Hypothetical Box, Hypothetical Shop, and Formal Code. In Box Open the child sees bricks being put into and taken out of a box, whose lid remains open, so that the child can see the bricks in the box all the time. The investigator discusses the number of bricks in the box with the child. In Box Closed, the lid is closed while discussion of the number of bricks in the box takes place, so that the child cannot see how many bricks are in the box. In Hypothetical Box the box and bricks were put away and the child was asked questions such as: 'If there were two bricks in the box and I put one more in, how many would be in the box altogether?' In another task form (Hypothetical Shop) the question referred to a context quite divorced from the immediate present. The child was asked, for example: 'If there were two children in the sweet shop and one more went in, how many children would be in the sweet shop altogether?' The most disembedded task form was when the problems were presented in the formal code of arithmetic: the child was asked questions such as, 'What does two and one make?'

> [There was] a statistically significant difference between the hypothetical problems and the Box task but for *small numbers only*. There was no such difference when the numbers were slightly larger.
>
> Why should hypothetical problems be harder from small numbers only? One possibility is that this difference is related to the different strategies proposed earlier for small-number and large-number versions of the Box task. If children do construct some sort of image or representation for small-number problems, then this may well be helped by having the bricks and the box actually present in front of them. If on the other hand they use the counting-on strategy for the large-number problems, then it would seem that the physical presence of the bricks gives only a minimal advantage. (Hughes, 1986, p. 32)

The Ritual and Conceptual/Verbal Hypotheses introduced in the preceding sections are two sides of the same coin. The Conceptual/Verbal Hypothesis emphasizes Man's potential ability to grasp the notions of number, 1, and successor. The Ritual Hypothesis emphasizes the importance of a physical/temporal exemplar, a conventional counting sequence, from which these notions are abstracted. Humans are able, and to some extent disposed, to recite stabilized lists of words while pointing to items in a collection. It is hard to imagine that the pioneering invention of conventional counting sequences was not motivated in some vague way by the rudiments of an understanding of number. Putting it simply, the first people to count probably had some extremely vague, totally inexplicit, idea at the back of their minds of what they were doing and why. The potential availability of number concepts provides meanings for the words in the recited sequence, and the availability of words in an established sequence

facilitates the successive differentiation of further number concepts. Together, the abstract conception and the concrete ritual work together to produce the strong, perhaps absolute, tendency for the basic numeral words in languages to correspond to a continuous sequence of numbers.

These low-valued words also, happily, turn out to be among the most useful in practical affairs, so that the factors mentioned under Referential Hypothesis also exert some pressure towards this result. If there are genuine counter-instances to the tendency towards continuity of basic lexical numeral sequences, as there may have been in the prehistory of Bantu, Ainu, and some other languages, these seem likely to be due to the fact that 'tenness' may be more salient than, say 'eightness', because of Man's ten fingers, so that the Referential Hypothesis may be useful in explaining counter-instances. The fact that Man has a set of extremely salient protuberances, conspicuously arranged in a sequence which is itself both salient and almost impossible to rearrange, would have helped in the very earliest development of numerals. If Man had been shaped like the whale, then even with Man's verbal and intellectual capacities, he would have been less likely to become aware of the possibility of associating a stable sequence of words with objects in his environment.

Allowing rival hypotheses to remain in the field, competing, rather than finally eliminating each other, is a model for the explanation of complex phenomena that is well worth exploring.

3.6 The Joint Acquisition of Numerals and Number

Starting from a linguistic phenomenon, namely the continuity of basic lexical sequences in numeral systems, the attempt to explain this phenomenon has arrived in territory usually argued over by philosophers of mathematics. Yet it is surprising how little the mathematicians and philosophers look at natural language when discussing the nature and basis of number, often simply dealing generally in terms of 'the notation' (for example Benacerraf, 1965, p. 50), or, if specifically, in terms of the written Arabic place-value notation (Blackburn, for example 1984, *passim*). It should be apparent that the hypotheses discussed in the previous section are not attempts to account *merely* for linguistic phenomena. They also make claims about the development of the number concept itself and about the involvement of natural language in this development. I believe that natural language plays a vital part in the development of the number concept in individuals. I give below a sketch of what particular capacities, linguistic and nonlinguistic, it seems reasonable to attribute to a child capable of acquiring numeral words and their meanings (number concepts). This will be followed by an outline of the kinds of stages the acquisition presumably takes. The whole can stand as a claim that linguistic considerations are central to the issue of the development of numerals/number. The inventory of given apparatus and developmental steps outlined below makes more explicit and systematic the concluding 'pluralist synthesis' of the previous section. Naturally, the same psychological and linguistic evidence are relevant here.

In what follows, I use the term 'concept', not to indicate essentially private entities, which may differ arbitrarily from one person to another, but rather in this sense: a concept of X or Xs is what an individual may be said to possess if he gives evidence in intelligent behaviour, talk, and so on, of knowing more or less the same basic or essential things about X or Xs as other individuals. The absence of such evidential behaviour does not necessarily indicate absence of a concept; other factors may prevent the relevant behaviour from manifesting itself. I assume, uncontroversially, that a concept is a complex psychological entity bringing together logical, encyclopaedic, and linguistic information along lines such as those sketched by Sperber and Wilson (1986, p. 86). I do not propose to be more precise than this, but rest on these admittedly vague statements to try to evoke in the reader my concept of a concept.

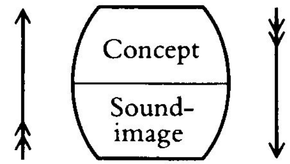
3.6.1 Given apparatus relevant to the acquisition of a basic numeral lexicon and the concomitant number concepts.

The child possesses at a very early stage:

(a) The concept of a word in general. that is the child is able to tell whether a stretch of acoustic signal is, or is not, a word, according to the previously acquired knowledge of the phonology, morphology, and syntax of its language.
(b) The potential to form concepts of particular words. That is the child is able to learn to decide correctly whether some stretch of acoustic signal is, or is not, for example, an instance of the word *one*, or *two*, or *three*, or *dog*, or *tree*, and so on.

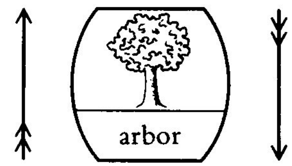
(c) Some form of the concept of an individual object in general. That is, given a collection of objects, the child is able to pick out the objects. [The concept of an individual event is already presupposed under (a) above, as a stretch of acoustic signal is an event.]
(d) Concepts of particular objects and object types, for example of a particular person or, say, of dogs in general.
(e) The Sign concept, a crucial ingredient of the language faculty. The child is able to establish a connection between particular word-concepts and particular nonlinguistic concepts. In Saussurean terms:

> The linguistic sign is then a two-sided psychological reality that can be represented by the drawing:



(Saussure, 1959, p. 66)

Saussure's example of a particular sign is (p. 67):



[Such drawings are not to be taken naively as anything more than suggestive expository devices. The concept of a tree, or treeness, is not necessarily anything like a picture, or an image, of a prototypical tree. Godel writes: 'even the editors of the *Cours* went astray in adding to the genuine diagram a second one with the design of a tree for the *signifié* of Lat. *arbor*, thus suggesting to the readers the very erroneous conception against which de Saussure warned his students, that is, the idea of the *signifié* being the image of an object' (1970, p. 486). No doubt concepts, such as that of a tree-in-general, are very complex. Rather than using a picture, one might say that the concept of a tree is an organism's (or a computer's) program for recognizing particular objects as trees. Or one might say that what occupies the significatum slot of the 'tree' sign is the characteristic function of the set of trees, a function from objects to truth values. Clearly, a (computer) program or a function is nothing like a picture. Similarly, a Saussurean sound-image is whatever internal apparatus it takes for an organism to recognize particular acoustic signals as instances of a particular word, say *tree*.]

(f) The concept of plurality or collection of more than one object. The child is able to distinguish a collection of several objects from a single object.

(g) [Present less strongly than (f).] The concept of twoness, or of a two-collection. The child is able to distinguish a pair of objects from a single object, and from a larger collection of objects.

(h) [Present less strongly than (g), if at all.] The concept of threeness, or of a three-collection. The child is able to distinguish a trio of objects from a single object, or a pair, and from a larger collection.

(i) The concept of stable sequences of words. That is, the child is able to recognize correctly a stretch of acoustic signal as an instance of some particular sequence of words in a particular situation type, for example 'eeny, meeny, miny, mo, . . .' in a dipping game situation, 'one, two, three, . . .' in a counting situation.

(j) The concept of the action of placing an object into a collection. The child is able to distinguish the act of placing an object into a collection from other acts, such as blowing his nose, reaching, and so on, either as carried out by the child himself or by others.

(k) A concept of the result of an action. The child is able to understand, for example, that a state of affairs after some action *results from* the action.

(l) The ability to form concepts, not only by experience of real-world exemplars, such as trees, objects, events, and utterances, but also by syntactic combination of existing concepts. (This will be justified below.)

These abilities, concepts, and so on, attributed to the child, are not offered wholly in the spirit of a set of axioms or postulates. That is, the intention is not simply to try to establish, on methodological grounds of parsimony, the barest set of characteristics which could logically form the basis for an organism's acquisition of number and numerals. The intention is, rather, to point out what, on the basis of mostly commonplace observation,

it seems reasonable to claim children are actually endowed with. Strictly, dogs do not need all four legs for locomotion, but a reasonable account of how they acquire locomotive ability assumes that they have four, with whatever advantages that may bring. None of the abilities attributed to the child in (3.6.1) should be particularly controversial. All are readily evidenced in the behaviour of prenumerate (that is pre-numeral system) humans, whether children or adult speakers of languages without numeral systems. Several of these capacities – for example (c), (d), and probably, (k) – can be uncontroversially attributed to higher animals, too. It is not necessary to delve into equivocations as to whether these capacities, in exactly the form described, are strictly innate. It is sufficient to point out that children possess these attributes at a very early age.

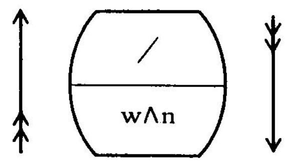
However rich the innate apparatus one attributes to the child, one cannot escape the conclusion that acquisition of more elaborate knowledge involves induction from experience in some form. The richer the pre-existing apparatus, the less work there is for the inducer to do, but it cannot be denied that induction from experience plays a role in the acquisition of language and of number. I now sketch the information which I presume an acquirer of English basic numeral words and the corresponding number concepts gleans by induction from relevant experiences.

3.6.2 Steps in the induction of a basic numeral lexicon and the concomitant number concepts.

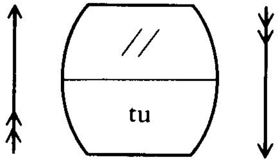
The child learns that:

(a) There is a word 'one' /wʌn/.

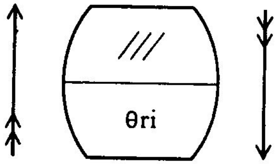
(b) The word /wʌn/ is associated with the concept of oneness, a concomitant of the concept of an individual object. Thus the child acquires the sign associating this concept with the word *one*. I suggestively represent this as below, but not claiming that at this stage the child has a fully fledged concept of the number 1. All she has is the notion of oneness that comes with the object concept.



(c) There is a word 'two' /tu/.
(d) The word /tu/ is associated with the concept of twoness, or two-collection, the child thus acquiring the sign



Again, I do not claim that the child has a full conception of the number 2 at this stage, but merely whatever distinguishes *pairs*, which she is able to recognize.
(e) (Perhaps) there is a word 'three' /θri/.
(f) (Perhaps) There is a sign



(g) There are other words (*three*), *four*, *five*, ... One environment in which these occur is the stable word sequence, *one*, *two*, *three*, *four*, *five*, ... The child learns to recognize this stable word sequence, analysing it into its component consecutive words. The first two (or three) words are words already associated by signs with non-linguistic concepts of an object or of collections of objects. The remaining words are at this stage unassociated by signs, that is effectively still nonsense words. (Fuson et al., 1982 give a very full account of the developmental details of children's acquisition of the counting sequence, which is clearly structured, and gradual rather than instantaneous. It is enough for my purposes that the counting sequence is actually acquired, however gradually.)
(h) Placing an object into an existing collection results in a collection. This applies to collections of any size, although the child at this stage does not have words for any but the smallest numbers. The information gained can be thought of roughly as the rule PLURAL + 1 = PLURAL, or MANY + 1 = MANY. Perhaps one should prefer to say that this rule is innate, rather than learned from experience: the issue is not central here. This rule states a principle basic to the underlying ontology of objects and collections, a matter taken up in Chapters 4 and 5.

(i) More particularly, placing an object alongside a single already present object results in a two-collection, or pair. The concepts of object, placing, result, and two-collection are, I have claimed, previously known to the child. What she is learning at this stage is a further truth involving them, a truth which could be expressed as 1 + 1 = 2. (The linguistic fact that the truth could be so expressed is not necessarily known to the child at this stage.) This begins to fill out the concepts of both 1 and 2, starting to fit them into an elaborated system of truths.
(j) (Perhaps also) in particular, placing an object into a two-collection results in a three-collection. 2 + 1 = 3. The child learns something new about the numbers 1, 2, and 3, and thus further fleshes out her concepts of these numbers.
(k) There is a parallel between the counting sequence (which the child now knows) and the elementary number rules just induced: *one* is followed by *two* in the counting sequence; and placing an object with an object (a oneness) results in a two-collection. Perhaps also: *two* is followed by three in the counting sequence; and placing an object with a two-collection results in a three-collection.
(l) Inductive generalization
If X is followed by Y in the counting sequence, placing an object in an X-collection results in what is called a 'Y-collection'. Thus, what results from placing an object into a three-collection is called a 'four-collection' (new concept). And so on, as far as the conventional sequence of words stretches. Pandora's box is open – a handsbreadth or two.

Steps (h)–(j) do not involve language, or words, and are logically quite independent of steps (a)–(g), which do involve language. The child makes parallel but independent progress on two fronts, that of the particular system of signs in the language she is learning, and that of the elementary truths about objects and tiny collections of objects. The relative timing of developments on these two fronts is not important for the principle of the argument here. 'It is not uncommon to find children who can rote count but who cannot properly count a set of objects and arrive at the correct answer for the number of objects in the set.' (Siegel, 1982, p. 123) Steps (k), (l), which involve bringing together linguistic knowledge and elementary knowledge of objects and tiny collections, are the crucial steps into the beginnings of a numeral system and the unlimited potential of number.

Steps (h), (i) and, if it happens, (j) are, I am claiming, inductions from experience. This concurs with Mill's opinion on the acquisition of such truths as 2 + 1 = 3.

> Three pebbles in two separate parcels, and three pebbles in one parcel, do not make the same impression on our senses; and the assertion that the very same pebbles may by an alteration of place and arrangement be made to produce either the one set of sensations or the other, though a very familiar proposition, is not an identical one. It is a truth known to us by early and constant experience – an inductive truth; and such truths are the foundation of the science of numbers.
>
> We may, if we please, call the proposition, 'Three is two and one,' a definition of the number three, and assert that arithmetic, as it has been asserted that geometry, is a science founded on definitions. But they are definitions in the geometrical sense, not the logical; asserting not the meaning of a term only, but along with it an observed matter of fact. The proposition, 'A circle is a figure bounded by a line which has all its points equally distant from a point within it,' is called the definition of a circle; but the proposition from which so many consequences follow, and which is really a first principle in geometry, is, that figures answering to this description exist. And thus we may call 'Three is two and one' a definition of three; but the calculations which depend on that proposition do not follow from the definition itself, but from an arithmetical theorem presupposed in it, namely that collections of objects exist, which while they impress the senses thus, $\substack{\text{OO}\\\text{O}}$, may be separated into two parts, thus, OO O. (1906, pp. 168–9)

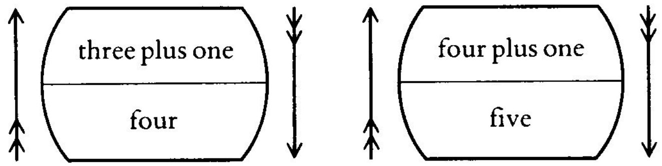
In my terms, Mill claims that a child learns from experience that adding an object to a two-collection results in a three-collection. Mill's exposition would have benefitted from maintaining a clear expository distinction between numbers and numerals, that is between the arithmetical objects (or concepts) themselves and the

words that are used to name them. Definitions, which Mill involves in his discussion, essentially concern words, or symbols in a language of some sort. What the child learns, I am claiming here, is essentially non-linguistic, a truth about what happens when collections of objects are manipulated. One might conceive of an alinguistic creature, such as a cat, learning this truth, without ever learning a language to express it in. (As a matter of fact, however, cats probably do not even learn this truth.) But, being alinguistic, a cat cannot possibly learn a definition. Thus an organism may know the truth of what we represent by 1 + 1 = 2, without knowing it in words.

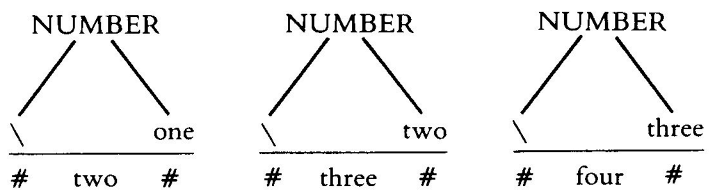
I restrict my claim to what numerical truths a child learns by induction from experience to a very small set: PLURAL + 1 = PLURAL, 1 + 1 = 2, and possibly 2 + 1 = 3. Mill goes further, too far, it seems. 'The fundamental truths of that science [of numbers] all rest on the evidence of sense; they are proved by showing to our eyes and our fingers that any given number of objects, ten balls, for example, may by separation and rearrangement exhibit to our senses all the different sets of numbers the sum of which is equal to ten' (1906, p. 169). Just how much Mill is claiming here depends on how one interprets the verbs 'rest' and 'prove'. But he seems to be saying, for example, that independently of any verbal definition or operation one can directly experience the specific fact that 7 + 3 = 10, by some means such as arranging pebbles. I take this to be very implausible. And, of course, if the view is applied to very high numbers, it becomes patently absurd. But down at the very bottom end of the number sequence, it seems to me quite plausible that induction of very elementary arithmetical truths from pure (non-verbal) experience of collections takes place.

[Frege's objections to Mill's identifying numbers as properties of collections will be discussed in the next chapter, where the rather vague suggestion that 'placing an object in an X-collection results in a Y-collection', made above in steps (k), (l) of (3.6.2) is also taken further, in response to Frege's views.]

Steps (k), (l) of (3.6.2) make the crucial leap from nonlinguistic truths about very low numbers to truths about higher numbers, involving words. As a result of the inductive generalization in (l), I must claim that the child has potential access to knowledge of signs of a new type, in which the significatum is not, as in previous cases, a non-linguistic concept, but rather a formula containing another significans, that is something roughly like (3.6.3).



(For the moment, I will continue to couch this discussion in terms of the Saussurean sign, as a simple and convenient metaphor for the relation between sounds and meanings.) The choice in (3.6.3), for the sake of an example, of formulae with the specific form 'numeral-word plus one' is fairly arbitrary here. It is sufficient to make the point, but other forms of words, such as perhaps 'the result of adding an object to a three-collection' would suit the purpose equally well. Or again, the significatum slot of the sign for *four* could specify a function from collections to truth values, the characteristic function for foursomes of things; the specification of this function would be structured in such a way that the function 'calls' the pre-existing functions for threesomes of things, adding an object to a collection, and result of an action. Signs as in (3.6.3) could be taken as meaning postulates, or, in Fodor's (1976) terms, a record of an abbreviation, in the internal language of thought, for a complex expression of that same language. It is also possible that the formulae in the significatum slot of the sign in these cases should be mixed representations of some kind, containing both linguistic and non-linguistic entities. Interestingly, the lexical entries for numeral words proposed in Hurford (1975) are notational variants of signs as in (3.6.3), but there, characteristically of an enterprise in the paradigm of generative grammar, a single uniform treatment was adopted for both very low-valued words (*two*, *three*) and higher-valued words (*eight*, *nine*).



(Hurford, 1975, p. 39)

I now claim a kind of psychological plausibility only for such lexical entries in the cases of *four* and above. Here is a case where a generative linguist's search for regularity and generalization caused him to postulate uniformity and continuity where in fact consideration of psychological evidence (such as that of Hughes, cited in Section 5 above) would suggest discontinuity.

I claim only that a child who has mastered the inductive generalization (l) of (3.6.2) *has access to* signs roughly along the lines of those in (3.6.3). In the next chapter, this inductive generalization will be interpreted as a semantic generalization about the denotation of noun phrases containing numerals. It only seems necessary to attribute knowledge of the *general* proposition to a child who has acquired the meanings of numeral words, and so it is not necessary to claim that *specific* signs as in (3.6.3), or specific lexical entries as in (3.6.4), are represented in the mind.

The inductive generalization (l) of (3.6.2) is at the heart of the human capacity to deal with number as a unified system containing both very low numbers (1-3) and higher ones, and involving the cardinality of collections of objects and an ordered sequence of expressions. It is a very daring generalization. It is made on the basis of at most two relevant cases. Children acquiring language and number usually get help and prompting to make this generalization, but the uniform and presumably genetically determined ability across the species to make this leap is possibly the single most important factor distinguishing basic human capacities in relation to number. The recursiveness of syntactic structure in language also plays an important part in a more developed notion of number, a topic to be taken up in Chapter 6.

4

Numbers: the Meanings of Numerals

This chapter sets out in detail a view of the way in which counting words can be put to use in making assertions about the world. This utility is crucial to the evolution and acquisition of numeral systems. The use of numerals referring to 'abstract' arithmetical entities derives from more concrete uses referring to aggregates and collections of objects. This chapter concentrates on the semantics of numerals, but necessarily some simple assumptions are made about the syntax of constructions containing them. The semantic view developed in this chapter provides a basis for the explanations to be offered in Chapter 5 for the rise of more complex syntactic constructions involving numerals.

4.1 Frege's Boots: Numbers and Collections

Mill regarded numbers as properties of collections, or aggregates. At face value, and at its simplest, this view holds that the aggregates or collections of things which may be observed in the world each objectively possess a particular property which may be interpreted as a number. For example, the pile of books now on the table in front of me has the property of fourness, and does not have the property of threeness or fiveness, because there are exactly four books in the pile. Frege ridiculed this view in the following and other, equally lively, passages. '*One* pair of boots may be the same visible and tangible phenomenon as two boots. Here we have a difference in number to which no physical difference corresponds; for *two* and *one pair* are by no means the same thing, as Mill seems oddly to believe' (1950, p. 33e). Frege expounded the faults in the idea of number as a property of

external things in sections 21–25 of the *Grundlagen* with great clarity and force. Reading these sections is enjoyable but exasperating, because one sees that the view he is attacking is in a sense clearly wrong, but one feels equally well that there is something right in it, which Frege utterly omits to give credit to. Here I will explore, largely following Frege, what is wrong with the view that numbers are properties of aggregates, but I will also point out the virtues of that view. This will lead to an analysis by Armstrong (1978) in which Frege's insistence on the importance of *concepts*, as opposed to external objects, is conceded, but numbers are still held to be properties of external objects. This analysis has the advantage over Frege's view in that it makes it relatively easy to see how number concepts could be *acquired* from experience of aggregates and collections of objects.

To bring out the problems with numbers as properties of external objects, nothing surpasses Frege's own arguments:

> It marks, therefore, an important difference between colour and Number, that a colour such as blue belongs to a surface independently of any choice of ours. The blue colour is a power of reflecting light of certain wavelengths; to this, our way of regarding it cannot make the slightest difference. The Number 1, on the other hand, cannot be said to belong to the pile of playing cards in its own right, but at most to belong to it in the way in which we have chosen to regard it; and even then not in such a way that we can simply assign the Number to it as a predicate. What we choose to call a complete pack is obviously an arbitrary decision, in which the pile of playing cards has no say. But if we examine a certain pile of cards in the light of this decision, we may discover, let us say, that we can call it two complete packs. Yet anyone who did not know what we call a complete pack would probably discover in the pile any other Number you like before hitting on two. (1950, p. 29e)

Quite right, but let me plant the seed of another view here. In this passage, Frege mentions *choice* several times and implies that the choice of what number we assign to the pile of cards is quite arbitrary. But he surely exaggerates when he claims that 'anyone who did not know what we call a complete pack would probably discover in this pile any other Number you like before hitting on two.' Really, *any* other number? Surely a significant number of people would, if given a pile made from two complete packs of cards, say the number concerned was 104 (assuming no jokers). There *is something* non-arbitrary about the 'hundred-and-fourness' of such a pile of cards. And I would also guess that the number 2 would come somewhere in the first 20 suggestions made by most people, if they could be persuaded to play this game for long enough. So even twoness, which admittedly is not immediately apparent in the pile, comes to the surface before some really arbitrary number like, say, 1,674,821. A similar objection can also be made to the next passage from Frege's argument.

> To the question: What is it that number belongs to as a property? Mill replies as follows: the name of a number connotes, 'of course, some property belonging to the agglomeration of things which we call by the name; and that property is the characteristic manner in which the agglomeration is made up of, and may be separated into, parts.'
>
> Here the definite article in the phrase 'the characteristic manner' is a mistake right away; for there are very various manners in which an agglomeration can be separated into parts, and we cannot say that one alone would be characteristic. For example, a bundle of straw can be separated into parts by cutting all the straws in half, or by splitting it up into single straws, or by dividing it into two bundles. Further, is a heap of a hundred grains of sand made up of parts in exactly the same way as a bundle of 100 straws? And yet we use the same number. The number word 'one', again, in the expression 'one straw' signally fails to do justice to the way in which the straw is made up of cells or molecules. (1950, pp. 29e–30e)

OK, but Mill's idea of a 'characteristic manner' is not completely wrongheaded. Frege picks his battlefields to advantage – bundles of straw and heaps of sand. Mill, naturally, since it suited *his* case, preferred the examples of the fingers on a hand and little collections of three pebbles. For some agglomerations, such as the fingers on a hand, we can at least say that there *tends to be* a 'characteristic manner in which the agglomeration is made up of, and may be separated into, parts'. It is the fact that humans tend to partition certain kinds of small aggregates in a characteristic and uniform manner that makes it possible for the notion of number to be exemplified concretely and passed on from one generation to the next. It is also this which, I would claim, facilitated the acquisition by invention of number in the first place.

The problem is similar to that of ostensive definitions. How can I successfully define *cat* by pointing to a cat unless I share with my interlocutor a characteristic manner of mentally classifying the objects of our perceptions? Holding up three fingers and saying 'This many is three' usually succeeds because in such a situation children are not so constituted (or so perverse) as to focus their attention on knuckles, or wrinkles, or rings.

But even if we could convince Frege that humans do possess a common characteristic manner of separating an agglomeration into parts, he would probably maintain that this still does not make numbers properties of the agglomerations themselves. How humans see things is a matter for psychology. 'But arithmetic is no more psychology than, say, astronomy is' (Frege, 1965, p. 37e). Frege maintained a strict fundamental principle 'always to separate sharply the psychological from the logical, the subjective from the objective' (p. Xe), and he was fond of citing astronomy as the paradigm objective science and astronomical facts as exemplary objective facts. Frege's strict separation of psychology from fields such as astronomy and arithmetic has the status of a methodological working principle, useful as long as it leads us to insights. But knowledge has advanced, and it is interesting to cite a famous astronomer writing roughly halfway between when Frege wrote and the present, on the separation of the 'objective' from the 'psychological'.

> Recognizing that the physical world is entirely abstract and without 'actuality' apart from its linkage to consciousness, we restore consciousness to the fundamental position instead of representing it as an inessential complication occasionally found in the midst of inorganic nature at a late stage in evolutionary history. (Eddington, 1928, p. 332)
>
> A rainbow described in the symbolism of physics is a band of aethereal vibrations arranged in systematic order of wavelength from about .000040 cm. to .000072 cm. ... But although that is how the rainbow impresses itself on an impersonal spectroscope, we are not giving the whole truth

and significance of experience – the starting point of the problem – if we suppress the factors wherein we ourselves differ from a spectroscope. (1928, pp 328–9)

The characteristic manner in which humans separate aggregates into parts may be a partly psychological fact, but it also reflects some external property of the aggregates themselves. All we really have is the interaction between us as subjects and the objects. The properties, number, colour, and so on emerge from this interaction.

Armstrong neatly turns a simple key in the door which appeared locked to Frege.

> Could number be a property of particulars? Frege (1884, SS 22) and others developed an argument, now widely used, to show that number is not such a property. The argument is that a particular has no definite number until it has been brought under some concept. Consider the particular which is the aggregate of the Fs. Given a suitable concept, it may have the number nought, one, two ...
>
> I believe that we ought to be suspicious of this argument. It amounts to saying that numbers are not properties of particulars because particulars have indefinitely many numbers. Might not the correct reaction be to say that the particulars have *all* these numbers as properties? ...
>
> [This] page is, of course, *three-parted, four-parted* ... perhaps *infinitely parted*. I suggest that these are all structural properties of the page. These properties will stand to each other as parts to wholes. ... But a particular which is absolutely indivisible, if there are any such, would have none of these properties. ... A particular which was nothing more than two absolutely indivisible particulars would have the property, *being two-parted*, but no other of these structural properties. (1978, p. 71–2)

Armstrong then gives an account of how threeness can be said to be a property of a group of three apples in a way which seems wholly to overcome the Fregean objections.

> Suppose it to be the case that there are three, and only three, apples in the room. The aggregate of these apples is a particular. This particular has an indefinite number of parts, perhaps an infinite number. But this particular has three and only three parts such that the predicate 'an apple' applies to them. ... the predicate will apply in virtue of certain properties which the three parts of the aggregate have. ... It is a matter of the aggregate having three parts, but three parts such that each one of them has properties which make that part exactly one apple. (1978, pp 73–4)

The importance of Frege's view of numbers is in its great emphasis on their relation to concepts. To use Armstrong's last example, the aggregate in question has the property threeness in relation to the concept 'apple'. To go back to Frege's boots example, once the relevant concept is tied down by a predicate, such as *pair of boots* or just *boots*, an extra dimension is added to this 'same visible and tangible phenomenon' and we can assign a number – 1 if the concept is 'pair', 2 if the concept is 'boots'.

It might be a reasonable move to meet Frege's most telling points by taking a *modal* view of the notion 'collection'. To take this view would be to assume that collections do not exist out there in the world, but that abstract, presumably mental, entities called 'collections' *can be constructed* out of (representations of) objects that do exist out there, by 'bringing them together' in some way. The mental act may or may not on occasion be accompanied and actually facilitated by physical manipulations of the objects concerned. If we naturally assume that bringing objects under a concept (for example 'apple', 'brick') is a part of the process of constructing a collection, then the collection arrives (emerges from the construction site) with a unique cardinality associated with it.

In many cases, aggregates present themselves to us as collections with sufficient force and clarity that we are aware of no particular constructive effort in seeing them as collections, and in such cases, particularly with very small collections, their cardinality seems like an overtly accessible property. In other cases, some mental construction is clearly involved. How many objects are there on my table? Do I count the bunch of tulips as one object or as several? In the non-modal way of talking, we attribute objective existence to any collection which could be constructed. So there are several different collections corresponding to the objects on my table. Where it is important, I will make a distinction between 'collections' and 'aggregates'. Where this distinction is observed, an aggregate is a typically physical, perhaps scattered, existing object, whereas a collection is an abstract entity in some sense constructed from an aggregate. An aggregate may have many cardinalities, as different collections are constructible from it. But a given collection, in this strict usage, only has one cardinality. In the construction of a collection out of an aggregate the identities of the member objects are not lost, so that a particular collection corresponds to exactly one aggregate. The aggregate-collection relation is one-to-many.

Just as, for concreteness, mathematicians prefer to speak of numbers as existing objects (that is non-modally), I will continue, as a matter of convenience, to speak non-modally of collections as existing objects. Adopting this non-modal way of talking, I attribute objective existence to three kinds of entities: individual or atomic objects, from which collections cannot be constructed (by division); aggregates, out of which collections can be constructed; and collections. Lest it be misleading to say that collections are 'constructed out of' aggregates, I mean this in the sense that a collection corresponds both to a way of dividing a single aggregate, and to a way of putting objects, which may be aggregates, together. The 'direction' of the construction (that is dismembering versus assembling) is unimportant here. This three-way ontology corresponds to that of Link (1983), who has: 'individual portions of matter' (his set D); 'singular objects' including non-individual portions of matter (his set A-D); and 'plural objects' (his set E-A).

In the case of the number 1, this can also be seen as an abstraction from a physical notion, that of an individual discrete object.

> The singular-plural-dual progression very likely derives from the primacy of the individual object in perception. Each object is perceived as an individual with an identity of its own, constant over time. Collections are conceived of as groupings of individual objects in which each object has its own identity. Hence the basic contrast is between an individual and a collection of individuals, between singular and plural. (Clark and Clark, 1977, p. 537)

There is a large literature on the 'object concept' from Piaget through Bower (1974), discussing detailed issues of what exactly is comprised in the concept of an object, at what stages in development the various components are acquired, and so on.

Clearly the object concept is itself complex. (See Chomsky, 1976, p. 203, Pulman 1983, pp. 53–78 for some relevant discussion.) But all that is to be used here is the acknowledged fact that humans, along with other higher animals, are equipped with the apparatus to develop the concept of a discrete physical object, on suitable triggering experience of the world. Oneness is inseparable from objecthood. An object without the property of oneness is inconceivable. The concept of '1' is an abstraction from the concept of an individual object.

This idea that the concept of oneness is abstracted from the notion of an individual object seems obvious and natural, but Frege argued strongly against it, as against the idea that higher numbers are properties of aggregates. I discuss Frege's arguments on this matter below. Arguing with Frege himself, rather than with his commentators, such as Wright (1983) and Resnik (1980) is, I believe, the best way of clarifying the main issue, since Frege expressed himself in the *Grundlagen der Arithmetik* with a clarity and strength which seem to me to have been muddied and sapped by more modern scholarly critics and apologists.

Frege's first specific argument against seeing oneness as a property of individual objects is:

> It must strike us immediately as remarkable that every single thing should possess this property [of oneness]. It is only in virtue of something not being wise that it makes sense to say 'Solon is wise.' The content of a concept diminishes as its extension increases; if its extension becomes all-embracing, its content must vanish altogether. It is not easy to imagine how language could have come to invent a word for a property which could not be of the slightest use for modifying the description of any object at all. (1950, p. 40e)

In this, Frege neglects the distinction between objects and stuff, between count and mass terms. The implication in this passage is that if the extension of a concept becomes 'all-embracing', then it becomes a set which contains all objects in the world (or universe of discourse), and therefore contains everything that one might want to refer to. If there is nothing that *one* does not modify, so the argument goes, modification by *one* is uninformative. But *one* does not modify mass terms. *One water*, *one air*, and so on are not used to refer to physical objects. The use of *one* can be informative. Imagine someone uses a new word, say *gadroil*, which is unfamiliar, in any of the following contexts:

4.1.1 (a) There's just one gadroil left in the box
(b) There's just some gadroil left in the box
(c) All he gave me was one lousy gadroil
(d) All he gave me was some lousy gadroil

(Assume unstressed *some* 'sm'.) From the context in which *gadroil* is used, one can tell a lot about its meaning, including whether a gadroil is an object, or gadroil is a substance.

This argument raises ontological questions about whether there is anything but objects. The ontology apparently reflected in natural languages is rich enough to countenance the existence of stuff, which, although it may be partitioned off into objects (lumps, clouds, drops, particles) is not itself *an* object. To the extent that the present study is concerned with ontological questions, it is concerned with them *as reflected in natural language*. I take it that ordinary language reflects the structure of the interaction between human consciousness and the external world, and that the entities referred to in natural language referring expressions are a product of this interaction. (This view is intended to be broad enough to permit the existence of non-physical entities, which might arise through language itself being treated as an external object of consciousness.) Frege's attitude to the evidence of natural language in his discussion of number is quite opportunistic; he cites linguistic examples which seem to go his way, and he dismisses cases which seem to go against him.

The view I take of the ontology reflected in language is well expressed in the following passage from Link.

> Our guide in ontological matters has to be language itself, it seems to me. So if we have, for instance, two expressions *a* and *b* that refer to entities occupying the same place at the same time but have different sets of predicates applying to them, then the entities referred to are simply not the same. From this it follows that my ring and the gold making up my ring are different entities. (1983, pp. 303–4)

The word *one* can be used to modify *ring* but not *gold*, because a ring is an object, whereas gold is not.

Frege's other main argument against oneness as a characteristic property of individual objects is an appeal to the incredibility of the idea of animals, such as dogs, possessing a concept of oneness, since, clearly, animals are capable of distinguishing individual objects. 'The point is strictly this: is the dog conscious, however dimly, of that common element in the two situations which we express by the word 'one', when, for example, it first is bitten by one larger dog and then chases one cat? This seems to me unlikely' (1950, p. 42e). It seems to me unreasonable to insist on consciousness of a concept as a criterion for its possession. Consciousness (as opposed to possession) of concepts in humans, let alone dogs, is in any case hard to determine, and I think that in the case of Frege's argument, the matter comes to little more than a reluctance to apply a certain honorific to 'brutes' Obviously humans know more about numbers than dogs do. But similarly a fluent French speaker knows more (about) French than a schoolboy who has learnt just one word of French. Should we deny the schoolboy's knowledge of the one French word he does know, just because he is ignorant of the rest? Whether the word is *merci* or *merde*, the boy can make at least some impact on the French-speaking world. The difference between the one word and the whole language is at least as great as the difference between a dog's apparent grasp of oneness and an adult human's grasp of numbers. Just as it seems clear that one can conceive of someone knowing a single word of a language, it does not seem unreasonable to envisage an animal having some kind of knowledge of just one number. Man, after all, evolved from brutes. The copious literature on number recognition in animals does not in general hesitate to use terms like 'concept' in relation to the limited capacities of animals in this domain. (For a tip of this iceberg, see Davis and Memmott, 1982; Thomas et al., 1980; Wesley, 1961; Salman, 1943.)

4.2 Numerals as Collection-denoting Expressions

Numeral words (with values of about 3 and over) occur first, I have claimed, in the conventional counting sequence. When recited in this sequence, they are not integrated into the syntactic structure of sentences, and are not, during the act of counting, used to make any assertion or commit the speaker publicly to the truth of any proposition. The counting procedure is a tool used in the construction of a collection from an aggregate. Its utility also lies in the possibility of calculating with it some conclusion about the constructed collection and the original aggregate, a conclusion which one might normally wish to assert publicly. This conceptual separation of the act of counting from any conclusions which might be drawn from it appears clearly from psychological research.

> In the Fuson and Mierkiewicz (1980) study, some children recounted sets as many as seven times in response to each repeated question of *How many blocks are there?* rather than giving the final word from the count. In such cases the *How many?* question seems to function as a request for the counting act rather than as a request for the information gained from the counting act. (Fuson and Hall, 1983, p. 64)

In phylo- or glossogenetic terms, once a recited sequence has been adopted and is conventionally used to draw conclusions about the cardinalities of collections, the need to express these conclusions prompts a need to integrate the words of the conventional counting sequence into the syntactic and semantic systems of the language concerned.

The previous chapter claims to explain the emergence from the counting ritual of a set of signs corresponding to a (short) continuous sequence of exact numbers, starting at 1. So we have a lexicon specifying numeral words and the associated number meanings. In that chapter, the specification of these 'meanings' was only given somewhat figuratively in terms of the Saussurean sign diagram. In the last section I have just argued that the meanings of numerals (above 1) should be conceived as properties of aggregates and collections. The ways in which these numeral meanings contribute to the meanings of the larger (non-numeral) phrases in which they become embedded have not yet been touched upon.

Linguists sometimes regard the goal of semantics as being to provide a mapping between natural language expressions and 'semantic representations', where the latter are often of a form identical to, or closely resembling, logical formulae. The semantic representation of a sentence is often thought of as a mental representation of its meaning, or sense. This approach has been criticized by Lewis (1972, pp. 169–70), *inter alia*, as merely providing yet another language, 'semantic markerese', which needs to be interpreted. The approach was followed in Hurford.

> I assume the semantic representation of any positive whole number *n* to be *n* marks on whatever material medium we can agree to talk about. ... the semantic representation of *one* is /, that of *two* is //, that of *three* is ///, and so on *ad infinitum*. In these terms, the semantic representation assigned to *one million* would be a million marks on the medium and the semantic representation of *zero* would be a complete absence of marks. (1975, p. 21)

In an obvious sense, these semantic representations were intended as models of the corresponding countable collections of things, and so perhaps a Lewisian critique of this particular form of 'semantic markerese' would be less severe, in that the mapping from collections of strokes on paper to collections of objects would be an extremely straightforward matter of one-to-one pairing. In Hurford (1975), a compositional semantics, involving addition and multiplication, was provided which assigned to complex numerals semantic representations with the appropriate numbers of strokes. Thus, the theory managed successfully to account for synonymy relationships, such as that between *eleven hundred* and *one thousand one hundred*, by assigning identical semantic representations (groups of marks) to synonymous expressions. And translation equivalents in different languages, for example *eighty* and French *quatre vingts* were also, appropriately, assigned identical semantic representations.

But this extremely simple representational language for numbers {/, //, ///, ////, ...} clearly has no direct psychological counterpart, except perhaps for the very lowest numbers up to about 3. There is no sense in which a speaker stores a million specially designated images, representations, markers, patterns, or whatever, to represent the meaning of *million*. The denotational semantics to be used in the present book avoids postulating representations of this sort, by talking directly in terms of collections of objects. In retrospect, the adoption of the ///... representations seems to have been a way of *avoiding* talking about the real denotations of numerals in terms of collections of objects. Once one undertakes to relate expressions to their real denotations, simply by talking of collections of objects, the need for such semantic representations actually falls away. Of course, marks on paper are themselves merely real-world objects of a particular kind, and an approach which treats them as specially significant in relation to the notion of number is guilty of a curious kind of 'paperboundedness'. The notion of 'whatever material medium we can agree to talk about' in the above quotation is entirely superfluous; since we have the real world to talk about, who needs a medium?

In writings on the syntax and semantics of number, plurality, and numerals, the standard 'Arabic' decimal place-value notation has enjoyed a special status as a component of semantic representations (for example in Bartsch, 1973; Hellan, 1980; Kempson and Cormack, 1981). Bartsch, for example, gives the following as a semantic representation of *Several hundred men are entering the arena*:

$$(\exists \mathcal{X})\ (\mathcal{X} \subset \{X\colon X \subset man' \ \&\ f_P^M(X) = 100\} \ \&\ f_P^M(\mathcal{X}) > 2$$
$$\&\ (\forall X)(X \in \mathcal{X} \longrightarrow (\forall x)(x \in X \longrightarrow \textit{enter the arena}'\ (x)))).$$

Putting aside the non-numerical symbols, the inclusion of the Arabic components '100' and '2' in this formula is a more objectionable case of semantic markerese than the stroke language {/, //, ///, ...}. Similarly, Kempson and Cormack (1981, p. 292) give a schematic logical form '$\exists\ S_n\ V_s\ P_s$ where n is a natural number'. This forgets the difference between numeral and number. The oversight is endemic in the literature. In modern literate cultures the Arabic notation seems to have become so familiar that it may be regarded as primitive and beyond analysis. But of course it is just another linguistic system, more useful for doing sums with than orthographic representations of ordinary spoken language, but no less in need of semantic interpretation. To say that *five* 'means 5' explains nothing to anyone's satisfaction, outside the English language classroom. And the compositional semantics of complex numeral expressions, such as *three hundred and sixty five* is in no way explicitly accounted for by the simple production of paraphrases, such as '365'.

The Arabic notation was preceded by centuries (perhaps millennia) of the use of spoken numerals. The evolution of a developed understanding of number proceeded through the growth of spoken systems, which should be the primary focus of study in any investigation of language and number. Many languages (for example Hawaiian, Yoruba) had ways of expressing exact numbers as high as the hundreds of thousands with constructions clearly interpreted by addition, multiplication, and subtraction, even though there was no way of writing these numbers down. These advanced numeral systems were totally oral.

I shall discuss the semantics of constructions involving numerals in extensional, or denotational, terms. That is, I am concerned

with how numeral expressions may be mapped onto features of a model, in particular onto collections and aggregates. I adopt a standard terminology according to which the denotation of a predicate is a *set*, for example the denotation of *cat* is the set of all cats, the denotation of *red* is the set of all red things, and so on. The denotation of *five cats* will be taken to be the set of all collections of cats with just five (cat) members. In general, I adopt a non-modal way of talking about collections, so that collections are deemed to have an existence distinct from the aggregates from which they are (can be) constructed. Thus the denotations of numerals and plural terms are said to be sets of collections. But to maintain the connection between numerals and physical objects (aggregates) I adopt the convention given as (4.2.1) below.

> *4.2.1* If a collection falls in the denotation of a numeral, then so does the aggregate from which it is constructed.

To avoid cumbersome phrasing, this convention will for the most part simply be assumed, and the denotations of numerals and plural terms will be referred to in shorthand as sets of collections. The treatment, though precise and explicit, will be relatively informal, avoiding the full formidable apparatus of modern formal semantics. The crucial thing is that the development of a system be described in such a way as to show how its users interact with the world in profitable ways. A method (function) telling, for a given entity, whether it is or is not a collection with some specific number of members is obviously a useful acquisition. It is this strict connection with the world that gives numeral systems their great usefulness, as the foundation for mathematics and hence for natural science, which interacts with the world in such spectacular ways.

The preferred grammatical position of a numeral is as a nominal modifier, as in *those five angry men*, but numerals are also common enough, as I shall show, used as grammatical predicates. I shall develop the view that numerals in either of these grammatical positions are expressions denoting sets of collections, and that the denotations of the larger phrases in which they are embedded are arrived at by the operation of set intersection for head-modifier constructions, and by the truth/falsehood of set inclusion for subject-predicate constructions. As the systems of expressions

I am interested in generally handle just the positive integers, I talk in terms of denotations which are sets of *collections*, rather than sets of sets. I use 'collection' here in the sense of 'finite non-null set'. Because of the way in which numeral words arise from the activity of counting out collections of objects, a particular numeral word becomes a predicate whose extension, or denotation, is a set of collections which can be put in a one-to-one correspondence with each other, that is all collections with some particular, exact, cardinality. (The meanings of ordinal numerals will be discussed in Section 4.4.)

Benacerraf (1965), following Frege in this particular, argued that numerals cannot be class-denoting predicates. His arguments rest on a sketch of some grammatical characteristics of English numerals, followed by an appeal to 'the traditional first-order analysis of sentences such as *There are seventeen lions in the zoo*' (pp. 60–1). He writes that numerals 'differ in many important respects from words we do not hesitate to call predicates' (pp. 59–60). There is an obvious fallacy in this line of argument. Cabbages differ in many important respects from (other) objects we do not hesitate to call vegetables, but this is not evidence that cabbages are not vegetables. Benacerraf continues:

> Probably the closest thing to a genuine class predicate involving number words is something on the model of 'seventeen-membered' or 'has seventeen members'. But the step from there to 'seventeen' being itself a predicate of classes is a long one indeed. In fact I should think that pointing to the above two predicates gives away the show – for what is to be the analysis of 'seventeen' as it occurs in those two phrases? (1965, p. 60)

No problem. There can be complex predicates composed from simple, atomic predicates. *Has red hair* and *red-haired* are such complex predicates, but we would not wish to deny that *red* and *hair* are predicates.

Benacerraf notes the 'similarity of function' between numerals and words such as *many*, *few*, *all*, *some*, *any*, traditionally seen as quantifiers. [This particular fact is also cited by Bostock (1974, p. 4), arguing the same point as Benacerraf.] Certainly there is a similarity of function, but this is no argument that numerals are not predicates denoting sets of sets. The similarity would militate against analysing numerals as *first-order* predicates which often

correspond to adjectives, like *red*. But nobody suggests that numerals are first-order predicates. First-order calculus quantifiers, such as $\forall$ and $\exists$, are in fact equivalent to second-order predicates denoting sets of sets, a fact brought out prominently by Barwise and Cooper (1981).

Quantifiers denote families of sets

> Quantifiers are used to assert that a set has some property. $\exists x S(x)$ asserts that the set of things which satisfy $S(x)$ ... is a nonempty set. That is the set of individuals having the property S contains at least one member. $\forall x S(x)$ asserts that the set contains all individuals. Finite $x S(x)$ asserts that the set is finite. It is clear that a quantifier may be seen as dividing up or partitioning the family of sets provided by the model. When combined with some sets it will produce the value 'true' and when combined with others it will produce the value 'false'. In order to capture this idea formally, quantifiers are taken to denote the family of sets for which they yield the value 'true'. (1981, pp. 163–4)

The similarity of function between numerals and words such as *all* and *some* surely, then, provides an argument *for* numerals being predicates denoting sets of sets. Bostock (and perhaps Benacerraf also) adheres to a different tradition in his view of quantifiers. 'I take it that no one will say that the word "some" sometimes names the class of some-membered classes or that the word "most" sometimes names the class of most-membered classes' (1974, p. 5). But this is what Barwise and Cooper, and their followers in research on generalized quantifiers, *do* say. The argument here rests on a basic difference between two research programmes, and one will therefore need long hindsight to judge between the two positions.

In fact there are clear grammatical differences between *all* and *some* and the numeral words. For example:

4.2.2 I gave them five books each
*I gave them all books each
*We are five here
We are all here
I have twenty-five thousand books
*I have twenty all thousand books

All five of my sisters are here
*Five five of my sisters are here
*All all of my sisters are here

Such examples could easily be multiplied in English. Barwise and Cooper state 'the familiar ∀ and ∃ are extremely atypical quantifiers' (1981, p. 260). Klein (1979) justifies a distinction between the 'classical' quantifiers, such as English *some, each, all* and words like *many* and *few* which he analyses as measure adjectives. Note examples such as *The problems are many* and *My friends are few*. McCawley (1981, pp. 103–4) points out the atypicality of ∀ and ∃ as quantifiers. And Prior (1985, p. 244) shows that there is across languages a syntactic distinction (albeit a fine one) between words for *all* and *some* and words for *few, many* and the numerals. The two classes occupy adjacent positions at one end of a hierarchy (mainly involving adjectives) predicting word-order relative to a head noun.

Benacerraf goes on: 'the nonpredicative nature of number words can be further seen by noting how different they are from, say, ordinary adjectives, which do function as predicates. We have already seen that there are really no occurrences of number words in typical predicative position, the only putative cases being along the lines of [*The lions in the zoo are seventeen*], and therefore rather implausible' (1965, p. 60). Personally, I do not find *The lions in the zoo are seventeen* too bad as an English sentence. And numerals can be used transparently as grammatical predicates in many languages. *Vous êtes quatre* in French clearly predicates fourness of a group of people, although this is most naturally translated into modern English with a somewhat disguised predication, as in *There are four of you*. But in French, *Vous êtes quatre*, with its straightforward use of a numeral as a grammatical predicate, is the only easy way of expressing this particular meaning. Similarly in German –*Wir sind vier* is clearly preferable to the dubious *Es gibt vier von uns*. The English of the Authorized Version of the Bible permitted even very complex bare numerals in grammatical predicate position. For example:

4.2.3 And his host, and those that were numbered of them, were three score and fourteen thousand and six hundred. (Numbers, 2.4)
All they that were numbered in the camp of Dan were an hundred thousand and fifty and seven thousand and six hundred. (Numbers, 2.31)

There are many more such examples in the Authorized Version. If it should be argued that this reflects the influence of the original Hebrew, that adds another language to the list of those allowing numerals as grammatical predicates. Givón (1972) gives evidence of numerals in ChiBemba (a Bantu language) used as bare predicates, and suggests an embedding process relating the predicate use to the nominal modifier use.

> The semantic arguments for deriving numeral modifiers from numeral predicates of embedded sentential modifiers parallel those given above for adjectives, with sentence (c) below seen as incorporating the meaning of the embedded (b) into that of the 'matrix' (a):
> (a) abaana baleeboomba 'the children are working'
> (b) abaana babili 'the children are two'
> (c) abaana babili baleeboomba '(the) two children are working'. (1972, p. 22)

Dixon writes of Fijian numerals: 'Like verbs, all numbers may be head of an intransitive predicate, e.g. ... *sa rua a waqa yai* there's two of these boats (lit: these boats are two)' (forthcoming). Barker (1964, p. 264) glosses a Klamath expression as 'My grandsons are nine'.

Languages give ample evidence of the use of numerals as 'clothed' (rather than bare) grammatical predicates, as in the examples below, in which the predication of cardinality is accompanied by predications of category membership, for example as friends, secretaries, and so on.

4.2.4 They are two of my best friends
We are three secretaries in the University
We are three soldiers
You are three experts
They are four idiots

Logically, I would argue, the numerals here are predicates. *We are three soldiers* seems to be equivalent to a conjunction of *We are three* and *We are soldiers*. [McCawley (1981, pp. 429ff) proposes a similar analysis of numerals as predicates.]

Benacerraf's arguments against numerals as predicates continues:

> The other anomaly is that number words normally outrank *all* adjectives (or all other adjectives, if one wants to class

them as such) in having to appear at the head of an adjective string, and not inside. This is such a strong ranking that deviation virtually inevitably results in ungrammaticalness:

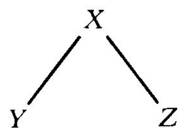
(6) The five lovely little square blue tiles

is fine, but any modification of the position of 'five' yields an ungrammatical string; the farther to the right, the worse. (1965, p. 60)

In Section 5 of the next chapter, a detailed explanation will be given for the ordering of numerals outside attributive adjectives in the noun phrase, which is a linguistic universal. The explanation to be given rests on the analysis of numerals as expressions denoting sets of collections.

I shall assume a relationship between syntactic structure and its compositional syntactic interpretation such that the question does not arise whether adjectives and numerals are logically predicates as opposed to nominal modifiers, in the sense in which this distinction is made and discussed by Kamp (1975), and Hoepelman (1983). I assume that for any syntactic structure as in (4.2.5), there will be a specified semantic operation yielding the denotation of the whole (X), as a function of the denotations of the parts (Y, Z). (Problems of intensional meaning do not arise in the structures to be discussed here.)

4.2.5



I do not insist that one of the constituents be itself the name of the required function. A typical formal semantic treatment is to say that a predicate (such as a common noun, or adjective) names a function from entities to truth values. In such accounts, in the semantic interpretation of a subject-predicate sentence, the predicate function is applied to the denotation of the subject (an entity, if the subject is a referring expression), yielding an appropriate truth value. But this makes a unified account of adjectives difficult, since one has to say that adjectives used in grammatically predicative position name functions from entities to truth values, while adjectives used as modifiers of nominals name functions from nominal denotations (sets) to nominal

denotations (sets). This approach to semantics assumes that, of two sister constituents in a construction, one can always be clearly identified as a function and the other as an argument. But this does not always seem to be possible. In *red cow*, for instance, the denotation of the whole is the intersection of the denotations of the two word-level parts. Intersection is commutative; there is no semantic reason for claiming that *red* here names a function applied to the set of cows, any more than there is reason to claim that *cow* names a function applied to the set of red things. Prior (1985) makes this point by noting that the mathematical notations '$\log(x)$' and 'xy' are different in an important way, in that in '$\log(x)$' the constituent 'log' is clearly the name of an operation, while in 'xy' the relevant multiplication operation is merely implicit.

The semantic interpretation rule that I associate with a subject-predicate structure, as in *London is dirty* or *They are three*, is a function from a pair (consisting of an entity and a set) to a truth value, for numerals and adjectives alike, such that the function returns 'TRUE' if and only if the entity (denoted by the subject) is a member of the set (the denotation of the predicate). Thus, for *They are three*, *they* denotes a particular collection and the denotation of *three* is the set of all collections of three individuals. We can describe the semantic effect of a linguistic form for plurality, for example the English {-s} morpheme, thus: the denotation of a noun combined with this morpheme is the set of collections of objects in the denotation of the singular noun. So, for example, *brick* denotes the set of all bricks, and *bricks* denotes the set of all collections of bricks. And for numerals and many adjectives in nominal modifier positions, the interpretation rule associated with a modifier-head structure, as in *red cow* or *three boys*, is a function from pairs of sets to the sets which are their intersections. Thus, *boys* denotes the set of all collections of boys; and the denotation of *three boys* is the intersection of this set with the denotation of *three*, that is it is the set of all collections of three boys. That is, the application of a numeral takes a plural nominal denotation – a set of collections of objects of some particular type – to one of its proper subsets – a set of collections of objects of that type with a particular cardinality.

The prefixing of a definite determiner as in *the three boys* has the effect of selecting one particular collection of three boys, whose identity is determined by the discourse context, so that such definite expressions denote entities (that is particular collections). With numerals in English, a plural indefinite determiner is phonetically null. Link (to appear) also analyses phrases such as *three apples* as containing a phonetically empty indefinite plural determiner. An indefinitely determined plural noun phrase can denote a particular collection, but its identity is not a function of discourse context. Link (1985) points out that the addition of *some*, as in *some fifty men* adds a sense of approximateness. This *some*, which is pronounced with an unreduced vowel, is to be distinguished from the indefinite plural determiner with a reduced vowel or syllabic nasal, often conventionally written by linguists as *sm*. *Sm* cannot appear with numerals.

Numerals and adjectives can be used *both* as sentential predicates *and* as nominal modifiers, and a semantic theory needs to relate to the fact that the *same* item occurs in *different* syntactic positions. The account given here assigns constant denotations to numerals and adjectives independently of the grammatical contexts in which they occur. The different grammatical constructions in which numerals and adjectives occur are associated with particular semantic interpretation rules, yielding, as appropriate, sentence denotations (truth values), or nominal denotations (sets). In what follows, I shall concentrate on the more typical use of numerals, that is as embedded in plural nominal phrases. I hope that the core of the account given will be applicable, with the appropriate additions, to accounts of all types of noun phrases containing numerals, such as definite referring expressions (for example *those six men*), sentences with mixed quantification (e.g. *two examiners marked six scripts* – see Kempson and Cormack, 1981; Tennant, 1981), noun phrases with mixed quantification (for example *all three men* – see Link, 1985), expressions used generically (for example *six men will fit in the back of a Ford*), intensionally interpreted expressions (for example *I'm looking for six men*), and so on. I will not pursue these complexities here. For convenience, I will mostly mention only the (relatively) more concrete denotations, namely sets of objects and collections, for the moment treating 'object' and 'collection' as basic ontological categories whose nature is intuitively clear.

In assuming that undetermined plural nominals denote sets of collections of objects, I exclude from consideration what Carlson (1977) calls 'bare plurals', which are plural nominals unaccompanied by any determiner or quantifier, excepting predicate nominals. Thus in *Beavers build dams*, both *beavers* and *dams* are bare plurals, while in *Beavers are mammals* only *beavers* is a Carlson-type bare

plural, and *mammals* is a plural predicate nominal. Carlson argues that bare plurals denote 'kinds of objects', ontologically to be distinguished from sets of collections of objects and I have no quarrel with that analysis here. The plural nominals that numerals combine with are not Carlson-type bare plurals, but rather the plural nominals that can occur as predicates or in (definitely or indefinitely) determined noun phrases (NPs). Thus the instances of *beavers* in *Three beavers appeared* and in *They are three beavers* are not Carlson-type bare plurals, and denote not kinds, but, I assume, sets of collections. Carlson argues for the desirability of a unified analysis of bare plurals in particular, but in thus divorcing bare plurals from other plural nominals, he ends up, ironically, with a conspicuously disunified analysis of plurals in general. Perhaps a way can be found to unify the general account of plurals, but I shall not make that my business here.

In some languages, numerals may modify *singular* nominals, rather than plurals. In many such cases it is not plausible to resort to the *ad hoc* postulation of a phonetically null plural morpheme used just with the numerals in question. Such nominals unmarked for number could be regarded as vague (or ambiguous) between singular and plural meanings. The denotation of such an ambiguous singular/plural term would be the union of the expected singular and plural denotations, that is the union of a set with its power set. When combined with a plural numeral, for example one meaning 3, the set intersection operation would yield the desired set-of-collections denotation for the whole NP; and when combined with a singular numeral, that is one meaning 1, set intersection would yield the desired set-of-individuals interpretation. This assumes that the singular numeral in question is not a kind of determiner.

I wish to avoid, if possible, postulating semantic representations in forms like that of predicate calculus, especially where these do not reflect natural language structure and word order. Thus, allowing the use of collection-denoting predicates in traditional predicate calculus-style formulae, one might suggest (4.2.6) as a representation of *Seven came*.

4.2.6 $\exists X[(\text{SEVEN } X) \ \& \ (\text{CAME } X)]$

That is, *Seven came* expresses the fact that there was a collection *X*, and that this collection was seven in number and came. But there is nothing in *Seven came* corresponding to the three instances of the variable X in (4.2.6); nor is there any indication of conjunction. Quite cumbersome rules would be needed to map the English sentence onto this putative semantic representation. I prefer to explore an approach in which the denotations (truth values) of natural language sentences are derived from representations much closer to the surface forms of the sentences themselves. I take a cue from Barwise and Cooper (1981). They translate

4.2.7 Some person sneezed
Every man sneezed
Most babies sneeze

as the logical representations

4.2.8 (some person) sneeze
(every man) sneeze
(most babies) sneeze

They comment 'These sentences will be true just in case the set of sneezers (represented by ... *sneeze* contains some person, every man, or most babies, respectively' (p. 165). Similarly *Seven came* is true, in my treatment, if some member of the denotation of *seven* (that is some collection of seven things) is a member of the set of things which came. Recall the phonetically null indefinite determiner postulated above, whose effect is to select one member of the denotation of the determined phrase. The entailment that all the individual members of the collection came can be derived from a logic for plural objects (collections), such as that developed by Link (1983) (referred to in more detail in Chapter 5, Section 6). I will not pursue detailed differences between Barwise and Cooper's proposals and mine. Mine are simpler and quite possibly less adequate than theirs for handling complex cases.

Another example where existential quantification over collections might be proposed would be *There are nine planets*, which can be given a translation such as (4.2.9).

4.2.9 $\exists X[{*}\text{PLANET } X \ \& \ \text{NINE } X]$

Here '*PLANET' is a plural predicate meaning 'is composed of planets' along the lines proposed by Link (1983) see Chapter 5, Section 6 below). (4.2.9) says that there is a collection composed

of planets and numbering nine. Representations such as this are linguistically unnatural. They are also subject to a subtle problem of interpretation if the quantification is taken to be over aggregates, as opposed to collections. The problem arises from the fact that a given aggregate can have different cardinalities depending on the concept under which it is brought. Unless some kind of binding between the predicates *PLANET and NINE in (4.2.9) is indicated, this formula could apparently also be true of any situation where there is an aggregate which consists of planets and which can be divided into nine parts, *whether these parts are planets or not*. This is because the simple proposition *PLANET(a) just says that *a* consists of planets, and the simple proposition NINE(a) just says that *a* has nine parts or members. In short, we have the problem of Frege's boots again.

It is possible to devise more complicated notations, still treating numerals as predicates, which indicate the necessary connection between the analysing of an aggregate into parts and the conceptual category (for example 'planet', 'apple') which is relevant to that analysis. Or one might find this enough to push one to enrich one's ontology with collections, as well as aggregates, in which case the problem does not arise, since collections, as opposed to aggregates, only ever have a single cardinality. On the other hand, one might take this difficulty to be an argument in favour of analysing numerals as subtypes of the existential quantifier (over aggregates). Subscripted existential quantifier treatments of numerals are frequently suggested (for example in Benacerraf, 1965, p. 61; Resnik, 1980, p. 126; Altham, 1971, p. 45; Field, 1980, p. 21; Bunt, 1985, p. 101). In such treatments, there is an infinite number of separate quantifiers, one for each number. Representing these as the existential quantifier with a subscript integer, *There are nine planets* would, for example, be rendered as:

4.2.10 $\exists_9 x \ [\text{PLANET } x]$

While such a treatment avoids 'the problem of Frege's boots', it is again linguistically very unnatural. The parts of the logical formula do not match up neatly with corresponding constituents in the natural language sentence of which this purports to be a translation. And such a treatment absolutely precludes the possibility of developing a compositional semantic account of the syntactically complex numerals in natural languages (for example *sixty four, two hundred and three*, and so on).

To try to meet these problems by searching for alternative logical formulae which somehow avoid them is actually superfluous, if one can simply deal directly with denotations. There is no need to find a representation in some logical language for *There are nine planets*, or any other sentence, if one can map the sentence directly onto a truth-value in relation to a model by means of semantic rules interpreting linguistic constituents in terms of their denotations. Thus the denotation assigned to *nine planets* is the intersection of the set of all collections of planets and the set of all collections of exactly nine things, that is it is the set of all collections of nine planets. The semantic interpretation rule associated with the existential *there is/are* construction assigns the value TRUE if and only if the universe of discourse contains at least one member of the denotation of the constituent noun phrase.

Collections which satisfy numeral predicates can themselves be quantified over. Thus a sentence such as *This elevator can take seven persons* might be given a translation like (4.2.11).

4.2.11 $\forall x[(\text{*PERSON } x \text{ \& SEVEN } x) \longrightarrow (\text{TAKE } e\ x)]$

where '*PERSON' is a Link-style plural predicate meaning 'is composed of persons' and '*e*' is the logical name of the elevator. (4.2.11) says that if anything is a collection composed of persons and numbering seven, then the elevator can take it. It is interesting that such universal quantifications over collections with specified cardinalities typically involve modal notions, as in the example given, where TAKE translated *can take*. An example without such a modal ingredient might be *Twenty soldiers make a platoon*, translated in predicate logic as:

4.2.12

$$\forall x[(\text{*SOLDIER } x \text{ \& TWENTY } x) \longrightarrow \text{PLATOON } x]$$

This attempts to say that any collection of twenty soldiers is a platoon (which may or may not be factually correct).

These logical representations, as before, are linguistically highly unnatural, and fall prey to the problem of Frege's boots if construed as quantifying over aggregates rather than over collections. (4.2.12), for example, in fact only says that if an object (such as an army) is composed of soldiers, and if it can be divided into twenty things (such as regiments), then it is a platoon. Thus the representation falls short of what is intended. Such sentences involving generic and modal notions are too complex to explore in detail here, but I assume that a method can be found of assigning denotations and ultimately truth values without postulating such unnatural and problematic representations. [Link (1985) discusses quantification over collections, specifically in connection with the German word *je*, which often expresses such quantification.]

The question arises why languages actually seem, on the whole, to prefer to use numerals as 'downgraded' predicates (nominal modifiers) rather than as full grammatical predicates. All languages have means of downgrading predications, for example of forming nominal modifiers corresponding to full predicate expressions. [The term 'downgrading' applied to predications is used by Leech (1969, especially pp. 26–8)]. Devices for downgrading predications include relative clauses, participial constructions, and attributive adjective constructions. And numerals conform to this pattern, being indeed most frequently used as nominal modifiers. Greenberg writes of the 'bare predication of numerals which is disfavored in many languages' (1975, p. 41). Certain other quantifiers, or terms denoting sets of sets, such as *all*, *some* and *most*, similarly tend to avoid grammatical predicate position.

One reason that immediately suggests itself for the infrequency of bare predication involving numerals, as in *They are three*, is the point, which Frege used against Mill, that unless it is clear what other predicate (concept) is involved, a particular aggregate can have many different cardinalities. Take the example *The army is six hundred*. Six hundred what? Divisions? Regiments? Battalions? Men? A nominal following the numeral specifies the particular concept with respect to which the analysis of the aggregate into members is to be applied.

Finally in this section I note that although numeral words indicate exact values, they are truth-conditionally compatible with real world situations involving lower values. That is, although numerals denote sets of collections which may be put in a one-to-one correspondence with each other, the aggregates which give rise to these collections may always be used to construct collections with lower cardinalities, simply by disregarding a bit of the aggregate. So if it is true that I saw three men, then I saw an aggregate from which could be constructed a collection with three members who were all men. And if I saw this aggregate, I saw a part of it which could be analysed into a collection of just two men. This gives rise to a valid inference from *I saw three men* to *I saw two men*, and generally to suggestions that the numerals indicate not exact values, but 'at least' values. The relation between a numeral and all lower-valued numerals is in this respect like the relation between the quantifiers *all* and *some*, at least for simple cases. In more complicated cases, such as those involving interaction with quantifiers, other numerals, and modals, the inference can go the other way. For example, *I managed to pack all my books into ten boxes* does not entail *I managed to pack all my books into five boxes*. I do not go into these complications here.

4.3 'Abstract' Interpretation of Bare Numerals

Frege's own alternative to the view that numbers are properties of aggregates is that numbers are abstract entities: 'Every individual number is self-subsistent object' (1950, p. 67e). Language does indeed provide the means to treat numerals like proper names, creating an impression of numbers as self-subsistent objects. This can be accounted for, but in a way which shows such constructions to be derivative of the constructions with numerals as nominal modifiers, prototypically used to refer to collections, or aggregates of objects.

Beside appearing as nominal modifiers, numerals also occur independently of any nominal, as NPs in their own right, with a structure as in (4.3.1).

4.3.1

```
NP
 |
Numeral
 |
```

Such bare numerals can occur as subjects and objects of sentences, as illustrated in (4.3.2).

4.3.2 (a) Q: How many cakes would you like?
A: I'd like three, please. (Two would be not enough, and four would be too many.)

(b) Seven is an odd number
Seven and five make twelve

The bare numerals in (4.3.2)(a) and (b) here are of apparently different types. The first type (a) is anaphorically elliptical for a full NP. So *three* stands for *three cakes*, *two* for *two cakes*, and so on, in the context of the given dialogue. It is recognized that there must be rules of discourse sanctioning such ellipsis under conditions statable in terms of surrounding linguistic and non-linguistic context. Such rules of ellipsis are well known for other types of expressions, e.g. verb phrases (*I wanted Shiela to kiss me but she wouldn't*), mass nouns modified by adjectives (*Bill ordered dark beer and I ordered light*), NPs after prepositions (*Shiela wanted it on the table, but I wanted it under*). So, on hearing, for instance, 'five', in a context where such ellipsis is permitted, a hearer in some sense tries to provide a missing nominal from the context.

The other type of bare numeral, given in (b) of (4.3.2) cannot easily be regarded as elliptical for a fuller NP. It would not seem immediately to make sense to try to interpret *Seven is an odd number* as a reduced form of, say, *Seven things is an odd number*. These are cases of an 'abstract' use of numerals, the kind of cases which give rise to talk of numbers as abstract Platonic objects. I think it is possible to give an account of such cases in terms of the cardinality of collections of objects, so that there is in fact no basic ontological difference between the ways elliptical and 'abstract' bare numerals may bear meaning. But there is an important *psychological* difference between bare numerals interpreted elliptically via context and the abstract cases.

Consider the following interesting dialogue between an experimenter (MH) and a child (Adrian).

> we had arrived at a situation where there were ten bricks on the box.
>
> MH: Let's just put one more in [does so]. Ten and one more, how many is that?
> Adrian: Er ... [thinks] ... Eleven!
> MH: Yes, very good. Let's just put one more in [does so].
> Eleven and one more, how many is that?
> Adrian: Twelve!

Five minutes later the following sequence of questions took place.

This time the bricks had been put away and there were no materials on the table.

MH: I'm going to ask you some questions.
Okay? How many is two and one more?
[No response.]
Two and one more, how many is that?
Adrian: Er ... makes ...
MH: Makes ... how many?
Adrian: Er ... fifteen [in a couldn't-care-less tone of voice].
(Hughes, 1986, pp. 38–9)

And again,

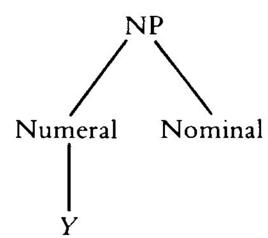
Adult: How many is two and one more?
Patrick (four years, 1 month): Four.
Adult: Well, how many is two lollipops and one more?
Patrick: Three.
Adult: How many is two elephants and one more?
Patrick: Three.
Adult: How many is two giraffes and one more?
Patrick: Three.
Adult: So how many is two and one more?
Patrick: [Looks adult straight in the eye] Six.
(Hughes, 1984, p. 9)

Meanings in a language can vary along a psychological dimension of perceptual/conceptual salience. We all know the distinction between abstract nouns, such as *ambition*, and concrete ones, such as *cat*. Typically, children acquire an understanding of many concrete nouns before they acquire an understanding of their first abstract noun. (See Clark (1979) for a summary of vocabulary acquisition studies, and Brown 1958, pp. 247–249, on the early acquisition of relatively concrete vocabulary.) On the scale of salience that I have in mind the denotata of concrete nouns are typically more salient than those of abstract nouns. But within concrete concepts themselves, there can be differences in salience. I claim that the denotation of a bare numeral word such as *five* is less salient than that of *five bricks*.

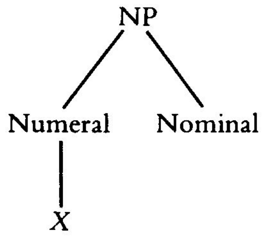
The scale of salience invoked here is a psychological scale, essentially involving the accessibility of referents by humans starting from contexts provided by their perceptions of their own physical spatio-temporal situations. Thus, since both *five* and *five* *bricks* denote sets of collections, one cannot claim that one denotation is in reality more concrete than the other. But it seems to be harder to conjure up in the imagination a collection of five unspecified things (that is not necessarily lollipops, nor elephants, nor giraffes, nor anything at all in particular – just *things*) than it is to envisage a collection of five things of a specified category (say, giraffes). The same is true of individual objects. 'Imagine a thing' is a much harder instruction to obey than 'Imagine a lollipop'.

The evidence of dialogues such as Hughes's above is that for the children in question, *three* as an NP is always elliptical: if they cannot supply a missing noun, they cannot interpret the question. Put another way, children find *five bricks* easier to understand than *five*. To a semanticist steeped in the Fregean tradition that the meaning of the whole is a function of the meanings of the parts, this is a paradox, and may perhaps even seem an impossibility: according to an obvious version of the compositionality principle, one could not grasp the meaning of *five bricks* without first grasping the meaning of the constituent *five*. But in fact it is possible to give a coherent account of the relation between the meanings of the parts and the meaning of the whole which makes an understanding of, for instance, *five bricks* logically prior to an understanding of *five*. The inductive generalization (I) of (3.6.2), proposed in the previous chapter as a piece of knowledge acquired by the child, can be seen as a rule for interpreting, or assigning a denotation to, undetermined NPs. Cast as a rule for interpreting such NPs, it might be rephrased as in (4.3.3).

4.3.3 If *X* is followed by *Y* in the counting sequence, the denotation of an undetermined NP of the form



is the set of collections which can result from adding an object in the denotation of the Nominal to a collection in the denotation of an undetermined NP of the form



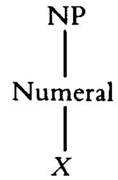
[Recall from (h) of (3.6.2) that placing an object into a collection results in a collection.]

The denotations of the very lowest-valued NPs, such as *one brick*, *two bricks*, and possibly *three bricks* are provided by other means. The child has perceptually representable concepts of oneness (singularity), plurality, twoness, and possibly threeness prior to learning the higher-valued numerals and their meanings. Plurality is a property characterizing collections in general. Twoness is a property of certain collections, pairs. The basic concrete denotation of *two bricks* is the intersection of the set of collections of bricks with the set of pairs of things in general, that is the set of pairs of bricks.

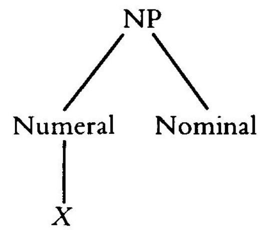
In rule (4.3.3) the denotation of an NP containing a numeral is expressed as a function of a conventional counting sequence. A speaker who did not know a counting sequence would not be able to apply this rule. Assigning a denotation to an NP containing a numeral is, in effect, the inverse of determining the cardinality of a collection of objects. Procedures actually used by speakers in practice for determining the number of objects in a collection invariably make essential use of a conventional counting sequence. Rule (4.3.3) thus provides a quite natural account of the knowledge underlying the ability to determine the cardinality of collections, and to interpret NPs containing numerals.

Rule (4.3.3) only tells its possessor how to interpret NPs containing numerals in construction with nominals, making reference to the counting sequence. It does not show how to interpret numerals in isolation. The interpretation of numerals in isolation can be accounted for by a further rule, which depends logically on the output, or prior application of rule (4.3.3). The rule for bare numerals is (4.3.4).

4.3.4 The denotation of an NP of the form



is the set of all collections in the denotation of an undetermined NP of the form



for any Nominal.

This second rule of interpretation cannot be practically used without the first rule interpreting NPs (4.3.3). This would account for the greater difficulty children have with *five* than with *five bricks*. Where it is possible to supply a missing nominal from context, a reconstructed NP can be interpreted by means of rule (4.3.3). But where such reconstruction is not possible, the bare numeral cannot be interpreted without rule (4.3.4). Acquisition of the second rule may be seen as a step of abstraction, a significant step towards an abstract conception of number.

This idea that an understanding of bare abstract numerals stems from a prior understanding of (Numeral Noun) constructions bears a resemblance to Goodman's (1952) proposal to account for the meanings of fictional nouns (such as *unicorn*, *leprechaun*). Though the denotations of the bare nouns are identical (the null set), the denotations of phrases containing them, such as *picture of a unicorn* and *picture of a leprechaun* are non-empty and distinct. Goodman's purpose was not psychological, but rather to suggest an extensional treatment of cases traditionally held to pose difficulties for an extensional view of meaning. He was not directly concerned with acquisition of concepts. But his proposal provides a plausible precedent, I believe, for claiming that access to an understanding of a grammatically simple form, such as a bare noun or a bare numeral, may sometimes proceed via prior understanding of a larger phrase containing it. This is in some sense a challenge to the Fregean compositionality principle.[(For a discussion of this principle in relation to learnability, see Chapter 10 of Hintikka and Kulas (1983).]

A theoretical point about rule (4.3.4) is that it makes the denotation of a bare numeral NP dependent on the nominal expressions that actually exist in the language in question and that can be modified by the numeral word concerned. A general consequence of this is that the notion of a countable collection is tied to the availability of nominal expressions, simple or complex, in a language. This seems right. There is no theoretical limit to the complexity of a nominal expression. An example of a somewhat complex plural nominal would be *piano-shaped things about as big as a cat, with little reddish green knobs on the underside, draped in scraps of Clan Stewart tartan*. If a class of objects is susceptible to description by some nominal expression, however complex, collections of such objects can be counted, but not otherwise. This point is close to Frege's insight that numbers belong to concepts; classes of objects that do not fall under a concept cannot be numbered. Frege would, of course, have rejected the association of 'concept' with 'possible nominal expression', since concepts were for him language independent. I believe that there are both language-independent concepts and concepts which are language-dependent by virtue of having been acquired via verbal definition (for example the concept of a bachelor). But even language-independent concepts, that is concepts acquired or acquirable independently of the corresponding predicates, are always in principle expressible by predicates in a natural language. Thus, in my account, Frege's insistence on the crucial role of concepts in the assignment of numbers is captured through a semantic denotation-assigning rule, in which predicate nominal expressions play a crucial role.

Could there be countable yet indescribable collections? That is, could there be a person whose linguistic and numerical abilities were in some sense the mirror image of the numeral-less Australian aborigines? Such a person would have, say, the separate concepts 'dingo', 'wombat', and 'koala', but, curiously, no words or nominal expressions corresponding to these concepts. Despite the linguistic lack, he could demonstrate his possession of

these language-independent concepts by elaborately differentiated behaviour towards the different species, behaviour that could not plausibly be put down to mere knee-jerk responses to the stimuli they present. (For example he might habitually draw pictures in the sand of dingos, but never of wombats; he might kill and eat wombats, but never koalas, even though they are easily caught and tasty.) Now this hypothetical person knows a numeral sequence and can count things (kinsmen, days to the next jamboree, and so on) for which he has nominal expressions. But can he count dingos, wombats, or koalas? Yes, clearly, in the sense that he could recite the counting sequence whilst pointing to a separate dingo with each word, stopping when he has pointed to each dingo in the pack in front of him. But, without the necessary nominal expression, he cannot report publicly the conclusion reached by his counting. In such a situation, one imagines that any normal person would quickly coin an appropriate nominal expression, so that he could report, for example, 'I saw six of those animals I draw pictures of'. With human languages, one cannot conceive of such expressions simply not being available. The expressive power of languages is well known to be such that any concept sufficiently differentiated to allow counting of the objects that fall under it is describable by some nominal expression.

A more specific, and peripheral, consequence of the dependence of numeral denotations on the nominals they modify is that in languages which use different numeral words to count different types of objects the denotations of particular numeral words would be restricted to the collections of objects in the denotations of the nominals with which they are specifically associated. 'There appear to be two dominant numeral systems in Kusaien [a Micronesian language]. One system is restricted to the counting of fish and to things related to fishing. ... Another system ... is restricted to the counting of other things than fish' (Vesper, 1969, pp 10–11). If such cases cannot be analysed as numeral classifier constructions (see Chapter 5, Section 4) or as agreement phenomena, this somewhat marginal problem may be likened to the lexicographer's favourite *addled*, whose definition includes the restriction 'of an egg', or to the Aristotelean example of *snub*, applicable only to noses.

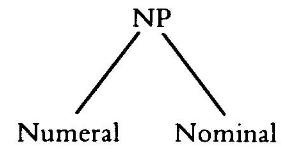
The provision of nominal expressions (predicates, not names) by language strongly facilitates the acquisition of number concepts, and subsequently allows for their widespread application.

That numbers can be applied to larger or non-physical collections – for example of molecules in a pencil, or stars in the sky, or of ideas, languages, and songs – is no objection to rooting their *acquisition* in the experience of small collections of physical objects. Knowing what bricks are, the exemplar *eight bricks* (plus knowledge of the counting sequence) allows a child to grasp what eight is in a particular instance. Having grasped what eight is in this instance, he can substitute any other plural nominal. If there is uncertainty in the correct application of any nominal, there will be difficulty counting the objects that fall under it, which is why it is not possible to give an exact number to the world's languages, or great philosophers.

The psychological difficulty, for a language acquirer, of grasping, for instance, *five* before grasping *five books* suggests an explanation for the non-invention of numerals in some communities. The child born into a numeral-possessing community has the advantage of being presented with both types of expression, that is, both bare numerals and NPs with numerals modifying nominals. The child grasps what he can first and progresses to more difficult types of expressions later. But this luxury is not available to the *inventor* of numerals, who has to acquire both the numeral signs and some rules for integrating them into linguistic structure more or less together. It is only when integrated into linguistic structure that the numeral signs become communicatively useful. The ordinary child language acquirer can memorize the counting sequence, trusting in the reassurances of his elders that there is some point to it. The point becomes clear to the child when he grasps rule (4.3.3) through observation of the use of numeral words to modify nominals. The inventor of numerals, on the other hand, has, in some sense, to have the foresight to provide this motivation for himself, to see one step ahead that a stable sequence of individual words will be useful in providing the basis for expressing things about collections of objects.

The fact that *five bricks* is more readily understandable by children than the bare *five* is accounted for by postulating the ordered acquisition of two semantic rules (4.3.3) and (4.3.4), the first of which relates to psychologically more concrete experience. But having once acquired the second, more abstract, rule, a child may reformulate or augment his knowledge of the interpretation of Numeral-Nominal structures by arriving at the following rule:

4.3.5 The denotation of an NP of the form



is the intersection of the denotations of the two constituents.

The addition of this rule may at first be essentially a reorganization of existing knowledge in more abstract terms. But it paves the way for interpretation of the numeral constituent in a way not directly dependent on a conventional counting sequence and without involvement of the interpretation of its sister nominal constituent. It is in fact necessary to postulate this rule in order to account for the subsequent development of constructions involving syntactically complex numerals, such as *seventy-two squares*. The conventional recited sequence of counting expressions can be extended into the hundreds, and so *seventy-two squares* can be arrived at as a correct description of some particular collection of shapes by laboriously counting up to *seventy-two*. Young children will count the little squares on graph paper in this way, without resorting to labour-saving calculations with abstract numbers. But adults can arrive at *seventy-two squares* as a correct description of the number of squares in a group without counting up to *seventy-two* and without involving the nominal *square* in the childish way implied by rule (4.3.3). (The development of syntactically complex numerals is taken up in Chapters 5 and 6.)

4.4 Ordinals

Languages often distinguish cardinal from ordinal numerals, for example English *twenty* verus *twentieth*. Frequently, the ordinals are morphologically derived from the corresponding cardinals, for example by the addition of a suffix (as, say, in French, German, or Tamil). Another method of indicating the ordinal as opposed to the cardinal is by a change of word order [as in Arabic (above 10), Japanese, and English]. Thus with certain nouns, one may say either *the fourth item* or *item four*, either *the fifth day* or *day five*. So a morphologically cardinal form may have either a cardinal or an ordinal semantic interpretation. We may define 'cardinal meaning' and 'ordinal meaning' as follows.

> *4.4.1* A numeral is used with cardinal meaning when applied to a class or set of objects, often in connection with a plural noun, for example *those five students*, a referring expression picking out on its occasion of use a collection whose cardinality is 5.

> *4.4.2* A numeral is used with ordinal meaning when applied to an individual object in an ordered sequence, often in connection with a singular noun, for example *the fifth student* or *student five*, referring expressions picking out on their occasion of use a particular student who is understood to occupy 5th position in an ordered sequence given in the context.

In the counting activity, the question arises whether the numeral expressions recited are used with cardinal or with ordinal meaning. It might seem obvious that in this case the numerals are used with ordinal meaning, since individual objects are pointed to in association with each numeral recited. But in languages which make a morphological distinction between cardinal and ordinal numerals, it is the cardinal forms which are used in the conventional counting sequence.

The counting activity is primarily a form of calculation by which an individual speaker arrives at a conclusion either about the cardinality of some set or about the position in an ordered sequence of some individual. Thus I can count to see how many people there are in the cinema queue, or I can count to find out my own position in the queue. Counting yields either a cardinal or an ordinal conclusion. Counting is not primarily a communicative act between a speaker and a hearer. Notions like 'assertion' and 'referring expression' are appropriate to communicative acts between speakers but it is not clear that in the act of counting the counter is making any assertion, committing herself to the truth of any proposition, or even referring to anything. Often the order in which the objects in a collection are pointed to when being counted is quite arbitrary, any order being equally effective if one is interested in the cardinality of the whole collection. But the *conclusions* arrived at by counting *can* be publicly asserted and thought of in terms of propositions, and such propositions are understood as involving particular numbers.

Thus the words used in the conventional counting sequence are neither inherently cardinal nor inherently ordinal, as far as the activity of counting itself is concerned. One need not ask the question, 'Why do we not count: first, second, third, fourth, ...?'. Brainerd (1973) discusses the question of whether cardinality or ordinality is more fundamental to number, both mathematically and psychologically. I have argued, in effect, that this question cannot arise, as the notions of 'collection' and 'sequence' are both necessary (and neither is sufficient) for a grasp of the full significance of numbers. The concepts of 'collection' and 'stable sequence of words' were both included in (3.6.1), the given apparatus relevant to the acquisition of numerals and number.

> It is natural to inquire at this point whether this subtle distinction between cardinal and ordinal number had any part in the early history of the number concept. One is tempted to surmise that the cardinal number, based on matching only, preceded the ordinal number, which requires both matching and ordering. Yet the most careful investigations into primitive culture and philology fail to reveal any such precedence. Wherever any number technique exists at all, both aspects of number are found. (Dantzig, 1940, p. 9)

The denotation of the cardinal numeral *five* is the set of all collections of five things. It is consistent with the general approach taken here to say that the denotation of the ordinal *fifth* is the set of all objects that are in fifth position in some ordered sequence. It might appear that an immediate problem with this is that it makes all objects fall into the extension of *fifth* (and any other ordinal, for that matter), since it is always possible to construct an ordered sequence in which some given object is the fifth member. In other words, if they have this denotation, ordinal numerals would seem to be totally uninformative. But in fact, out of context, so they are.

> *4.4.3* Ivan was (the) fifth

This sentence, with no context given, tells us nothing about Ivan that we did not already know. One wants to know *the fifth WHAT?* So try:

4.4.4 Ivan was the fifth person
Ivan was the fifth Ukrainian

Here, *Ukrainian* is more informative than *person*, which might legitimately be inferred from *Ivan*, but even with *Ukrainian* there is still something missing. A particular ordered sequence in which Ivan was the fifth Ukrainian is presupposed. (The sequence need not consist entirely of Ukrainians: there might be 100 people of all origins in a queue, with Ivan somewhere in the middle, but he could still be the fifth Ukrainian.) Finally, something like *Ivan was the fifth Ukrainian in the queue* is more complete in that it seems not to presuppose the existence of anything not mentioned in the sentence itself. So, to interpret an expression containing an ordinal numeral, a particular ordered sequence of objects must be present, explicitly or implicitly, in the context in which the expression is used. I will call such a sequence the 'context-given' sequence.

The necessary presence of a context-given sequence has the effect of making expressions containing ordinals definite. Compare the ordinal case with the others in the following:

4.4.5 We saw a queue. The fifth person was a Ukrainian.
We came to a house. The door was open.
We discovered a body. The head was missing.

In such mini-discourses, an object as yet unmentioned (for example a door) can nevertheless be introduced with the definite article because the context created by previous expressions (for example *house*) allows one to anticipate its existence. (The anticipation may sometimes be incorrect, as in the case of a house with no doors, or a queue with only four people in it.) In non-ordinal cases, the object whose existence is anticipated may not be unique: a house may have several doors. But in the ordinal case, what is anticipated, on the basis of some context-given sequence, is the existence of objects uniquely identifiable by their position in the sequence. Compare the following.

4.4.6 We saw a queue. *A fifth person was a Ukrainian.
We came to a house. A door was open.
We discovered a body. An arm was missing.

So expressions containing ordinal numerals are typically definite. Indefinite expressions such as:

4.4.7 A fourth man came.
I saw a fourth man.

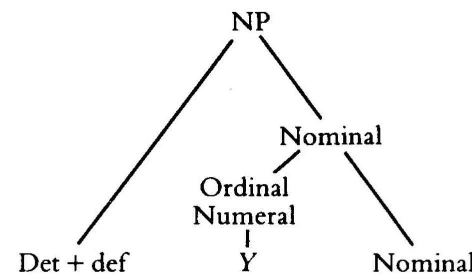
are in fact compressed ways of saying quite complicated things such as:

4.4.8
For the fourth time, { a man came / I saw a man } – a different man from the others.

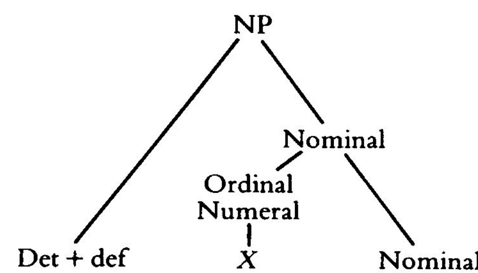
The compressed expressions in (4.4.7) are only appropriately used where construction of context-given sequences has already begun.

The context-given sequence definitizes an ordinal expression and makes its interpretation possible. Therefore, the semantic interpretation rule for NPs with ordinals can be formulated specifically for a definite NP and needs to mention the context-given sequence. A rule which will do the job is as in (4.4.9)

4.4.9 If *X* is followed by *Y* in the counting sequence, the denotation of an NP of the form



is the object in the context-given sequence and in the denotation of the lower Nominal, first arrived at by moving along the sequence from the object denoted by an NP of the form



This rule, modelled on rule (4.3.3) interpreting cardinals, similarly appeals to certain concrete operations, such as moving along a sequence. I assume that a child learning her first language grasps the concrete concept of moving along a sequence, just as she grasps the idea of adding an object to a collection. Less psychologically salient sequences, such as sequences of distant events, are, like less psychologically salient collections, harder to grasp and operate with. The structures in rule (4.4.9), like all such structures in rules in this book, are simply for concrete illustration and mirror the structures found in a particular language (here English). The structures used in other languages for expressing ordinality, perhaps with no determiner, and with alternative word-order, can be substituted. As with the rule for cardinals, I assume that the meanings of the first few ordinals are learnt independently of the general rule, which builds on this primitive knowledge of the first few ordinals. The first few ordinals are often suppletive in languages, for example Spanish *primero*, *segundo*, not morphologically related to *uno*, *dos*. In many languages, the word for *second* is also the word for *other*, for example Arabic *taani*.

I suggest that the meanings of ordinals, after about 3, are first learned in this wholistic way, with a rule such as (4.4.9) being acquired for interpreting a relatively complex structure, and based on the rote-learnt counting sequence. So, as in the case of cardinals, a child may in some sense know the meaning of, say, *the fifth man* without actually knowing a meaning for the isolated word *fifth*. Indeed even for adults, if ordinals generally only occurred within structures as in (4.4.9), it would not be possible to distinguish between knowing the meaning of an ordinal numeral word and knowing the meanings of structures in which it occurs. For other configurations in which ordinals occur, they could conceivably all be interpreted as derivative, elliptical, or compressed versions of a basic structure as in (4.4.9).

There are no bare ordinals used in arithmetical statements the way cardinals are. Thus while *five and two is seven* is well-formed in English, **fifth and second is seventh* is not. We do not use ordinals to do our sums.

It is possible, as a logical exercise without any necessary psychological validity, to concoct denotations for the individual constituents of a structure as in (4.4.9) in set-theoretic terms, and to define compositional rules that produce an appropriate result. Thus, for example, the denotation of *fifth* could be the set of all ordered pairs such that each pair consists of (1) a sequence of at least five objects and (2) the object which is the fifth in that sequence, for example ⟨⟨A, B, C, D, E, F, G⟩, E⟩. The rule combining this ordinal denotation with a (singular) nominal denotation would narrow this set of pairs down to a subset such that each pair now contained as its second member an object in the denotation of the nominal. For example if the nominal were *Greek letter*, the pair given above would be eliminated, but if the nominal were simply *letter*, that pair would survive to be included in the denotation of the larger nominal, for example *fifth letter*. Finally, the semantic effect of the definite determiner, or its equivalent, would be to give as the denotation of the whole expression that object which is the second member of the pair whose first member is the context-given sequence. Thus, if the context-given sequence were the Roman alphabet in its conventional order, the denotation of *the fifth letter* would be the letter 'E'. As a declarative reconstruction of the relations between the denotations of the parts and the whole, this may be correct. But as a psychological story of how a speaker interprets an expression containing an ordinal, it is mightily implausible.

In the view taken here, the meanings of ordinals are not derived from the meanings of the corresponding cardinals. There is no simple compositional rule accounting for the meaning of *seventh* as a function of the meanings of the parts *seven* and the suffix *-th*. Instead, I claim that both cardinal and ordinal meanings are derived from the conventional recited counting sequence. The relationship between cardinal and ordinal is precise, but indirect. The universal linguistic markedness of ordinals, relative to cardinals, can be taken to arise from the clearly greater complexity of the interpretation of ordinals, in particular the necessary involvement of a context-given sequence.

4.5 Knowledge of Number

I am sceptical about the possibility of rational agreement on what constitutes the 'full concept' of number. The question could drag one into fruitless essentialist wrangles. Such considerations do not arise in this work, which is not concerned, as Frege and Russell were, with *defining* numbers and number in general, but rather providing an account of how people can use, and acquire the use of, numeral expressions. I am only concerned with the essence of numbers to the extent that ordinary linguistic usage appears to reveal a shared belief by speakers that certain properties and relations are essential to number(s).

Of the abstract notions of particular numbers, we can say that such notions come in two versions, cardinal and ordinal, both of which are abstractions from the possible conclusions of acts of counting. Cardinality 5, for example, is just the property ascribable to any aggregate or collection of objects to which one can apply the counting activity in the orthodox, conventional way, concluding (in English) with the word *five*. Russell rejected an essential connection between counting and number: 'In counting, it is necessary to take the objects counted in a certain order, as first, second, third, etc., but order is not of the essence of number: it is an irrelevant addition, an unnecessary complication from the logical point of view' (1919, p. 17). Russell's logical point of view is a pinnacle reached by evolution of a language and a culture (not to mention the species) and the learning of individuals. Sitting on this pinnacle, Russell, like Frege, has kicked the ladder away, and described how the top of the pinnacle can be charted. I wish to say that the pinnacle is nothing but a high rung that we have reached on the ladder, and if we detach ourselves from the ladder we have nothing to stand on. Ordinary understanding of number is inextricably bound up with counting, and is by no means 'an irrelevant addition'.

It follows from the view that numbers are abstractions from conclusions reached by counting that for any given numeral word, the cardinality associated with that word will only be ascribable to collections which can be put in a one-to-one correspondence with each other. Thus numeral words come to have what we call, *ex post facto*, 'exact' numerical values.

The notion of one-to-one correspondence pre-dates linguistic methods of counting, being used, and therefore presumably in some sense understood, for example, in reckoning on the fingers, or in keeping tallies of notches on a stick. This fact, as Wright points out in response to Benacerraf (1965), shows that '*sameness* of number – 1–1 correlation – is the fundamental thing; that is, its possession is necessary for possession of *any* understanding of cardinal number, and sufficient for the rudimentary apparatus of number judgements illustrated in the case of the tribe [a hypothetical tribe who can communicate cardinalities by means of tallies] (1983, pp. 120–1). But it is doubtful whether one can reasonably attribute understanding of particular cardinalities (above about 3) to anyone who has not mastered a counting sequence. A person who commands a counting sequence can be said to possess a knowledge of how to use, say, *seven, eight*, and *nine* permanently and simultaneously, even while he is not actually using them. This knowledge is part of his linguistic and arithmetical competence. (In this area, knowing the basic arithmetical facts is *the same thing as* knowing the relevant linguistic facts.) But the pre-counting tribesman who uses tallies, while he may *perhaps* be said to have the notion of some particular number in mind on a particular occasion when matching a set of tallies with a collection of objects, 'forgets' that number (assuming he ever 'knew' it) as soon as his matching operation is finished and he has turned his attention to something else. But if he learns a counting sequence, he puts an instant tally-making kit inside his head, and then and only then can it be said that the set of possible tallyings is in some way known to him permanently and simultaneously. Kleene writes that 'the idea of such [1–1] correspondence is *more primitive than* the idea of "cardinal number"' (1967, p. 175, emphasis added), and also gives an example of a hypothetical tribe who compare large collections by one-to-one pairing of the members.

In the next chapter, where I discuss the development of syntactically complex numerals interpreted by addition and multiplication, I will give an account of sentences with bare numeral NPs expressing arithmetical truths, such as *Six threes are eighteen*. The interpretation of bare numeral NPs as denoting sets of collections allows a straightforward account of such sentences. Meanwhile, we have enough machinery to see how a sentence such as *Seven is a number* comes out TRUE if the denotation of the word *number* (in its sense of 'positive integer') is taken to be the set of all sets of collections. This set contains as a member the denotation of any particular numeral word. So whereas a particular numeral (for example *seven*) is a second order predicate, denoting a set of collections, the predicate *number* itself is third order, denoting a set of sets of collections.

Thus while, I claim, the word *seven* is no more abstract in its denotation than the word *bricks* (the sets intersect), the word *number* is indeed more abstract than the word *seven*. It is not clear that many ordinary speakers of English (and no doubt other languages) possess a clear conception of this abstract denotation of the word *number*. Ordinary speakers will happily point to a symbol on a page and declare it to be a number, and will find difficulty in expressing any more abstract notion of what a number is. For such speakers *Seven is a number* expresses a received truth, having a somewhat distant relationship to its interpretation by mathematicians and philosophers.

The abstract usage of mathematicians and philosophers is a different language game from the everyday use of words like *seven* and *number*. The two language games do, however, bear a close family resemblance to each other, and, something which is potentially confusing, they use substantially the same syntax. Hodes distinguishes 'a strict notion of logical form ... from a looser localized notion applicable to particular arguments or particular kinds of discourse (e.g. that of the number theorist [or the elementary arithmetic teacher]) in isolation from the rest of the language' (1984, p. 142). The looser kind of language permits, for instance, *seven plus five equals twelve*.

Hodes gives some revealing quotations from Frege's diary for 1924, near the end of his life. In a diary dated 23 March, 1924, Frege wrote:

> My efforts to become clear about what is meant by number have resulted in failure. We are only too easily misled by language and in this particular case the way in which we are misled is little short of disastrous. The sentences 'Six is an even number', 'Four is a square number', 'Five is a prime number' appear analogous to the sentences 'Sirius is a fixed star', 'Europe is a continent' – sentences whose function is to represent an object as falling under a concept. Thus the words 'six', 'four', and 'five' look like proper names of objects, and 'even number', 'square number', and 'prime number' along with 'number' itself, look like concept-words; so the problem appears to be to work out more clearly the nature of the concepts designated by the word 'number' and to exhibit the objects that, as it seems, are designated by number-words and numerals. (Hermes et al., 1979, p. 263)

Then in an entry dated March 24, he continues:

> Indeed, when one has been been occupied with these questions for a long time, one comes to suspect that our way of using language is misleading, that number-words are not proper names of objects at all and words like 'number', 'square number' and the rest are not concept-words: and that consequently a sentence like 'Four is a square number' simply does not express that an object is subsumed under a concept, and so just cannot be construed like the sentence 'Sirius is a fixed star'. But how then is it to be construed? (1979, pp. 263–4)

(Note that astronomical facts are still the standard to which Frege compares statements about numbers.)

The ordinary mature speaker of a language certainly knows no more than Frege. He knows in some sense that the numeral expressions which he uses in counting correspond to numbers. What kind of entities numbers are might be quite mysterious to him. The speaker who knows a numeral system may fairly be said to know the meanings of the numerals, but in this case, at least, 'knowing the meaning of' must be interpreted in a quite limited way. In particular, except for the very lowest-valued numerals, the speaker does not have readily available a mental model (in the sense of Johnson-Laird, 1983) of a number. Whereas it might be plausible to say I have a mental model of a table, which comes to mind with the word *table*, what comes to mind with the expression *twenty-three*, if anything comes at all, is not a model of the number itself, but some kind of paraphrase, synonym, or translational equivalent, such as perhaps the Arabic representation '23'. '23' is no more the number itself than *twenty-three*. It appears that speakers know the meanings of numerals in a proof-theoretic way, that is by being able to make (arithmetical) inferences with sentences involving numeral expressions without necessarily going to any extralinguistic model. The view taken here is the one J. S. Mill (1906) eloquently holds up as plausible, though he then rejects it:

> What has led many to believe that reasoning is a mere verbal process is, that no other theory seemed reconcilable with the nature of the Science of Numbers. For we do not carry any ideas along with us when we use the symbols of arithmetic or of algebra. In a geometrical demonstration we have a mental diagram, if not one on paper; AB, AC, are present to our imagination as lines, intersecting other lines, forming an angle with one another, and the like, but not so *a* and *b* [symbols of algebra]. These may represent lines or any other magnitudes, but those magnitudes are never thought of; nothing is realised in our imagination but *a* and *b*. The ideas which, on the particular occasion, they happen

> to represent, are banished from the mind during every intermediate part of the process, between the beginning, when the premises are translated from things into signs, and the end, when the conclusion is translated back from signs into things. Nothing, then, being in the reasoner's mind but the symbols, what can seem more inadmissible than to contend that the reasoning process has to do with anything more? (906, p. 167)

Although Mill's discussion here focuses on algebraic symbols as variables over numbers, it is clear that he is describing the same view concerning symbols for particular numbers used in calculations.

In fact, non-linguistic, non-mental, physical models or interpretations for numerals can be constructed by assembling collections of objects or making series of marks; although any competent speaker can do this, it is a laborious task and involves certain skills in addition to just a knowledge of a numeral system. Speakers of languages with well-developed numeral systems know that numerals denote numbers, and they will readily agree that numbers (whatever they are) form a continuous sequence whose starting point is 1, with all successive numbers reachable by iterative applications of a successor function. The successor function may be conceived of by the speaker as in some way analogous to the physical operation of adding another object to a collection. But, except in the case of quite low numbers, humans do not have the apparatus to construct in their heads (mental models of) numbers, that is of the non-linguistic entities denoted by numerals. This applies to the vast majority of numbers and numerals, as known by an adult, but I do not discount the possibility that mental models for the very low numbers play an important part in the acquisition of numerals and number concepts by children.

In the next chapter, dealing with syntactically complex numerals, whose interpretation involves addition and multiplication, I shall insist on the evolutionary development of numeral syntax in strict parallel with a clear denotational semantics. But this is not to insist that the meanings of complex numeral expressions are necessarily acquired and known, psychologically, by the ordinary non-inventive speakers of each generation as direct representations of possible collections of objects from the world. It is well known that one cannot bring to mind collections

with large cardinalities. I cannot distinguish an image of 100 horses in my mind from an image of 101 horses. The semantic information that the knower of a numeral system actually retains in his mind is usually a quite uneconomical set of rules stating various equalities between numeral expressions, as in the rote-learnt formulae of multiplication tables. These formulae are theorems whose truth and consistency derive from basic truths about collections. Language acquirers are simply given these formulae as part of their cultural patrimony; they are inventions, like the wheel, passed down culturally, and which there is no need to reinvent. The underlying presence of a solid denotational semantics at every stage in the evolution of complex numeral expressions is the guarantee that the developed system works and is practically useful. Occasionally, a new construction involving numerals is invented. At every such new stage in the evolution of syntactic constructions and their accompanying semantic interpretation rules, there must be a clear method for determining what particular sets of collections of objects the new expressions denote. It is just because there is a sound denotational interpretation of each construction that the inheritors of the rote-learnt formulae get no nasty surprises when relating their sums to the collections of real-world objects they are interested in.

Numbers seem to be real, since talk involving them engages with the world in convincing ways, but psychologically this is puzzling, as we cannot easily see what sort of things numbers could be. The introduction of an evolutionary dimension relieves the ordinary individual user of numerals of the responsibility for knowing what numbers are. He can say of arithmetic, as he says of electrical appliances and computer programs, 'I know it works – I don't know how or why, but it does work.' Clearly, someone in a community needs to keep a more rationally based grip on how and why arithmetic works, which involves technicalities and abstractions. Society needs some philosophers and mathematicians. Without them to perform a watchdog role in the transmission of knowledge involving numbers, systems could degenerate to the kind of situation where there are 'correct misinterpretations' as mentioned in Chapter 1, Section 2, for example where a language seems to treat 20 as 9×2, or 6 as $5 + 3$.

> The knowledge of an individual is grounded in the knowledge of the community authorities. ... Behind them is a

> sequence of earlier authorities. ... we must understand how the chain of knowers is initiated. Here I appeal to ordinary perception. (Kitcher, 1984, p. 5)

In the view just outlined, the speaker who knows a numeral system possesses what might be called a 'concept of number', since, through the system of expressions, he shows that he knows something about numbers. But it must be emphasized that this concept of number is obviously weaker, and less specific than the concept studied by Piaget (1952). For Piaget, a full concept of number involves a full awareness that the cardinality of a set of objects cannot be changed by operations other than inserting or removing objects. The following quotations bring out the fact that, for Piaget, someone may be able to count, and have some kind of verbal competence with numeral expressions, but not possess the concept of number.

> It is as though, for the child [at this stage], quantity depended less on number (a notion which, if our hypothesis is correct, is still only verbal, although the child can count correctly) or on the one-one correspondence between the objects, than on the global appearance of the set, and in particular of the space occupied by it. Even Mül, for instance, who could count, thought that 'there are more where it's bigger', although he had counted that there were six glasses in the group and six bottles in the row.
> ... it is possible that at this level the correspondence between numerals and objects is still purely verbal, and that the child has not yet acquired the notions necessary for the construction of number itself, i.e. permanence and equivalence of sets irrespective of the distribution of the elements of which they are composed. (1952, 45–6)

It is ultimately a matter of the theorist's definitional fiat how much he decides to include in the criteria for 'possessing a concept of number' (or any other concept). Certainly, young children appear to know less about number than adults, in that they apparently believe that the number of objects in a set decreases when the objects are moved closer together. But full adult awareness of the relation between numbers and sets is not, for me, a condition for some knowledge of a numeral system. It seems reasonable to envisage partial competences.

Goody (1977) observes that a grasp of the abstract senses of numerals is in some sense facilitated by a graphic method of representation. On the basis of his observations of the LoDagaa people of Northern Ghana he writes:

> The idea of multiplication was not entirely lacking; they did think of four piles of five cowries as equalling twenty. But they had no ready-made table in their minds by which they could calculate more complex sums. The reason was simple, for the 'table' is essentially a written aid to 'oral' arithmetic. The contrast was even more true for subtraction and division; the former can be worked by oral means (though literates would certainly take to pencil and paper for the more complex sums), the latter is basically a literate technique. The difference is not so much one of thought or mind as of the mechanics of communicative acts, not only those between human beings but those in which an individual is involved when he is 'talking to himself', computing with numbers, thinking with words. (1977, p. 12)

Goody is careful not to carry this view to an extreme dichotomy.

> Literacy and the accompanying process of classroom education bring a shift towards greater 'abstractedness', towards the decontextualization of knowledge, but to crystallize such a developmental process into an absolute dichotomy does not do justice to the facts either of 'traditional' society, or of the changing world in which the LoDagaa now find themselves. (1977, p. 13)

The full abstraction of the notion of number from a conventional counting sequence includes the insight that any particular conventional sequence is arbitrary. Thus the mathematician or philosopher does not make the number 5 dependent in any essential way on the contingent fact that *five* is the fifth word in the English counting sequence. Whatever counting sequence one begins with is taken as merely an exemplar, and one realizes that other sequences in other languages would serve the same purpose. The ability to envisage an abstract sequence corresponding in its formal properties to *any* conventional counting sequence is the ability to conceive of the relatively abstract notion of number. But for an ordinary speaker, his native language's conventional counting sequence, learned by rote, together with other handy rules of thumb, such as multiplication tables, is quite sufficient to get by in everyday life, and any more abstract conception buys no practical advantages.

The account proposed here might have satisfied Mill in its attention to the connection between numbers and collections of objects. One sympathizes with the spirit, if not the letter, of the following: 'All numbers must be numbers of something; there are no such things as numbers in the abstract. *Ten* must mean ten bodies, or ten sounds, or ten beatings of the pulse. But though numbers must be numbers of something, they may be numbers of anything' (1906, p. 167). Where Mill seems to have gone wrong is in neglecting two vital functions played by language. He neglected the role of the counting sequence in *creating* numbers in the abstract; in this respect, his predecessor Locke, who stated clearly the importance of numeral expressions and the counting sequence, had the better of him. And Mill also neglected the importance of language in fixing concepts beyond the reach of perceptual experience to which numbers may apply; in this respect, Frege's subsequent criticisms of his naive empiricism are well taken.

Russell, while stating that Frege 'correctly answered' the question 'What is a number' (1919, p. 11) nevertheless gives a definition of number which moves significantly away from (early) Frege in the direction of Mill's position: '*The number of a class is the class of all those classes that are similar to it*. Thus the number of a couple will be the class of all couples. In fact the class of all couples will *be* the number 2, according to our definition' (1919, p. 18). Perhaps in acknowledgement of Russell's debt to Frege, his view is sometimes characterized as 'the Frege/Russell view'. But it is important to note the clear difference between Frege's position, that 'every individual number is a self-subsistent object' and Russell's definition of numbers as (extensionally) sets of sets or (intensionally) properties of sets.

Kitcher (1984) has recently resurrected a position of 'mathematical empiricism', which adopts a number of Mill's ideas about mathematical structure reflecting the structure of the real world. At the same time, Kitcher espouses a psychologistic epistemology for mathematics, and comes to terms with a position that could be called 'enlightened conceptualism'. An empiricist ingredient in Kitcher's position emphasizes, as I have done, the role played by physical operations, such as adding objects to collections, in

the acquisition of number concepts. 'Children come to learn the meanings of "set", "number", "addition" and to accept basic truths of arithmetic by engaging in *activities* of collecting and segregating. Rather than interpreting these activities as an avenue to knowledge of abstract objects, we can think of the rudimentary arithmetical truths as true in virtue of the operations themselves.' (Kitcher, 1984, pp. 107–8). Without going into any detailed examples, or mentioning the concept of number, Kitcher gives the following picture of the role of language in building mathematical knowledge:

> When we learn our language a complex set of dispositions is set up in us. In virtue of the presence of these dispositions, which comprise our linguistic ability, we become able to entertain certain beliefs. Let us now suggest that exercise of our linguistic ability generates in us particular beliefs and that it warrants those beliefs. ... linguistic training induces psychological changes, and these changes make available processes which can generate warranted belief. (1984, pp. 70–1).

I take Kitcher's use of 'disposition' to be compatible with the attribution of mental representations. That is, the appeal is not to dispositions 'without structured vehicles'. Chomsky's (1980a, Chapter 2) discussion of Kitcher (1978) suggests that Kitcher's type of dispositionalism is one lacking structured vehicles, but Kitcher (1984, p. 70n) withdraws somewhat from his 1978 position criticized by Chomsky. Kitcher prefers to deal with dynamic notions such as 'process', 'activity', and 'operation', rather than with the notions of their static results and bases such as (mental) 'state', 'object', and 'representation'. 'One central ideal of my proposal is to replace the notions of abstract mathematical objects, notions like that of a collect*ion*, with the notion of a kind of mathematical activity, collect*ing*' (1984, p. 110). But I take it that any account of permanent human knowledge needs to admit states and representations, which are (mental) objects. The activity of collecting inevitably yields collections, (unless it fails because it is interrupted). If I construe Kitcher's (1984) remarks correctly, the account which I have given of the acquisition of a basic numeral lexicon and numeral meanings can be seen as a somewhat detailed working out of his empiricist/conceptualist programme at the foundations of arithmetic, the number sequence itself. The crucial role of language and conventional counting sequences are to be noted.

Whilst accepting the empirical basis of an understanding of number in the perception and manipulation of collections of physical objects, one need not reject knowledge of number as knowledge of abstract objects.

> Nor is it even necessary to forego the claim that mathematics studies abstract objects – *so long as we regard that claim as ultimately interpreted in terms of ideal operations* [on physical objects]. What is central to my account is a scheme for recasting mathematical language, so that it can dissolve the mysteries which Platonism spawns, and this, I suggest, is consistent with viewing Platonism as a convenient *façon de parler*, a position which errs by adopting a picture of mathematical reality without recognizing the route through which the picture emerged. (Kitcher, 1984, p. 42)

Having articulated how expressions containing numerals are interpreted in terms of the world, it seems of secondary importance to ask whether 'numbers exist'. Some philosophers utter provocatively paradoxical sounding formulations on this topic. Thus Benacerraf concludes his article with: 'if the truth be known, there are no such things as numbers; which is not to say that there are not at least two prime numbers between 15 and 20' (1965. p. 73). And I believe that Field, in his *Science without Numbers*, is not clearly being faithful to the notion of truth as usually understood, when he writes: 'To explain even very complex applications of mathematics to the physical world it is not necessary to assume that the mathematics that is applied is true' (1980, p. vii).

More illuminating than such paradoxical formulations are proposals that we conceive of the existence of abstract mathematical objects in what Putnam (1975a) calls a 'modal' way.

> To talk about the 'existence' of numbers would be simply to talk about the logical possibility of the corresponding formal properties. This seems a reasonably plausible view of mathematical existence. (Armstrong, 1978, p. 73)

> There is not, from a mathematical point of view, any significant difference between the assertion that *there exists a set of integers* satisfying an arithmetical condition and the assertion that *it is possible to select* integers so as to satisfy the condition. (Putnam, 1975a, p. 49)

> Mathematics is higher-order modal logic. (Hodes, 1984, p. 149)

As Hodes points out, the adoption of a modal view of mathematics solves a technical difficulty with the Russellian view of numerals as denoting sets of sets (or sets of collections in my variant). If the universe happened to be finite, propositions involving numbers higher than the number of entities in the universe could not be guaranteed their appropriate truth values. To replace 'existence of' by 'possibility of constructing' sets of collections of objects solves this difficulty.

Although to a large extent the issue seems to be one of devising an appropriate *façon de parler* for talk about numbers, it seems in most cases unobjectionable to treat numbers (as opposed to collections) as real, but abstract, objects created through an interaction of people, language, and the world. 'The natural numbers are the work of men, the product of human language and of human thought' (Popper, 1972, p. 160, see also 1973, p. 22) This is not to say straightforwardly that numbers didn't exist before Man got the ability to deal with them. Starting with the simple and vague idea that numbers, whatever they are, are the meanings of numerals, an account of the meanings of complex numerals involving addition and multiplication, the beginnings of mathematics, can be provided by taking the denotations of numerals to be sets of collections of objects (see next chapter). So numbers can be thought of as sets of collections, as Russell thought of them.

Presumably objects have always in some sense existed, from long before Man got his numerical ability. But even the notion of an object has a certain modal element of potentiality and relativity to particular organisms in it. Tables and chairs and other familiar middle-sized things show up saliently on our sensory radar, but no doubt there are organisms totally unable to conceive of these as objects. Low-level organisms lack any concept of physical object at all. If humans had evolved differently, our concept of physical object might have been different. Right now we have to say that the objects in the universe are those things that correspond to our concept 'object'. So if our concept had developed differently, the objects in the universe would be a different set from those that (we say) are objects now.

The relativity to particular organisms applies also to collections of objects. The notion of a collection in some sense presupposes a collector and an act (not necessarily physical) of collecting. It is a matter of including some objects in the collection concerned and excluding others. Sometimes the criteria for inclusion or exclusion are thrust forcefully upon us by our senses and cultural classifications, as when one sees ten students round a tutorial table. We also have the ability to conceive of more abstract or whimsical collections. But here too we are constrained by our biology. The collections of things in the world are those which we are able to conceive of or construct, given our peculiar discriminatory powers of inclusion and exclusion.

We *could* say that the collections of objects which are in the denotations of numerals have been lying (or perhaps hurtling) around in the universe since the beginning of time, waiting for Man to get complicated enough to conceive of them. In some sense this must be true (apart from the metaphorical 'waiting'). But to an outside observer who had not foreseen the course of evolution it would not have been clear just what in the universe Man would end up being able to regard as a collection and what not.

5

Syntactic Integration of Counting Words

Building on the view of numerals as collection-denoting expressions, the most obvious cross-linguistic generalizations about the syntactic distribution of numerals are argued to emerge naturally from the basic meanings of numerals. Basic universal patterns in the embedding of numerals in larger constructions and the formation of complex numerals are explained in terms of natural parallels and analogies with non-numeral constructions. The naturalness of these parallels and analogies facilitates the invention and transmission to ordinary language acquirers of these numeral patterns.

5.1 Counting Words Become Adjectives or Nouns

> Numerals are generally treated as adjectives ... but not infrequently the higher ones or some of them are substantives. (Jespersen, 1969, p. 119)

5.1.1 CORBETT'S UNIVERSALS

> (1) simple cardinal numerals fall between adjectives and nouns
> (2) if they vary in behaviour it is the higher which will be more noun-like.
>
> (Corbett, 1978a, p. 368)

Corbett (1978a, 1978b) has systematized the impressionistic generalization expressed by Jespersen, and backed it up with carefully assembled data from a range of languages (though with a marked emphasis on Slavic languages). His method is to take the numeral words from a language and match them against recognized morphosyntactic characteristics of adjectives and nouns in that language. The results can be expressed in a 'squish' diagram, of which the following is an example.

Table 5.1 Syntactic behaviour of Russian cardinal numerals.

	odin 1	dva 2	tri 3	pjat' 5	sto 100	tysjača 1,000	million 1,000,000
1 Agrees with N in syntactic number	+	–	–	–	–	–	–
2 Agrees in case throughout	+	–	–	–	–	–	–
3 Agrees in gender	+	(+)	–	–	–	–	–
4 Marks animacy	+	+	+	–	–	–	–
5 Has own plural	–	–	–	–	(+)	+	+
6 Takes agreeing determiner	–	–	–	–	+	+	+
7 Takes N in genitive plural throughout	–	–	–	–	–	±	+

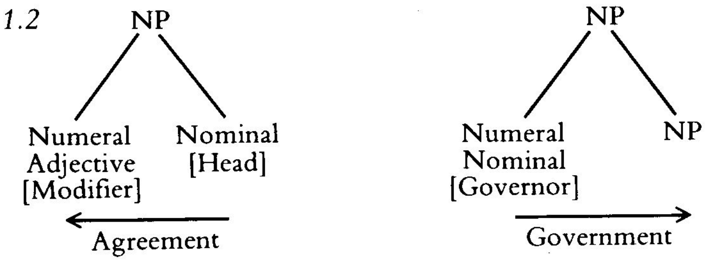
Tests 1–4 are features we normally associate with the syntax of adjectives. These first four tests do not split the numerals into two clear groups (adjectives and non-adjectives); rather they show that the numerals are more or less adjectival, the more adjectival being those on the left. Similarly, tests 5-7, features of the syntactic behaviour of nouns, show that some numerals are more noun-like than others. (Corbett, 1978a, p. 358).

(The parenthesized '+' signs and the '±' sign in the table indicate that the form in question has the feature in some environments or uses but not in all.)

Corbett's methods and conclusions are interesting, and certainly bring us closer to the truth about the syntactic categorization of numerals, but I believe his generalizations need some reorganizing and overhauling. The main revision I would propose is a strict separation between morphological and syntactic criteria.

The two basic types of syntactic structure into which a numeral sister of a nominal may fit are given schematically in (5.1.2). [I use 'nominal' as a cover term for the head of a noun phrase (NP), which may be a bare noun or sometimes contain other elements, such as determiners and other, non-numerical adjectives.]

5.1.2



Latin examples would be *duae feminae* (note case and gender agreement) versus *sex milia feminarum* (note obligatory genitive plural in the governed noun). The linear order of the elements is not represented in (5.1.2). Adjectives may precede or follow a head nominal, as governed NPs may follow their governors. The crucial structural relationships are 'head/modifier' and 'government'. The head of a construction usually determines the case and perhaps the number and gender of its modifiers. The governor in a construction determines the case, but not other features such as number and gender, of its sister constituent. The syntactic case of the nominal head or governor in a construction is determined from outside the construction itself, by the structure of the sentence into which it fits. I simply assume these typical structures in languages and make no attempt here to explain why such relations as head/modifier and government should exist in language. But, taking their existence as an established fact, along with other known typical characteristics of (non-numeral) adjectives and nouns, I will investigate, in this section, *how* numerals align with adjectives and nouns in their syntactic behaviour, and, in the next two sections, *why*.

To take examples from the Russian squish above, in agreeing in gender with the following noun, a numeral behaves adjectivally and reflects this head/modifier relationship, whereas in the more nouny construction the numeral imposes genitive plural case and number on the following nominal, while taking its own case from outside the construction. The government relationship may be realized by other means than the imposition of morphological features such as case. In English, it can be expressed by the insertion of the preposition *of*, as in *millions of cats*.

It is possible for two constituents in a construction to influence each other 'in opposite directions' via the agreement and government relationships. For example, in a case-marking language which also has auxiliary verbs agreeing with their subjects, the auxiliary can be said to govern the subject NP, and to impose nominative case on it, while the subject NP imposes its features for number, gender and person on the auxiliary via the agreement mechanism. As far as constructions with numerals are concerned, however, the adjectival and nouny structures tend to be relative complements of each other. Either the government relation goes one way or the agreement relation goes the other. Instances can be found where a numeral both governs (or seems to) and agrees with its sister constituent. G. Corbett has given me the example of Russian *dve sosny*, 'two pines'. Here *sosny* is genitive singular because of the numeral 2, and *dve* is feminine singular nominative, feminine because of *sosny*. Another example might be Welsh *tair o ferched* '3[feminine] of daughters', where the preposition *o* can be interpreted as a genitive marker. But where case is concerned, an element cannot both govern and agree with its sister, since government involves imposing case.

For this reason, there must be an omission in the Russian numeral squish given above, in the cell where *tysjača* and 'Agrees in case throughout' intersect. In this cell, there should be a '+' sign, since this cell has to complement the cell at the bottom of the column, corresponding to the criterion 'Takes N in genitive plural throughout'. According to Corbett's commentary, *tysjača* may either impose the genitive plural on the following noun or agree with it in some instances in the oblique cases. Thus, for *tysjača*, there should be a '+' sign in both the 'Agrees in case' cell and the 'Takes genitive plural N' cell. But this would upset the nice continuity of the squish, which is the property Corbett most wished to emphasize. The continuity can be saved by replacing the second criterion 'Agrees in case throughout' by 'Agrees in case in direct (non-oblique) cases', but I would be uneasy about the general validity of squishes if too much epicyclic ingenuity had to be used in devising the tests to be applied. And even if one can make the syntactic data conform, the cross-language generalizations which appear to hold for syntax do not hold for morphology.

Latin and Ancient Greek present difficulties for a method of analysis which does not separate morphological from syntactic criteria. I will discuss these briefly before proposing alternative generalizations of the data. The Latin forms up to 99 behave well in Corbett's terms. But then there is a discontinuity. *Centum*, 100, is an indeclinable adjective, thus showing no overt signs of agreement with its governing noun; but the forms *ducenti*, *trecenti*, ..., *octingenti*, *nongenti*, 200, 300, ..., 800, 900, are adjectives inflecting fully for gender and case, in agreement with the

governing noun. True, Corbett's universals (5.1.1) are carefully worded to apply only to *simple* numerals, and *ducenti*, and so on are clearly bimorphemic. But there is adjectival behaviour going on here and it seems a pity not to frame an account which can say something about it. The Ancient Greek case does actually involve only simple numerals.

In Ancient Greek the form for 1 inflects for all four cases and all three genders; the form for 2 inflects for case only, not gender (and collapses nominative with accusative and genitive with dative); the forms for 3 and 4 return partially to inflection for gender, distinguishing masculine from feminine in the nominative and accusative cases; the form for 4 makes all the distinctions made by the form for 3, plus one more, as it distinguishes between nominative and accusative in the masculine. (Data from Rutherford, 1930; Goodwin, 1879.) In trying to form a squish diagram of these facts, we should have to do as follows:

5.1.3 Inflectional morphology of low numerals in Ancient Greek

	1	2	3	4
Agrees in gender	+	–	(+)	(+)
Agrees in case	+	(+)	(+)	(+)

If showing overt agreement in gender is to be taken as a feature of adjectival behaviour, there is no getting away from the fact that the form for 2 appears less adjectival than the forms for 3 and 4. In general, Ancient Greek adjectives in the dual inflect for gender as well as case, although some do not.

The solution to these problems (which unfortunately substantially undermines the basis of Corbett's elegant squishes) is to recognize that if an adjective has a defective inflectional paradigm, or even if it is downright indeclinable, it is no less an adjective syntactically. Latin *centum* is indeclinable, but it does not impose syntactic case on a following noun, so it is called an adjective. In traditional Latin grammar, the syntactic head/government relation is paramount, even to the extent of classifying the singular form *mille* as an adjective (indeclinable) because it does not impose genitive plural on a following noun, and the case-inflecting plural forms *milia* and so on as nouns, because they take the case imposed on them by the structure of the sentences they appear in, and in their turn impose genitive plural on the following nouns, as illustrated in (5.1.4).

5.1.4

Consul venit cum mille viris
Consul came with 1000 men
(case imposed by cum)

Consul venit cum sex milibus	virorum
Consul came with six thousand	men
(case imposed by cum)	(genitive plural imposed by milibus)

Corbett is aware of such facts. Of them, he writes: 'they do not constitute clear counterevidence to our claim. The invariable forms do not behave as nouns. They are indeclinable adjectives'. (1978a, p. 364). Overt agreement is positive evidence of adjectival behaviour. Lack of overt agreement, as with *centum* and *mille*, is not evidence against adjectival behaviour, but simply lack of evidence one way or the other. Corbett's squishes represent the clear presence of adjectival behaviour in some particular respect by a '+' sign in the appropriate cell. Correspondingly, a '-' sign should mean (and does mean in some cases) the clear absence of adjectival behaviour. But sometimes the '-' sign is used to indicate, not the absence of adjectival behaviour, but lack of overt evidence for the presence of such behaviour.

It boils down to this. If the fact that Russian *tri* shows no overt gender distinction earns it a '-' sign in its cell for 'agrees in gender', then Latin *centum* and *mille* should get '-'signs in their cells for that feature too. But since *ducenti*, ..., *nongenti* would clearly get '+' signs for 'agrees in gender', the cline from adjectival to nouny behaviour which Corbett wants to display in both morphological and syntactic features is interrupted.

Adopting the separation of morphology and syntax, Corbett's universals can be restated as follows:

5.1.5

(1) SYNTAX

(a) Cardinal numeral words sometimes agree with a noun in the same construction (that is behave adjectivally), and sometimes govern it (that is behave in a noun-like way).

(b) If they vary in behaviour it is the higher which will be noun-like, governing rather than agreeing with a noun in the same construction.

(2) MORPHOLOGY

Lower-valued numeral words tend to have more complete inflectional paradigms than higher-valued ones. This applies especially strongly to simple (monomorphemic) words.

Only the syntactic statements (1a,b) here pertain to a difference between noun and adjective. In particular languages, the difference between governor (noun) and modifier (adjective) will correlate with other syntactic features, especially word order. The morphological generalization (2) has no implications for nominal or adjectival categorization. Declinability, *per se*, is not a criterion discriminating between adjective and noun.

The Russian and Latin data conform to the syntactic statements in (5.1.5), with Russian *tysjača* and Latin *milia* being the first words to impose genitive plural on a following noun. In these cases, noun-like behaviour first emerges at a quite high numerical value (1000). Behaviour which is somewhat noun-like can emerge much lower, as in Serbo-Croat, where numeral words with value 5 and above impose genitive plural on a following noun (Corbett, 1978b, p. 48), but this is unusual. And Serbo-Croat *pet* '5', which takes genitive plural, is avoided in oblique cases without a preposition; since it does not decline, the case would not be marked; so it is not fully noun-like. By far the great majority of numeral words behaving syntactically as nouns have values which are used as multiplicative bases in complex numeral expressions, for example (in a typical decimal system) 10, 100, 1000. This fact will be significant in explaining the correlation of higher values with nouniness.

The morphological statements in (5.1.5) apply to all numeral *words*, both simple and complex, but express no more than a tendency, as illustrated in (5.1.6). Note that some of the complex forms here clearly go against the tendency for lower-valued words to have fuller morphological paradigms, while the simple forms follow the tendency.

5.1.6

	Simple	Complex
Inflecting	unus, duo, tres 1 2 3	ducenti, trecenti, nongenti 200 300 900
Indeclinable	quattuor, novem, 4 9 decim, centum, mille 10 100 1000	undecim, septemdecim, 11 17 triginta, nonaginta 30 90

The universals given do not apply straightforwardly to syntactically complex numerals (numeral phrases) formed from several words. Thus, both the following express *2500 horsemen* in Latin.

5.1.7

duo	milia	et	quingenti	equites
2	1000	+	500	horsemen (nom/acc)

equitum	duo	milia	et	quingenti
horsemen (gen)	2	1000	+	500

'The substantive is not put in the genitive if separated from *milia* by numerals that do not qualify *milia*' (Hayes and Masom, 1928, p. 60).

We now have two things to explain; a syntactic correlation between high/low and noun/adjective, and a morphological tendency for lower-valued words to exhibit more conditioned morphological variability, that is to have richer inflectional paradigms. The former actually helps to explain the latter. For some features, in particular gender and animacy, conditioned variability can only be manifested through the mechanism of agreement. A form will only exhibit variability in features such as gender and animacy if it occurs in syntactic environments which require it to agree with some other form, for example if the syntax treats it as an adjective. For other features, such as case and number, variability can be manifested via the mechanisms of agreement and also through other syntactic devices (such as government). Thus for case and number also, the adjectival syntactic behaviour of a form enhances the likelihood of it displaying variability for these features. In short, the syntactic treatment of low-valued forms as adjectives gives them greater opportunity to exhibit conditioned morphological variability. Some other factors reinforcing the tendency for low-valued numerals to have richer inflectional paradigms can also be suggested .

The very lowest-valued numerals, for 1, 2, and possibly 3, may well be integrated into a language as nominal modifiers of some kind (perhaps as singular, dual, and trial affixes) before the rise of a conventional recited counting sequence (as argued in Chapter 3), while the linguistic integration of the next few numeral words may follow their introduction via the conventional recited sequence. These words derive their numerical significance from their position in the recited sequence (as argued in Chapter 4). Use in a recited sequence would exert a pressure towards invariance, since the essence of a conventional recited sequence is its invariability. On the other hand, variable low-valued forms

would exert a pressure of analogy on higher-valued forms, but it seems reasonable to suppose that this analogical pressure would affect numerically adjacent forms most strongly and numerically more distant forms only weakly. Thus a tug-of-war along the number sequence would take place, with the original very low-valued variable forms pulling towards variability, and the higher-valued forms introduced from the invariable counting sequence pulling towards invariability. Analogical pressure towards variability would also be exerted from variable non-numerical modifiers, such as adjectives, and it seems likely that the numeral forms most susceptible to such pressure would be those that were already variable to some extent, that is the lower-valued ones. These analogical forces can only be weak and can be upset by other factors, so we observe only a tendency for lower-valued forms to show inflectional variability.

I turn now to the correlation between low/high numerical values and adjective/noun syntactic behaviour. At this point a further reservation about Corbett's and Jespersen's low/high-adjective/noun generalization must be stated. There are languages which are alleged to have no adjectives, but only the categories noun and verb. Nootka is often cited as such a language. In such a language, would low-valued numerals resemble nouns or verbs? Corbett says he would expect numerals to come between verb-plus-adjective and noun in such languages, but I don't know what principle generates this expectation. Paul Schachter (personal communication) writes 'Numeral stems in Wakashan languages (Nootka, Kwakiutl, etc.) are verbs'. Unfortunately, I have had no success in finding further facts about such cases, and will have to leave them undiscussed here. In Fijian, 'numbers behave like verbs' (Dixon, forthcoming), although they can also be used as nominal modifiers in a way which does not seem to conform fully to the rules for adjectives. Information on such cases is scarce, and again I will leave them undiscussed, as they do not represent the main tendency in the syntactic categorization of numerals. Many African languages have only three or four basic adjectives, that is adjectives that are not clearly denominal or deverbal formations; in such languages numerals are nevertheless usually syntactically distributed like these basic adjectives.

Corbett tentatively suggests an explanation for the low/high-adjective/noun generalization:

> It is less easy to see why numerical value should correlate with nouniness, though the beginnings of a solution can be

> suggested. In the course of history the need has arisen for successively higher numerals; nouns referring to a vague large number have taken on a specific numerical value, larger than that of the previously largest numeral. Thus the higher numerals are nouns pressed into service as numerals. An example would be Old Church Slavonic *t'ma* 'multitude' which came to mean '10,000'. As in the course of cultural development new numerals are introduced, naturally at the top of the earlier system, the previously highest numeral may be further integrated into the system and lose some noun-like features. One must still ask why the items pressed into service as higher numerals are always nouns. This regularity is linked to the notion of individuation. Nouns such as *t'ma* 'multitude' originally denote a number too large to grasp, conceivable only as an undifferentiated group. (1983, p. 245–6)

[Biblical Welsh *myrdd* is like Old Church Slavonic *t'ma*, sometimes indicating just a very large number ('myriad') and sometimes perhaps exactly 10,000 (Hurford, 1975, p. 138–9), but *myrdd* did not survive in later Welsh with a precise numerical meaning.]

This hypothesis, then, is that numeral words, invented in strict numerical order, all enter a language as nouns, but move towards being adjectives. The words that have been in the language longest (that is the lowest-valued) have moved the furthest along this route. So the static synchronic picture is like a snapshot of a road with travellers moving along it from their common landing-point (noun) towards their common destination (adjective); the intervals between them reflect differences in their arrival times at the original landing-point.

There is indeed evidence that in some languages (and language families), numerals *have* shifted historically away from nouniness and towards adjectivality. In Slavic, 'the numerals 5-10, 100, 1,000 have lost some of their noun-like properties' (Corbett, 1983, p. 236). The noun-like properties Corbett alludes to include syntactic, as opposed to morphological, properties. And in Biblical Hebrew 'the character of the numeral tended more and more to become adjectival rather than substantival' (Kautzsch, 1910, p. 432; this assertion is backed up by a detailed factual footnote). Here again, the properties alluded to are syntactic, specifically word-order, properties.

A problem with Corbett's tentative explanation (which he readily admits to) is that it is not clear why the high-valued vague terms which become precise numerals should necessarily start as nouns, rather than as adjectives or quantifiers, like English *numerous, many*. Corbett's suggested link with the notion of individuation is not detailed enough to be convincing. Also, one needs to explain why further integration into the system should necessarily involve loss of some noun-like features. In the next two sections I will explore the links between the meanings of words and their syntactic categorization, arriving at a picture which shows the category 'adjective' to be the naturally appropriate one for number words, but showing how higher-valued words, in particular, can come to be used as nouns.

5.2 Semantic Motivation of Adjectival Categorization

Why do numeral words universally tend to resemble adjectives or nouns in their syntactic properties? And why is there a correlation between the value of a numeral and its syntactic behaviour, as an adjective or as a noun? I assume that the *meanings* of numerals are the most significant determinants of their syntactic categorization. In this section I shall develop an argument that the meanings of numerals make 'nominal modifier' (adjective) the primary most natural syntactic category for them. And in the next section I will show that numeral meanings also naturally give rise to a secondary kind of use, as nouns, and the simple arithmetic of numbers tends to select only the higher-valued words to fulfil this secondary function.

Hopper and Thompson (1984) discuss the origin of the universal syntactic categories 'noun' and 'verb', and distinguish between semantic and discourse factors leading to the categorization of a given item as a (more or less prototypical) noun or verb. The distinction is approximately this. Semantic factors are permanent intrinsic features, or 'propensities' of the meaning of a word, while discourse factors involve features attaching to a word by virtue of its function in a particular discourse. Thus, 'denoting a set of middle-sized concrete animate objects' would be a semantic feature of *fox*; while 'used generically' would be a discourse feature of the same word in *The fox is a cunning animal*. Hopper and Thompson's thesis on the origin of syntactic categorization is that it is partly rooted in the semantic propensities of a form, and partly, they emphasize, in the typical use, or discourse function, to which a form is put. The role of semantic factors in syntactic categorization has been known and discussed for a long time (see, for example, Lyons, 1977, pp. 438–52); the idea of a distinct role for discourse factors is new with Hopper and Thompson. Their distinction between semantic and discourse features is probably methodologically useful, even if it is not easy to apply in all cases. Strawson (1959), in a particularly profound and thought-provoking study, discusses the relation of linguistic categories to our conceptual picture of the world in a way which gets to the common bases of both discourse and semantic factors. Gupta (1980) also relates the syntactic class 'common noun' to a semantic property in a way which closely echoes the conclusions both of Strawson and of Hopper and Thompson, although none of these writers actually refer to the others. I will discuss the syntactic categorization of numerals in both semantic and discourse terms, starting with what Hopper and Thompson would classify as a discourse factor.

Reciting the counting sequence is a verbal calculating trick, or rule of thumb, by which a speaker arrives at conclusions which he may then assert. The conclusions pertain to the cardinality of collections or to the position of objects in sequences. The collections or sequences involved are present to the attention of the calculating speaker *before* the conclusions which he draws from the counting activity. Although one can count without having anything to count (for example 'Can you count to 10 in Latin?'), one can only draw an inference (other than one about the speaker's abilities) from the word arrived at in some counting activity in the context of some collection or sequence of objects being counted. So, in the mind of the speaker during the practical use of numeral words, the objects counted preexist the numerical conclusions about them. And if the speaker is to communicate his conclusions informatively to other speakers, the objects themselves are also likely to be present to the attention of his interlocutors, who are likely to be unaware of the numerical conclusions he is to express.

Consider also the relative naturalness of the following two questions: (1) How many children have you got? (2) What have you got six of? To answer question (1), one simply has to assemble the objects concerned (one's children), either physically or mentally, and count them; to answer question (2), one has to go through a process of trial and error, assembling various

collections of things and counting them until one finds a collection numbering just six. To use a computational metaphor, our knowledge about collections of objects is not indexed by their numerical properties, but rather by what nonnumerical categories (for example, child, stick, pebble) the objects belong to. This presumably follows from the fact that the cardinality (above 3) of collections cannot be perceived directly without counting or calculation of some sort, whereas other categories are, or may become, categories of direct perception. (These observations apply only, of course, to what may be called the 'situation-new' uses of numerals in larger constructions, in particular not to cases of reported usage, as in *John told me that he has six children*.)

The correlation between sentence structure, in terms of subject and predicate, and the structure of the world as it impinges on human consciousness is, of course, extremely complex and elaborate. But it seems safe to claim that one principle at work, among many others, is that if a connection is asserted between elements of two categories (object, state, action, property, and so on) and one element is less salient to perception than the other, then the term corresponding to the less salient element tends to occur in the predicate part of the assertion, and the term (usually a nominal) corresponding to the more salient element tends to occur in the subject part of the assertion. Thus the 'fiveness' of a handful of pebbles is a property less salient than the 'pebbleness', and so we would expect it to be more natural to assert 'These pebbles are (number) five' than something like 'Five is this collection of pebbles'. One cannot experience fiveness without first experiencing some collection. The numerical property is what one predicates of some given collection.

Interpreting this in Hopper and Thompson's terms, forms indicating the cardinalities of collections would tend to be less 'manipulable' in discourse than, say, forms indicating middle-sized concrete objects.

> Forms which are presented as playing a role in the discourse – which are manipulable or 'deployable' there – are universally able to adopt the full range of grammatical attributes of N's. . . . A N which falls short of fulfilling this pragmatic function is often, as we have shown, 'less of' a N in the range of morphological oppositions which it can manifest. (1984, p. 718)

On discourse grounds, then, one can expect numerals not to exhibit a full range of nouny characteristics.

The earliest numeral words are often borrowed from body-part terms, words for the hand, various fingers, and sometimes elbows, shoulders, eyes, ears, and so on (see Lean, 1985–6; Saxe, 1982a, b). As body-part terms, these words would be nouns. Hopper and Thompson (1984, pp. 724–6) note that in many languages body-part terms are not fully nouny, a fact which they explain in terms of body-parts tending to belong to the discourse background, rather than being salient, fully manipulable objects. But when such words are borrowed for use as numerals, they take on completely new denotations. Thus, for example, in Melamela – a language of New Britain (Lean, 1985–6, Vol. 4) – the form *lima-* denotes the set of hands (that is means 'hand'), but, pressed into use as a numeral, *lima* denotes the set of collections of five things (that is means '5'). With such a large difference in meaning, one would not expect the borrowed form necessarily to retain aspects of its former morphosyntactic behaviour. In seeking semantic explanations for the syntactic categorization of numerals, we have to look at the meanings of the numerals themselves, rather than of their etymological antecedents.

Predicates of all sorts, whether they function grammatically as nouns, verbs, adjectives, or numerals, have denotations. For example:

5.2.1 WORD	CATEGORY	DENOTATION
cat	noun	the set of all cats
sleep	verb	the set of all sleeping things
red	adjective	the set of all red things
five	numeral	the set of all collections of five things

The denotation of a numeral is a set of collections. The prototypical members of the collections denoted by numerals are physical objects. Children learn to count physical objects first, and events (such as noises), actions (such as kicks), and more abstract entities (such as aspects of a problem) later. I will restrict the discussion to the basic case of numerals as denoting collections of physical objects. This will not harmfully distort the argument.

Why, first of all, do languages never (except languages such as Fijian and perhaps Nootka, see previous section) treat numerals like verbs? One can easily conceive of a hypothetical language in which the numerals behave like verbs. In such a language, the translation of *There were eight in the team* would be something like *The team eighted*; the downgraded predication could be handled participially, so that *a team of eight* would be *an eighting team*. What the elements in the denotation of a verb, in any language, typically have in common is participation in some particular dynamic process, action, or activity (for example, walking, sneezing, assembling, dispersing, circulating) – this is uncontroversial: see Sapir (1921, p. 119), Brown (1958, pp. 249–52) – or, somewhat less typically, the fact that they are in some particular state (for example, sleeping, sitting, standing). Clearly, the dynamic notions are in no way involved in the meanings of numerals. But what of the notion of state? One might safely say that for a collection to have some particular number of members is for it to be in a certain state. Thus, the denotation of *five* is the set of all collections that happen to be in the state of having five members. However, the typical states involved in the meanings of (intransitive) verbs (such as *lie*, *sleep*, *sit*, *stand*) are volitional states of animate subjects. Clearly, numerals do not necessarily or typically involve states achieved by the volition of animate subjects.

(I assume that the uses of verbs such as *lie* and *stand* with inanimate subjects, as in *the pencil is lying on the desk*, are derivative of the basic uses involving animate subjects, and are quasi-metaphorical. When such verbs are applied to inanimate subjects, they often lose much of their meaning. Thus *The house lies|sits|stands in the valley* could all describe the very same situation, with slightly different poetic overtones. But, applied to an animate subject, as in *There is a man lying|sitting|standing on the stairs*, the verbs retain quite distinct meanings.)

Having eliminated verbs (and prepositions, which are typically inherently relational), we are left with the categories of noun and adjective. Many nouns and adjectives express quite similar concepts to those expressed by numerals. There are nouns, for instance, which, like numerals, denote collections of objects (for example, *group*, *team*, *party*, *gaggle*); and there are many adjectives which describe non-volitional states, even states, like that described by *numerous*, very similar to those involved in numeral meanings. So, fairly naturally, numerals figure in the syntax of languages in roles similar to those of adjectives and nouns. I will now push on into the question of the somewhat greater appropriateness to numerals of the category 'adjective', as opposed to 'noun'.

Kamp (1975) makes an observation on a semantic difference between adjectives and nouns. Kamp wonders why many adjectives, but hardly any nouns, have a comparative form. *This is bigger than that* is fine, but *this is more of a table than that* is odd. Kamp suggests that adjectives are often 'one-dimensional' in the sense that for an object to satisfy an adjective, it often need only satisfy a single criterion, for example, of being red, or hot, or whatever. On the other hand, 'in order for an object to satisfy a noun it must in general satisfy all, or a large portion of, a cluster of criteria. None of these we can promote to the sole criterion without distorting the noun's meaning beyond recognition, (Kamp, 1975, p. 148). Clearly, possessing a particular cardinality is a single criterion, or dimension. Thus in this respect numeral meanings resemble adjectival meanings rather than nominal ones.

Strawson's work *Individuals* (1959) is in a style and tradition utterly different from the work of linguists such as Hopper and Thompson. Yet clearly they are addressing very much the same issue. Strawson is concerned with the identification and reidentification of particulars by speakers and hearers in discourse. Although Hopper and Thompson make no mention of Strawson, their concern with forms which are readily 'manipulable' clearly resembles Strawson's concern with the ways in which language users reidentify particulars.

Strawson make an interesting distinction between two types of universals (and hence of predicates), a distinction which appears to be related to grammatical categorization. Strawson's distinction is between 'sortal' universals and 'characterizing' ones. 'Roughly, and with reservations, certain common nouns for particulars introduce sortal universals, while verbs and adjectives applicable to particulars introduce characterizing universals' (1959, p. 168). Chomsky (1976, p. 44) also picks up the connection between 'sortal predicate' and 'common noun'. The topic is elusive, because of its nearness to the bases of ontology and epistemology, and Strawson's discussion is delicate and cautious. So, in using his distinction to any argumentative purpose, one must also be cautious (and try to be delicate!). Wright (1983) has elaborated somewhat on Strawson's notion of 'sortal' and related it specifically to numbers.

Strawson's notion of a sortal is this (not in his words). If a universal applies to a particular of necessity throughout its whole existence, the universal is a sortal. So *dog*, *terrier*, *animal* are sortal predicates. As soon as a dog comes into existence, it is a dog. When it dies it becomes a dead dog. If a dog is vaporized, it ceases simultaneously to exist and to be a dog, or to *exist as a dog*. On the other hand, a brown thing can cease to be brown without ceasing to exist. So *brown* is not a sortal, but a characterizing predicate. The tie between a particular and a universal of the sortal kind is an 'instantial' tie: Fido is an instance of the dog universal. The tie between a particular and a universal of the characterizing kind is itself 'characterizing': Fido may be characterized as brown. If Fido were to change his colour from brown to white, it would seem odd or infelicitous to say that Fido had 'ceased to exist as a brown thing' or 'as something brown'.

Strawson's wider discussion on the bases of understanding assertions about particulars emphasizes the importance of the possibility of reidentifying particulars. Wright, explicitly and perhaps less circumspectly, draws a close connection between sortality and reidentification. 'Understanding a sortal concept involves ... an understanding of what it is for any object, *a*, exemplifying the concept to be the same as, or distinct from, any object, *b*, exemplifying the same sortal' (1983, p. 2). Gupta, without using the term 'sortal', also make the link between the syntactic category 'common noun' and reidentification.

> We have seen that common nouns differ from predicates in that they have associated with them, as part of their intension, a *principle of identity*. A common noun such as 'man' divides all objects (in a world and at a particular time) into those of which it is true and those of which it is false; and in this respect it resembles predicates. These, too, have associated with them, as part of their intension, a *principle of application*. But common nouns have more to their intension than a mere principle of application. The common noun 'man', for instance, provides a principle for determining when an object at time t (and/or world w) is *the same man* as an object at time t' (and/or world w'). (1980, pp. 67–8)

Gupta's classification of common nouns outside of predicates can be taken as a terminological matter. What is important is their additional ingredient involving (re)identification, which Strawson and Hopper and Thompson, in their various ways, also consider crucial. I will continue this discussion with more explicit reference to Strawson.

Obviously the notion of sortality is difficult. The territory Strawson lays out may be a minefield or a quagmire for the analyst, but I do not think it is a mirage. There is something to the notion. Obvious problems are fairy-tale happenings, like people being turned into animals, which seem to tell us that *person* and *dog* cannot be sortals. Without going to fairy-tales, there is the fact that caterpillars become butterflies, and tadpoles frogs. A caterpillar ceases to be a caterpillar without ceasing to exist, and so fails our test for sortality, yet caterpillars make very much the same kind of impression on us as dogs do, in respect of objecthood, permanence, and so on. But surely it is rather *surprising* that caterpillars and tadpoles should be so radically transformed. This is why these creatures are such hardy annuals in the primary school curriculum. The notion of sortality seems to have to do, then, with a normal conceptual scheme which we, to some extent, impose on the world, rather than the world simply presenting it to us.

Are particular numeral predicates sortal? Can a pair of things cease to be a pair of things without ceasing to exist? Take a pair of ordinary pencils. Send one off to Alaska and keep the other one here. They are still in some sense a pair. Burn one. That particular pair has ceased to exist. But add a further pencil to the original pair. On the one hand it seems acceptable to say that the pair has *become* three. But on the other hand, the original pair can still be identified, as part of the new threesome, and thus, clearly, has *not* become three. The problem is that we are using the expression *the pair* in two different ways here. In the sense in which 'the pair becomes three', the twoness is incidental to the collection, whose more permanent identifying features presumably involve the fact that it is made up of pencils, and that we are talking about it. In the sense in which 'the pair is part of the threesome', we have chosen to regard the twoness as the more permanent identifying feature. (But it is not a sufficient identifying feature, as there are two other distinct twosomes of pencils in the newly formed threesome.)

It seems reasonable to talk of a group of a hundred soldiers becoming a-hundred-and-one soldiers by a new soldier joining their ranks. Here one might wax somewhat pedantic and say that by the addition, the original group of a hundred has indeed ceased to exist, and has been replaced by a group of a-hundred-and-one. But it is impossible to deny some kind of continuity of existence here. Something which was a hundred has become not-a-hundred without ceasing to exist. The continuity is maintained by the fact that the group's members are all soldiers, indeed, except for one, the same soldiers as before. It is not clear that the addition of a soldier would prevent reidentification of the particular group.

A collection with a given cardinality can certainly be viewed as a whole self-contained object, just as a dog can. But the cardinality of a collection is not crucial to reidentifying that group on a second occasion: other properties may suffice, in particular properties of its individual members, or the manner of their distribution over space. Indeed, the cardinality of what is perceived, and treated in discourse, as a continuously existing particular collection may change over time, just as a particular dog's colour can conceivably change. And we can reidentify a collection as the same without knowing its cardinality, although the cardinality can be a useful clue in the same way as the colour of a dog can be a useful clue. But we could not reidentify a dog as the same object we had seen before without in some way ascertaining, consciously or unconsciously, that it was indeed (still) a dog. The particular cardinality of a collection at a given point in time seems to bear the same relation to the collection as a dog's colour bears to the dog. Twoness, fiveness, hundredness are, like brownness and hotness, incidental properties inessential to reidentification. In short, numbers, like colours and temperatures, seem to be characterizing rather than sortal. Hence, with reservations, one would expect numerals to behave less like common nouns, and more like adjectives.

To sum up: in Strawsonian terms, the cardinality of a collection is not normally sufficient for reidentifying it as a particular; in the alternative, and not exactly equivalent, terms of Hopper and Thompson, numerals are normally not markedly manipulable terms in discourse. So they are not good candidates for noun status. As Frege insisted, an aggregate can only be associated with a unique number when brought under some other concept. A particular aggregate has as many cardinalities as there are ways of dividing it into parts. But if the type of the parts into which a given collection is to be divided is specified, for example, 'apples', 'apple-quarters', 'basketfuls', then the collection usually has just one cardinality brought under that concept. Hence, a numeral is normally only useful when associated with the word for the concept relevant to the occasion of use. Such words are usually nouns, denoting sets of individual objects. So numerals very naturally fall into a category for satellites to nouns – that is attributive adjectives. This is not to say that they cannot be used as predicative adjectives or as nouns, but to claim that the attributive adjective slot is their natural niche.

[The adjectival niche, which numerals naturally evolve to occupy, may well have other, rather different inhabitants. There are many different kinds of adjectives, whose semantic interactions with their head nouns are of quite varied types – see Siegel (1980), Kamp (1975), and Dixon (1977)). In Siegel's terms, numerals are of the same type as adjectives with 'intersective' meanings.]

The conclusion just reached relates to the ordinary use of numerals as in *five bricks*, but not to the abstract use of numerals in the formulae of arithmetic, as in *seven plus two equals nine*. Interpreted abstractly, as in these arithmetical contexts, where all information about collections except for their cardinality is suppressed, or abstracted away from (as argued in Chapter 4, Section 3), the cardinality naturally becomes crucial. Some collection of physical objects may be able to maintain its identity over time while changing its cardinality, but it is clear that a number considered as an abstract arithmetical entity cannot change by becoming greater or less. 5 = 5 = 5 and can never become 6. Correspondingly, in these usages, we find, as is to be expected, that the numerals are not adjectival, but nominal. Indeed, they are 'extremely' nominal, being in effect used as proper names. In fact, numbers considered as abstract entities, rather than as (incidental) properties of collections, are probably sortal, as Wright argues, 'number-theoretic platonism is just the thesis that *natural number* is a sortal concept' (1983, p. 4). The fact that abstracting the arithmetical notion of number from experience of counting collections involves such a radical conceptual shift, from characterizing to sortal, almost certainly contributes to the difficulty of grasping arithmetical notions, which have to be deliberately inculcated into the young. The nouny nature of numerals as names of abstract arithmetical entities helps considerably to explain the fact that it is the higher-valued numerals which, if any, tend to behave as nouns. This is taken up in the next section.

5.3 Collections as Objects, Numerals as Nouns, and Multiplication

The intuitive notion of 'collection' has various properties that may be argued to give rise to the development of constructions expressing multiplication. Five relevant properties, implicit in much of our discussion so far, are now given explicitly below, expressed in a way reminding of the relation between aggregates and collections.

5.3.1 COLLECTION/OBJECT BIVALENCE

Any aggregate from which a collection can be constructed can also be (constructed as) a single object, which can itself be a member of a larger collection. Hence, there can be collections of collections.

5.3.2 TRANSITIVITY OF MEMBERSHIP

Any aggregate from which can be constructed a collection of collections of objects, can also be taken for constructing a collection of those objects. Thus an aggregate can be in the denotation of numerals of different values.

5.3.3 DISJOINTNESS OF MEMBERS

In the construction of a collection of collections, the member collections do not overlap, that is they share no common members.

5.3.4 PLURALITY OF COLLECTIONS

Every collection has at least two members. There are no one-member collections.

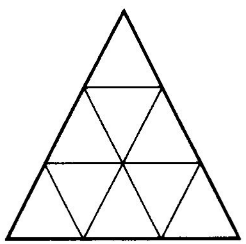
5.3.5 IRREFLEXIVITY OF MEMBERSHIP

Nothing is a member of itself.

Principles (5.3.1)–(5.3.5) may seem natural enough to a layman innocent of set theory. In fact, they draw a picture which is clearly at odds with classical set theory. Set theory insists on a distinction between 'member' and 'subset'; a member of a set is *not* a subset of it. (5.3.1)–(5.3.5) override this distinction, appealing only to the notion (proper) 'member'. The theory of aggregates and collections as depicted here will be shown to provide a basis for explaining the development of several numeral constructions. In particular, (5.3.1)–(5.3.3) will be used in this section to explain the rise of 'nouny' multiplicative constructions. (5.3.4)–(5.3.5) above were mentioned to reassure the reader of the consistency of the proposed scheme. In subsequent sections further principles of the nature of collections will be spelt out as needed to explain numeral-adjective word order and additive constructions. The picture adopted here is close to the ontology proposed by Link (1983, forthcoming) and also to the basic intuitions about part-whole relations appealed to by Leonard and Goodman (1940). None of these authors, however, discuss the construction of collections from aggregates.

The assumption in (5.3.3) of the disjointness of members of a collection can be justified in psychological terms. The most prototypical collection of collections is a thing like a group of piles of coins on a table, in which the subcollections are disjoint. A collection of (partially) overlapping collections is harder to conceptualize. A graphic example is the following diagram. Count the triangles in it.

5.3.6



When one has not seen a puzzle like this before, one's first natural reaction is to count the non-overlapping triangles. Counting triangles which overlap with other triangles is clearly more difficult. The prototypical object in a collection has (or is psychologically regarded as having) its own proper 'skin' or defining boundary which never crosses, and is never shared with, the skin of a neighbouring object, though it may touch it. [Diagram (5.3.6) thus does not represent a typical collection, because a single (ideally widthless) line serves as a boundary for more than one triangle.] A platoon of soldiers on a parade-ground is an aggregate from which may very readily be constructed a collection of three rows of ten soldiers, or a different collection of ten files of three soldiers. (Rows and files are kinds of collections.) But no single collection of subcollections can be constructed in which a given soldier is a member of more than one subcollection.

A consequence of (5.3.2) is that aggregates do not necessarily have unique cardinalities. A given aggregate may have many different cardinalities, depending on the concept under which it is brought, as argued by Armstrong (quoted above in Chapter 4, Section 1) in response to Frege. Thus an aggregate consisting of five piles of ten coins on a table has (at least) the cardinalities 5 and 50. It falls in the denotations of both *five* and *fifty*. It also falls in the denotations of the plural nouns *piles* and *coins*. Only when numerals are used with nouns can anything numerically informative be said about the aggregate on the table. The aggregate on the table falls in the denotation (derived by set intersection) of *five piles* and of *fifty coins* but not in the denotation of *fifty piles* or of *five coins*. And, crucially for the argument of this section, the aggregate in question also falls within the denotation of *five tens* and, more informatively, of something like *five tens of coins*.

Psychologically, whether or not something is treated as an object or a collection depends on the kind of attention one is giving it at the time. Thus, the pile of raspberries on the kitchen table could be treated as a single object if I am setting the table for a meal, avoiding putting plates, cutlery, and so on, in contact with it. But the same pile is treated as a collection if I am working over it systematically, inspecting each raspberry for creepy-crawlies.

(Having been careful so far in this section to distinguish between aggregates and collections, I will revert below to the less cumbersome shorthand usage in which the denotations of collective and plural nouns and numerals are said to be sets of collections.)

The prototypical denotation of a noun is a set of objects. Since a collection can be an object, a set of collections can be the denotation of a noun. Familiar nouns whose denotations are sets of collections include *army*, *team*, *group*, *committee*, *flock*, *herd*, and *collection* itself. The characteristic property of some such collections is the category to which individual members belong (for example, sheep, cow, and so on); the characteristic property of others is the function which the collection typically serves (for example, to fight wars, play games, make decisions, and so on). *Group* and *collection* are superordinate nouns for collections with, in ordinary language, slightly different implications about the way the members have come, or been brought, together.

The characteristic properties of collections denoted by collective nouns often include specific cardinalities. *Pair*, *trio*, and *quartet* are examples.

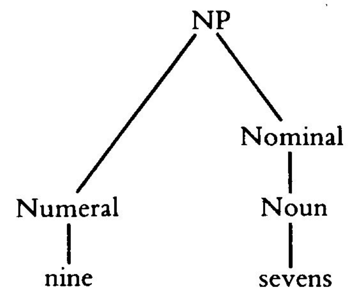
> A number word is often used for a thing that is made up of the number of parts shown by the number word used ... Thus ... a *five* is used for a basketball team (which has five men), an *eight* for a racing crew in rowing, a *nine* for a baseball team, and so on. (*Webster's New Collegiate Dictionary*, 1959, p. 576)

Examples such as these can be found widespread across languages, and probably in all languages which have numerals. Familiar definite referring expressions also use numeral words as nouns, as in *The Magnificent Seven*, *The Famous Five*, *The Top Twenty*.

These types of example are frequent, and familiar from particular contexts, for example, of some specific sport. In such contexts, use of a numeral as a noun has been conventionalized almost to the point of being idiomatic. But in general, in any context, languages allow the ad hoc use of numeral words as nouns where there is a felt need to refer to collections with a specific cardinality. Thus, given, say, a need to refer to collections of seven things, and an existing numeral lexicon *one*, *two*, ..., *six*, *seven*, *eight*, ..., it is possible to treat *seven* as a noun, as well as a numeral. This gives rise to expressions such as *that seven over there* (note the singular demonstrative), *those sevens on your right* and *my sevens* (note the plural suffixation). Clearly, such expressions would only be useful in a context where the actual categories (for example, pencils, marbles) of the objects in the collections could be relegated to the background of the participants' attention and the numerical properties of the collections involved temporarily assumed greater salience. This would be a relatively unusual situation, since, as argued earlier (Section 5.2), numerical properties of collections are usually less salient than the types or categories of the objects comprising them, but such situations obviously do occur, and indeed are permitted to occur by the prior development of numerals as expressions denoting sets of collections with particular cardinalities.

Given the emergence of numeral words as nouns, the kind of structure shown in (5.3.7) becomes available.

5.3.7

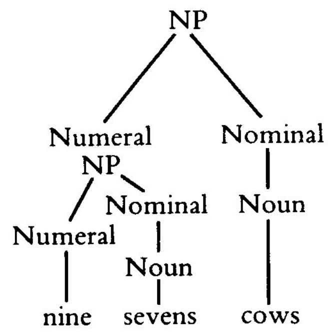


The denotation of this expression, by intersection (4.3.5) and the principle of Collection/Object Bivalence (5.3.1), would be the set of all collections of nine collections of seven objects (that is, the intersection of the set of collections of nine things with the set of collections of collections of seven things). This would be equivalent, by the principle of Transitivity of Membership (5.3.2), to the set of all collections of 63 objects, assuming (5.3.3), which ensures the disjointness of all the collections involved. Reverting to the strict aggregate/collection distinction, the aggregates in the denotation of (5.3.7) are the set of aggregates from which a collection of nine collections of seven objects can be constructed. Such aggregates can also be used to construct collections of 63 objects. The interpretation of such a structure is, evidently, equivalent to arithmetical multiplication. Simple sentences expressing arithmetical truths involving multiplication, for example, *four twos are eight*, can now be assigned the value 'TRUE'.

The account of the rise of multiplicative constructions is not in any sense built on an account of the rise of additive constructions (which will be discussed in the final section of this chapter). So the evolutionary relationship between addition and multiplication is not as conveyed by the usual picture of multiplication as serial addition. Multiplication emerges from pluralization, and addition from conjunction. In principle, although both multiplication and addition arise, I claim, from the same psycho-ontological scheme of aggregates and collections, a language could possibly develop multiplicative constructions before additive constructions.

Situations will naturally arise where a speaker wishes to refer to some large collection, drawing attention both to its cardinality and to the category of the individual objects in it. Given NPs, like *nine sevens*, which in effect express higher numbers, it would be quite natural to treat these expressions also as numeral nominal modifiers. This might be done in a variety of ways, of course, but the most obvious possibility is to parallel the existing numeral-plural constructions, as in *five bricks*, simply placing the NP *nine sevens* in the Numeral slot. So, one might expect structures such as (5.3.8):

5.3.8



The complex node [Numeral, NP] here seems the most straightforward way of representing the fact that an existing NP is inserted into an existing Numeral position. The NP *nine sevens* is interpreted as (5.3.7) before, and its interpretation is taken as the value of the Numeral which contributes to the interpretation of the whole NP *nine sevens cows*. By the rules and principles already developed, the denotation assigned to this structure would be the set of all collections of 63 cows. The fact that a speaker chooses to use this expression, rather than, say *seven nines cows* might possibly carry a pragmatic implication that the cows are assembled in groups of seven, but this matter will not be pursued here. Structures of the type of (5.3.6) will be further discussed in the next section in connection with numeral classifier constructions. For the present, it is sufficient to note that we have seen how a [Numeral, NP] phrase, containing a nouny numeral, can naturally arise. (I do not claim, of course, that the actual expression *nine sevens cows* is acceptable in English; the argument is about the development of multiplicative structures in general, and *nine sevens cows* is a schematic example of the whole class of such structures, which will differ in details, such as word order, the inclusion of structure markers such as English *of*, and so on.)

An explanation has been given for the emergence of the use of numeral words as nouns in constructions interpreted by multiplication. It remains to be shown why the numeral words pressed into this nouny use should be just the higher-valued ones. Two distinct cases must be considered.

One kind of case of a numeral used as a noun is the case of a word from the original short continuous numeral lexicon (Chapter 3) being so used. Say a community develops a sequence of counting words up to the value 10. (I will use the English words for convenience.) Now in principle, any of these words can be pressed into nouny use in multiplicative constructions, for example, *two fives*, *five twos*, *seven tens*, *ten fours*, and so on. Thus a very wide range of alternative expressions becomes available to express a relatively small range of numbers. In fact, numeral systems tend strongly to standardize on a small set of multiplicative base numbers. For reasons of economy, the multiplicative base word chosen from the original short counting sequence is almost always the last (highest-valued) word (10 in a typical case). The mechanics of this process of standardization on the highest number word available as a multiplicative base will be discussed in detail in Chapter 6. Once a form for, say, 10 has been adopted as a nouny numeral in a multiplicative construction, it may, in the subsequent history of the language, lose its nouny characteristics as the whole construction becomes idiomatized and its internal structure becomes opaque.

The other kind of case of numeral words used as nouns is the case of much higher-valued words, usually powers of the base number (for example, 100, 1000), again used in multiplicative constructions. Such words are presumably 'invented' and given their precise values at a historical stage after the development of the initial short lexical sequence. How can the values of such words be fixed so exactly? I suggest that the values become fixed by the adoption and transmission to subsequent generations of quite explicit and abstract verbal definitions, such as *a hundred is ten tens* and (later) *a thousand is ten hundred*, and paraphrases of these. By the time one is dealing with such large numbers, the umbilical tie to physical examplars in the form of small collections of concrete objects is impractically stretched. Progress in the fixing of exact values at this high level can only proceed via the emerging abstract arithmetical language game played with bare numeral NPs, as in these definitions. Being originally introduced in this way, as names for abstracted entities, and not as nominal modifiers, words for high numbers are naturally nominal. Again, these too may shift historically away from this nouny behaviour, to the extent that they become commonplace and frequently used in complex numeral constructions, modifying nouns.

5.4 Numeral Classifier Constructions

In this section, I take up the nouny structures, like (5.3.8), interpreted by multiplication, and show how such structures are essentially the same as the numeral classifier structures found in many languages, and interpretable by the same semantic rules.

In English and many other languages, the noun in the middle of a construction such as *five hundred (of) cows* can only be a noun formed from a numeral word, or a collective such as *group*. *Seven cows sheep*, for example, would not be interpretable in the required way by the semantic rules proposed. Technically, the interpretation assigned to *seven cows sheep* would be the empty set. This would be arrived at in the following way. The denotation of the constituent *seven cows* is the set of all collections of seven cows, by rule (4.3.3); the denotation of the whole expression *seven cows sheep* is the intersection of this set with the set of all collections of sheep. Since no collection of sheep is also a collection of cows, the intersection of the two sets is empty. But this brings out the point that if a non-numeral noun were used whose denotation intersected non-emptily with that of the second, modified noun, there would be a nonempty interpretation of the whole expression. So, for example, *seven animals sheep* would receive, by the machinery described here, the same interpretation as the less redundant *seven sheep*. In fact, a large number of languages allow expressions of just this kind. These languages are known as 'numeral classifier languages'.

Greenberg, in a wide-ranging and stimulating discussion of numeral classifier languages, gives the following comment and preliminary definition.

> A considerable number of the world's languages including almost all of these in Southeast Asia exhibit the following characteristic. An English phrase such as 'five books' is rendered in translation by a phrase containing, outside of possible grammatical markers, not two but three elements. The kind of literal translation often supplied in grammars of such languages might be something like 'five flat-object book'. The second item in such a phrase is often called a numeral classifier in allusion both to its occurrence in a numeral phrase and to its providing a semantic classification of the head noun. (1972, p. 2)

I suggest that the evolution of such numeral classifier constructions is facilitated by the independent development of [[Numeral Noun] Noun] constructions, as in (5.3.8), interpreted by semantic apparatus which, as it happens, assigns non-empty interpretations when the middle Noun is a superordinate term to the modified (head) Noun.

Greenberg hastens to point out that there is, within so-called numeral classifier languages, a great variety of particular characteristics which show that this simple preliminary definition rarely represents the whole story. The construction may be optional or obligatory, it may be more or less widespread in the modified Nouns it is used with, the implicit semantic classification may be more or less transparent, the order of numeral and classifier may vary (though they are always adjacent), and so on. The classifier element is not always a superordinate term to the modified Noun.

> The word for 'tail' is sometimes used as a classifier for animals (e.g. *ekor* in Malay) but we cannot consider a dog a kind of tail though of course we can devise a property 'having a tail'. On the other hand we could define the class meaning of *ekor* in Malay as that which is common to all nouns which take *ekor* as a classifier.
>
> Furthermore in some languages such as Burmese and Thai, there are a fair number of words which are, as it were, their own classifiers. An example is Burmese *ʔein ta-ʔein* 'house one-house' in which *ʔein* in its first occurrence is a head noun and in its second occurrence a 'classifier'. (1972, p. 3)

Of these kinds of examples, the former, something like *five tail dog* (ignoring the lack of plural suffix), could not be interpreted by the semantic rules here, as the set of tails does not intersect with the set of dogs. But the latter kind of example, in which a noun 'is its own classifier', as in something like *five house house* could be interpreted by our rules, as, obviously, the set of houses does intersect with itself. Towards the end of his article, Greenberg summarizes the semantic relation of the classifiers to the head nouns.

> There seem here to be three main types. A) Superordinate terms such as 'person' as a classifier for humans and 'tree' for individual 'species'. B) Items in one-to-one relation to

> the objects being counted; among the most common of these are 'head' for animates and 'trunk' or 'stalk' for trees. C) Words which themselves designate arbitrary or insignificant units like 'piece', 'grain', etc. (p. 17)

Greenberg's Type A classifiers, the superordinate terms, present no problems for the interpretation rules proposed here.

The type B classifiers, such as, *head, tail, trunk, stalk,* which are in a one-to-one relation with their head nouns, cannot be straightforwardly interpreted by the semantic rules proposed. But it would not be implausible to take into account a process of synecdoche, by which words for parts of things, such as *tail, head, trunk, stalk,* and so on, take on an extra sense, synonymous with the words for the wholes to which they belong, such as *animal, plant,* and so on. If such an interpretation of these classifiers is allowed, the semantic interpretation of numeral classifier expressions such as *three tail dog* can proceed like that of *three animal dog*.

The type C classifiers, words such as *grain, piece* and so on, fit well with Greenberg's own hypothesis for the evolutionary origin of numeral classifier constructions. Greenberg conjectures that the classifier words have evolved from words functioning as 'individuators' before 'collective' nouns, where by 'collective' he means a class of nouns which collapses the mass/count distinction and in which the singular/plural distinction is weak; an English example of such a collective is *cattle*.

> We have seen what might be called, anthropomorphically, the aversion of collectives to direct construction with a numeral and the intervention of an individuated noun, the classifier, as one of the devices to avoid this direct confrontation. This aversion has, therefore, as its natural counterpart, the corresponding attraction to the classifier and an immediate constituent structure in which the numeral goes directly with the classifier while the numeral + classifiers combination as a whole enters into a more remote construction with the enumerated noun. (p. 14)

Greenberg's account here and in a later paper (1975) postulates a constituent structure [[Numeral + Noun] Noun] for numeral classifier constructions, exactly as in (5.3.8), assuming, as Greenberg does, that the classifiers are derived from Nouns. His account in terms of an evolution from constructions such as *five grains of rice* is plausible, especially for classifiers such as *grain, piece,* and so on, which are used with materials which fall, psychologically, on the mass/count borderline, such as rice, (swarms of) gnats. Although individual grains and insects can be picked out, in practice there is seldom any need to, and the collections are treated more or less as portions of stuff, like water. But it is less obvious why well individuated objects such as books, houses, largish animals, and people should be named by words which have an aversion to direct construction with a numeral, so Greenberg's account is less compelling here.

Greenberg's account and mine are not exclusive alternatives; in fact they complement each other. The form of explanations in evolutionary terms does not require a single causing factor. Rather, the more facilitating factors that can be pointed to for the rise of some construction, the better that construction is seen to fit into its ecological niche. The rise of [[Numeral Noun] Noun] constructions, where the middle noun is adapted from a numeral word, as in *six fours dogs*, facilitates the (perhaps simultaneous) rise of cases in which the middle noun is of some other sort. Given the semantic rules for assigning denotations to such constructions, other nouns can fill the middle slot and give rise to nonempty interpretations if they are semantically superordinate to the modified noun.

The convergence of Greenberg's account of classifiers and my account of multiplicative constructions on the structure [[Numeral Noun] Noun] is shown neatly by the following facts:

> It is particularly common for classifiers not to occur with higher units of the numerical system and their multiples e.g. 10, 20, 60, 100, 300. (1972, p. 3)

Burling reports in regard to the most common unit counter of Burmese (1965, 262), the so-called general classifier *-khù*, that it is included by some Burmese speakers 'in the same series as the classifiers for the powers of ten ... *-khù* indicating only one individual object'. It was noted earlier that multiples of higher numerical units do not take classifiers. This also occurs in Burmese and shows clearly the function of the unit classifier as meaning 'times one'. Thus in Burmese 'two-ten book' = 20 books, i.e. 2 × 10 books while, following the interpretation by native speakers just cited 'two-*khù*

> (classifier) book' = 2 books, i.e. 2 × 1 "books". Many analysts consider words for 'ten', 'hundred' etc. in these languages as a subtype of classifiers. (pp. 5–6)

Thus, given the structure [[Numeral Noun] Noun], the middle noun may be either a multiple – that is a noun adapted from a numeral word – or another classifier. If the multiple constructions and the classifier constructions were regarded as two separate constructions, one would expect to find complex combinations of the two of them, as in the hypothetical [[[Numeral Noun(Multiple)] Noun(Classifier)] Noun(Head)]. But, apparently, such combinations do not occur.

In showing a way in which numeral classifier constructions in general can arise, no explanation is given for the specific differences between classifier languages and non-classifier languages. That is, it is not explained, for example, why English 'chose' not to allow expressions like *seven animals sheep*. Explaining the forms taken by individual languages is a different task from explaining the forms that languages in general tend to take (and is, I assume, practically impossible). It must be presumed that certain background factors relevant to the evolutionary cycle of invention and reanalysis were present in the histories of classifier languages, but absent from the histories of non-classifier languages. (For further recent discussion of numeral classifiers, see Allan, 1977; Craig, 1986; Killingley, 1982, 1983.)

5.5 Adjective-Numeral Word Order

Having argued that numerals fulfil a basically adjectival role, it is necessary to explain certain differences between numerals and other adjectives. The most obvious difference, already mentioned in a quotation from Benacerraf (Chapter 4, Section 2) involves constraints on the relative ordering of numerals and adjectives. *Three French hens* is OK, but **French three hens* is bad.

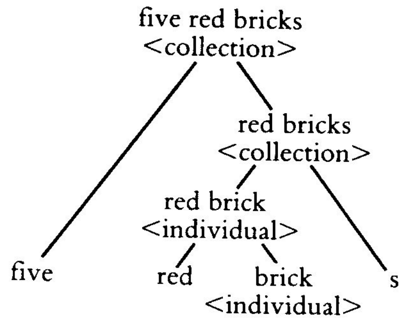
Modifying a singular noun with an adjective usually takes its denotation from a set of individual objects to another set of individual objects, a subset of the first. A book is an individual object; a big red book is also an individual object. Pluralizing a noun takes its denotation from a set of individual objects to a set of collections of such objects. This is also true of nominals composed of a noun and modifying adjectives (not marked as

plural). That is, in languages like English, where adjectives show no number concord, the pluralization is semantically clearly 'outside' the adjectival modification. Both *books* and *big red books* denote a set of collections. Modifying a plural nominal expression with a numeral takes its denotation from a set of collections to another set of collections, a subset of the first. Both *books* and *three books* denote sets of collections. Typical adjectives indicate properties of individual objects, while numerals indicate properties of collections.

> Adjectives like *red, big, built in* 1900, etc. are not plural adjectives but are, rather, adjectives which are basically singular. They are expressions of properties of single objects. The application of a singular adjective to a basic or complex singular noun yields a new complex singular noun to which the plural operator can then be applied. That means, if singular adjectives occur with plural nouns, they are not applied to the plural noun but to the singular noun before the plural operator is applied. (Bartsch, 1973, p. 57)

Thus the compositional semantic structure of *five red bricks* is as in (5.5.1).

5.5.1



[Terminology: I use 'individual' for any object which is not a collection. Recall (5.3.1) that all collections are objects, but not conversely.]

Sapir (1949, pp. 103–7) gives an interesting discussion of the range of grammatical treatments accorded to plurality in languages, especially in relation to other nominal modifiers. But Sapir seems not to have wished to tie plurality too tightly into a universal ontological scheme in terms of categories such as 'individual object' and 'collection'. In keeping with the spirit of much of his work, he attempted to discern language-particular ontologies through the syntactic patterns of particular languages. Thus plurality is classifiable as a more, or a less, 'material' concept, depending on the language, according to Sapir. But the fact that a concept of plurality is apparently identifiable *across* languages means that, even for Sapir, some kind of universal category exists, presumably to be identified with the notion of 'collection'.

Both structure (5.5.1) and the cases discussed by Sapir illustrate the lack of a simple parallelism between the morphosyntactic structures of plural nominals and their semantics. This kind of lack of parallelism between semantic and morphosyntactic structure is known through the more familiar example of the English possessive *-'s*, as in '*the man I saw yesterday's*, where the meaning shows that the [-z] is in construction with the entire preceding phrase' (Bloomfield, 1933, pp. 178–9). An equally pervasive, though less familiar, example is that of tense, often morphosyntactically realized as an affix on a verb, but semantically plausibly regarded as modifying a whole sentence.

But in other cases involving nominal modifiers, languages behave in a much more uniform way, a way which can in fact be explained by reference to a compositional semantic structure as in (5.5.1). Such a structure provides a direct and simple explanation for a universally attested syntactic relationship between numerals and adjectives.

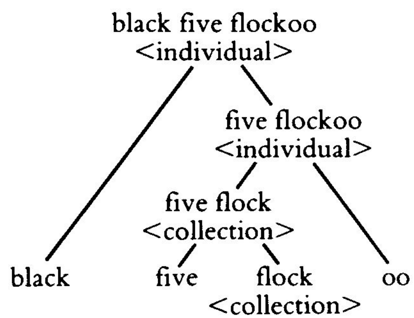
> The order within the noun phrase is subject to powerful constraints. When any or all of the three types of qualifiers precede the noun, the order among them is always the same: demonstrative, numeral, and adjective, as in English. 'these five houses' [*sic*].
>
> When any or all follow, the favorite order is the exact opposite: noun, adjective, numeral, demonstrative. A less popular alternative is the same order as that just given for the instances in which these elements precede the noun. An example of the latter is Kikuyu, a Bantu language of East Africa, with the order, 'houses these five large', instead of the more popular 'houses large five these'. (Greenberg, 1963, p. 87)

Numerals only modify collection-denoting (that is, plural) expressions. And adjectives unmarked for plural typically only modify individual-object-denoting (that is, singular) expressions. Pluralization is the only linguistic operation relating the two types of expression (that is, singulars and plurals). It follows, if semantic interpretation is compositional, and reflected in syntactic structure, that the position of a numeral must be outside that of an adjective unmarked for plurality. Put simply, **red five houses* is ill-formed because a collection cannot be red; but *five red houses* is OK, because individual objects, such as houses, can be red. Exceptional cases, such as the Kikuyu mentioned by Greenberg, present a problem. Perhaps they can be explained by appeal to some kind of topicalization mechanism deriving *houses these five large* from *these five large houses*. Or possibly an explanation can be given in terms of plural number agreement on adjectives, as for some Latin and Arabic examples, to be mentioned presently.

The explanation just given for the Greenbergian universal depends crucially on plural being the morphologically marked category, as opposed to singular. The semantic compositional structure in (5.5.1) suggests that adjectives assumed to be basically singular combine with singular nouns before pluralization of the whole resulting nominal. In fact, English adjectives are not marked for singular or plural, so *black*, for example, does not appear overtly to be either one or the other. On semantic grounds, I assume that most English adjectives are basically singular, in that they denote sets of individuals, rather than sets of collections. (An exception would be *numerous*, as in *the exceptions are numerous* as opposed to the ungrammatical **the exception is numerous*. The data are complicated in that there is often a valid inference from the fact that all the individuals in a collection have some property to the fact that the collection itself has that property.) The compositional semantic hypothesis suggested here would allow expressions of the form [Numeral [[[Adjective + singular] Noun] plural]], in which an overtly singular-marked adjective precedes a noun and a plural marker. I do not know of languages which permit such constructions, but the fact that they perhaps do not occur does not shake the compositional semantic hypothesis put forward here. Universal patterns of syntactic agreement, as described by Corbett (1983), make attributive modifiers the most likely of all elements to agree syntactically with their heads, and thus make such constructions unlikely on other grounds.

Singular, rather than plural, is universally the morphologically unmarked category presumably because individual objects tend to be more salient in human perception than collections. But imagine, if possible, a creature (say a Martian) for whom collections loomed into consciousness before individual objects. Such a creature might have in its language a 'singularization' process, such that, say, *flock* meant a collection of sheep and *flockoo* an individual sheep, *herd* a collection of cattle and *herdoo* an individual cow or bull, and so on. If in this language there were only the singularization processes indicated by suffixing *-oo* (that is, no pluralization), and, as in English, numerals could only modify collection-denoting expressions, and adjectives could only modify individual-object-denoting expressions, we would expect to find expressions such as *black five flockoo*, with compositional semantic structure as in (5.5.2).

5.5.2



This would be translated into English as the relatively circumlocutionary *black member of a collection of five sheep*. The main effect of this circumlocution is to manoeuvre the adjective to a position outside the numeral.

The Martian would presumably have to be able to determine the cardinality of collections other than by counting their members, say by measuring or weighing the collections. But this would only be possible if the Martian environment provided individual objects in neatly quantized sizes. The inevitable increasing bizarreness of the hypothetical Martian example, when elaborated consistently, points up, on the human side, the inevitable naturalness of the observed human pattern of numerals outside adjectives, given human perception and the human environment.

[Actually, Egyptian Arabic goes a small way in the direction of the hypothetical Martian pattern, with what Mitchell (1962, p. 42) calls the 'singulative', used to refer to individuals singled out from masses, for example, *xoox* 'peaches as a mass', *xooxa* 'one peach', *basal* 'onions as a mass', *basala* or *basalaaya* 'one onion'. But this dialect cannot modify the basic mass terms (for

example, *xoox*) with a numeral (in the range 3-10) and resorts to what might be called 'repluralization' to indicate the cardinality of collections of such individuals (for example, *talat xooxaat* 'three peaches'. And it is as awkward to refer in Egyptian Arabic to a ripe member of a collection of 100 peaches as it is in English.]

The occurrence of adjectives morphologically marked for plurality (in concord with the head noun) complicates the hypothesis that Numeral-Adjective-Noun (or the reverse) word order is explainable by a straightforward appeal to compositional denotational semantics. In fact, the hypothesis predicts the possibility of numerals ordered *inside* adjectives just when these adjectives are plural. I will explain with an example.

In Latin *homo* ('man') is singular, *homines* ('men') is plural. The former denotes the set of men, the latter denotes the set of collections of men. It would seem reasonable to make parallel statements about the adjectives *sapiens* ['wise' (singular)] and *sapientes* ['wise' (plural)]. Thus *sapiens* would denote the set of wise objects, and *sapientes* would denote the set of collections of wise objects. In fact, *sapientes* can be used as a noun, as in *tres sapientes* ('three wise ones'). In English the numeral cannot semantically combine directly with an adjective, because most adjectives are inherently singular in English, denoting sets of objects, rather than sets of collections. Thus there cannot be any non-null intersection of the denotation of a numeral and the denotation of an adjective. But with a plural adjective, such as Latin *sapientes*, the denotations of numeral and adjective can combine directly, as in *tres sapientes*. Since *homines* denotes another set of collections, numeral, adjective, and noun are in principle combinable in any order, with semantic interpretation producing the intersection of the three sets concerned. And in fact, of the six logical possibilities, all are acceptable, although some would (according to a Classicist informant) only be used to give special emphasis to the final word in the phrase. The possibilities my Classicist informant was least happy about were the third and the fifth below, interestingly just the ones that violate Greenberg's universal.

5.5.3 Tres sapientes homines
Tres homines sapientes
Sapientes tres homines
Sapientes homines tres
Homines tres sapientes
Homines sapientes tres

Greenberg's universal 19 (above), locating numerals outside qualifying adjectives, relates only to what he calls the 'dominant order' in languages. In other languages beside Latin, it is possible to have an alternative word order, as in the following Arabic example.

5.5.4 arrijjaal aθθalaaθ almuʕamiriin
the men the three the old plural
the three old men

(Adjectives in Arabic agree in definiteness, which accounts for the repetition of 'the' in the gloss here.) Note that in this example also, where the numeral intervenes between noun and adjective, there is a plural marker on the adjective. I have not examined the correlation between counter-instances to Greenberg's universal and plural agreement of adjectives, but the Latin and Arabic examples suggest that it might be just in cases where adjectives agree in plurality with their head nouns that a freer word order is possible. The hypothesis linking Numeral-Adjective-Noun word order with compositional semantic interpretation would predict such a correlation.

These remarks on numeral-adjective word order only apply, of course, to cardinal numerals, which modify collection-denoting expressions. Ordinal numerals, which modify individual-object-denoting expressions, may be freely ordered with respect to adjectives, given a suitable context. Thus one may find both *the fifth fat boy* and *the fat fifth boy*, although situations in which the latter would be appropriate are less likely to occur.

In expressions such as *The Magnificent Seven*, *The Famous Five*, *The Top Twenty*, and so on, the adjective actually precedes the numeral (the numeral used as a noun, that is), unlike the cases discussed above. But this is possible just because in examples like *The Magnificent Seven*, the referent of the numeral (noun) is a collection *viewed as an object*. In the movie, it was the band of seven considered as a unit that was magnificent, though some of its members fell short of individual magnificence. *The magnificent seven* is not an exact paraphrase of *the seven magnificent ones*.

The point is perhaps clearer with *the top twenty*. Taken individually only one of the top twenty is strictly top; the other nineteen of the top twenty are not strictly top, that is, they are second, third, and so on. So *the top twenty* is not used to refer to a collection of individuals which, as individuals, are all top, unlike the *five red bricks*, which is used to refer to a collection of individuals which, as individuals, are all red. But even in a relatively clear case, such as *the top twenty*, because the question of whether something is viewed as an object or as a collection is subjective, the data in this area tend to be subtle, and speakers will not always go along with this strict analysis, and will allow themselves a kind of 'doublethink' according to which something may appear as both an object and a collection simultaneously. The issue is made more complex by differences between predicates, such that some are 'distributive', allowing a valid inference from a property of a collection to a property of each individual member. *Sleeping* is distributive in this sense, so *the sleeping twenty* is in fact a paraphrase of *the twenty sleeping ones*.

I have discussed the ordering of numerals and descriptive adjectives in relation to each other. There is another generalization concerning the ordering of numerals and adjectives, relative to their head noun.

> When the descriptive adjective precedes, then the demonstratives and numerals virtually always precede the noun likewise. ...

	NA	AN
Num-Noun	8	10
Noun-Num	11	0

(Greenberg, 1963, p. 86)

Thus, in Greenberg's sample, in languages where the adjective follows the noun the numeral follows or precedes the noun with roughly equal frequency. But in all languages (in this sample) where the adjective precedes the noun, the numeral does so too. Looking at a wider sample, the facts are more complex, and involve the question of whether a language has prepositions or postpositions. Prior (1985), working with a different and larger sample chosen to represent the world's languages as systematically as possible, gives the following table for prepositional languages.

(Pr)	NumN	either	NNum	Total available
NA	10	3	10	23
AN	9	–	2	11

(1985, p. 283)

This is roughly in keeping with Greenberg's table above, though it shows a somewhat less clear-cut situation. For postpositional languages, Prior gives:

(Po)	NumN	either	NNum	Total available
NA	1	1	16	18
AN	10	4	-	14

(1985, p. 284)

In these languages numerals tend strongly to go on the same side of the noun as adjectives. These data, particularly the involvement of the pre/postpositional parameter, are quite perplexing. I offer no explanations for them.

5.6 Addition and Conjunction

It is a very familiar fact that languages tend strongly to use the same word to indicate addition in numeral constructions as they use to indicate logical conjunction. Examples include English *and*, French *et*, Diyari *ya* (Austin 1981, pp. 56–7), and Arabic *wa*. Other words are found sporadically, such as Welsh *ar*, a preposition meaning 'on', but the commonest indicators of addition, across languages, are words which also happen to indicate the logical conjunction of propositions.

From the point of view of work in the foundations of mathematics and logic, the familiarity and apparent naturalness of this fact is quite puzzling. No version of logicism, which attempts to derive arithmetic concepts – such as addition – from logical primitives – such as sets, or classes, or ordered pairs – gives any hint of a connection between arithmetical addition and logical conjunction. Indeed, if a relationship between addition and any logical connective is suggested, it is between addition and *disjunction*, which, interestingly, is sometimes called 'logical sum' (for example, in Whitehead and Russell 1962, p. 93). Wall (1972, p. 186) links addition with set-theoretic union. The standard links between propositional connectives and the operators of set theory are:

5.6.1 Conjunction – intersection
Disjunction – union

On the basis of formal properties, addition is no more closely linked with intersection than with union. Addition, multiplication, intersection, and union are all both commutative and associative. The Distributive Law which applies to addition and multiplication is paralleled by two theorems of propositional logic:

5.6.2
$$(x + y) \star z = (x \star z) + (y \star z)$$
$$(p \,\&\, q) \vee r \longleftrightarrow (p \vee r) \,\&\, (q \vee r)$$
$$(p \vee q) \,\&\, r \longleftrightarrow (p \,\&\, r) \vee (q \,\&\, r)$$

These parallelisms fail to reveal any *special* connection between addition and conjunction.

On the other hand, it must be said that there are instances of English *and* (and corresponding words in other languages) which cannot be easily accounted for in terms of the logical conjunction of propositions.

5.6.3 John and Mary are a happy couple.

This example is well known, and is the first of several which Partee and Rooth (1983, p. 361) mention as illustrating 'special' rather than 'central' uses of *and*. Their distinction between central and special uses of *and* allows Partee and Rooth to 'forestall a quick negative answer' to the question of 'whether we can give a single meaning for *and*'. Keenan and Faltz (1985) unify the treatment of conjunctions of sentences with that of conjunctions of phrases by means of 'homomorphisms' between, for example, *John is sleeping and Mary is sleeping* and *John and Mary are sleeping*. But examples like (5.6.3) elude their treatment also, and Keenan and Faltz admit a class of 'non-homomorphic predicates' such as *be two teachers*, *be a happy couple*, and *love each other*. Like Partee and Rooth, Keenan and Faltz note the need to 'distinguish two *and*'s in English, the lower order *and* we have been using which forms intersections [corresponding to the connective of propositional calculus] and a higher order one which (roughly) forms sets' (1985, p. 270). Note that it is not only in English, but probably in all languages that this embarrassing need to distinguish two conjoining particles arises.

There is room for hope that an account of *closely related* meanings can be given for *and* and its equivalents in other languages, an account which explains the naturalness of the use of these words to express arithmetical addition. Partee and Rooth refer their readers to an article by Link (1983). Link proposes an account of sentences in which predications are made collectively of groups of individuals. He postulates a logical connective ⊕ defined as follows: 'let *a* and *b* denote two atoms [atomic elements in the universe of discourse . . . *a* ⊕ *b* is the *individual sum* or *plural object* of *a* and *b*' (1983, p. 307). Informally, then, *John* ⊕ *Mary* denotes the plural object constituted by the two individuals John and Mary; this object might perchance also be a happy couple. Link's ⊕ is not the same as the traditional propositional connective '&', but he does propose a logical system in which a number of theorems relate this plural-object-forming operator to the conjunction of propositions. One such theorem is:

T.11 $\bigwedge x \bigwedge y(\star Px \wedge \star Py \rightarrow \star Px \oplus y)$

Here $\star P$ is any plural predicate, formed from a singular predicate *P*. If the denotation of *P* is a set of individuals, the denotation of $\star P$ is the union of that set of individuals with the power set of that set, that is, the set of individuals and sets of such individuals. [Link refers to this set by the technical term 'complete join-subsemilattice', from work in lattice theory by Grätzer (1971).] Theorem T.11 says that if some object *x* satisfies $\star P$ and some object *y* satisfies $\star P$, then the plural object $x \oplus y$ also satisfies $\star P$. This is an inference from a conjunction of propositions to a single predication involving a plural object.

Link's system also allows inferences in the other direction, from a predication involving a plural object to a conjunction of propositions. Link defines a class of *distributive* predicates as follows:

(27) $Distr(P) \longleftrightarrow X(Px \rightarrow Atx)$

(here the predicate *At* stands for the property of being an atom in the model). To illustrate take the intuitively valid inference from a) to b) in (28).

(28) (a) John, Paul, George, and Ringo are pop stars.
(b) Paul is a pop star.

This inference can be formally represented if we consider *pop star* as a distributive predicate *P* in the sense of (27). In this case the extension of $\star P$ is closed under non-zero i-parts [i.e. every non-

null part of an element in the extension of $\star P$ is also in the extension of $\star P$], so every atom of an i-sum [i.e. of a plural object] which is $\star P$ is itself $\star P$, hence it is a *P*. In symbolic form the inference (28) looks like this.

(29) *(a)* $\star P(a \oplus b \oplus c \oplus d)$
(b) *Distr* (*P*)
(c) $b \ \Pi \ a \oplus b \oplus c \oplus d$
[i.e. *b* is an individual part of the plural object]
(d) $\star Pb$
(e) Pb

(1983, p. 309)

Clearly, Link's logical machinery plus the standard machinery of propositional logic will allow inferences from single predications involving plural objects to conjunctions of propositions, inferences as in (5.6.4).

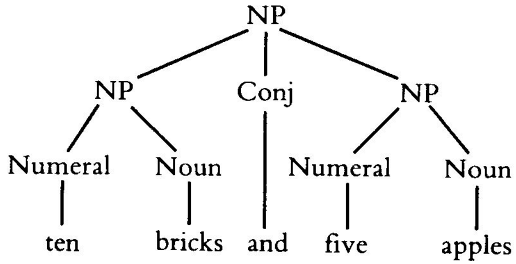
5.6.4 John and Mary are happy → John is happy and Mary is happy

This kind of inference is, of course, only valid for the distributive class of predicates, so we will not get the invalid inference '*John is a happy couple*' from '*John and Mary are a happy couple*', as *be a happy couple* is not distributive. But the great majority of predicates in languages are in fact distributive in this sense. Given this and the *general* validity of the inference from conjoined propositions to predications involving plural objects, as in Link's theorem T.11 above, it is not surprising that languages should choose to use the same word (for example, English *and*) to express both the logical conjunction of propositions and the operator which forms an expression for a plural object out of other, singular or plural, names.

The *and* used to form expressions for plural objects seems, intuitively, closely related to the *and* used to express arithmetical addition, as in *one and one make two*. Furthermore, Link's 'plural objects' or 'i-sums' correspond closely to my 'collections'. I will investigate whether the semantics of collections, or plural objects, can be used to show the naturalness of the use of words such as *and* to form expressions for higher-valued numbers out of the words available in a simple numeral lexicon. That is, given a numeral lexicon (for example, *one*, *two*, . . . , *nine*, *ten*), I will try to answer the question of why it seems natural for expressions such as *eight and three*, *two and nine*, *ten and one*, and so on, to have the meanings they do, that is, to express the number 11.

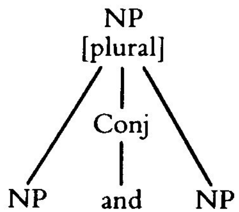
The denotation of the NP *ten bricks* is the set of all collections of ten bricks. Consider now a conjunction of NPs with numerals, for example, *ten bricks and five apples*, with a structure as in (5.6.5).

5.6.5



The denotation of this expression is the set of all collections consisting of ten bricks and five apples. A rule assigning this denotation systematically is (5.6.6).

5.6.6 The denotation of an NP of the form



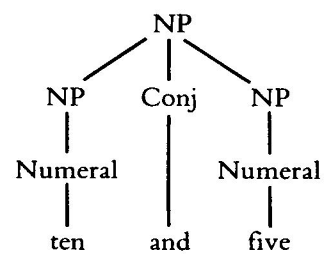
is the set of all collections which are the union of a collection in the denotation of one constituent NP with a collection in the denotation of the other constituent NP.

I adapt set-theoretic terminology, for example, 'union', to apply to collections, although collections as envisaged here do not behave exactly like sets in classical set theory. A significant difference between my use of 'union' and the set-theoretic use is that I permit the union of individual objects with collections or other individual objects. This will be discussed in more detail below. As an informal example, the union of the collection consisting of all my children and the collection consisting of all my neighbour's children is another collection, consisting of just those individuals who are either a child of mine or a child of my neighbour's. Or if my neighbour and I have just one child apiece, the union of my individual child and her individual child is the collection of our two children. In the case of disjoint collections, for example, of bricks and of apples, or (as it happens) my children and my neighbour's children, using the ordinary union operation in rule (5.6.6) obviously gives the desired result.

Now take a case such as *I saw ten bricks and five bricks*. This sentence can be interpreted in such a way as to entail a conjunction of propositions, as in *I saw ten bricks and (then) I saw five bricks*. In this case the five bricks may or may not include some of the same bricks as the ten bricks. Consider the case where ten bricks are shown and then, while the observer turns away momentarily, five bricks are removed so that when he turns back, he sees five of the original ten bricks. In this case, although there might be said to be fifteen distinct acts of seeing an individual brick (some involving the same brick), there are less than fifteen bricks. In fact the number of bricks involved may be any number from ten to fifteen. Rule (5.6.6) accounts for this correctly, as the collections which are the union of a collection of ten and a collection of five may have any cardinality between ten and fifteen, due to the possible non-disjointness of the collections. Perhaps cases where non-disjoint collections are involved are unusual or statistically exceptional in actual practice, so that the most normal interpretation of *ten bricks and five bricks* is a collection of just fifteen bricks, but nevertheless the possibility of less than fifeen being involved definitely exists.

In Chapter 4, Section 3, a rule for assigning a denotation to bare numerals was given (4.3.4). This rule operated on NP structures dominating just the category 'Numeral'. Given the syntactic co-ordination of NPs by *and*, the means now exists to assign denotations to bare numerals co-ordinated with *and*. The denotation of the structure in (5.6.7), for example, will be the set of collections which are the union of a collection of ten things with a collection of five things.

5.6.7



This could occur in the elliptical *I saw ten and five*, and could be used in the same situation as the one described above, where five bricks were removed from an original collection of ten. Thus, here again, the number of objects involved is not necessarily exactly fifteen, but could be any number between ten and fifteen.

I claim that syntactically complex numeral expressions involving addition naturally arise as NP co-ordinations of NPs dominating bare single-word numerals. These co-ordinations use whatever connective the language concerned uses to form expressions for plural objects by conjoining names for individuals, for example, English *and*, Arabic *wa*, and so on. In this respect, Fijian seems to present a problematic counter-example. 'Numbers are combined using the conjunction *'a* 'and', which is used to link clauses and phrases (and *not* using *'ei* 'together with' which combines nouns, names and pronouns); this is one piece of syntactic evidence linking numbers with verbs'. (Dixon, forthcoming). In the light of such facts, it must be admitted that additive constructions do not always arise from conjunctions of NPs, although this may well be their most typical evolutionary source. The Fijian numerals are not fully integrated into the language as nominal modifiers like adjectives; Dixon has to postulate a special *ad hoc* rule accounting for the apparently idiosyncratic positioning of numerals functioning as nominal modifiers. It would be interesting to investigate cases like Fijian further, rare though they are.

It remains to explain how such syntactically complex numerals come to occupy the range of syntactic environments occupied by single-word numerals and how they take on *exact* numerical values, for example how an expression such as *ten and five* can come to act as a nominal modifier, as in an expression such as, *ten and five bricks* and to denote the set of all collections of exactly fifteen objects. I suggest that this would naturally come about by children reanalysing the language of their parents, assuming, in effect, that syntactically complex numerals have the same linguistic properties as the single-word numerals first acquired.

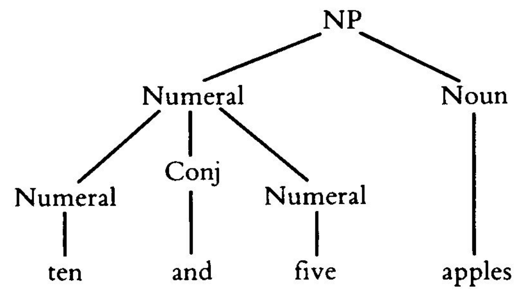
Given the optional ellipsis of head nouns, with the resulting domination of bare numeral words by NP, and the co-ordination of NPs, the following kinds of pattern will be found.

5.6.8 (a) Give me ten apples
(b) Give me ten
(c) Give me ten apples and five apples

(d) Give me ten apples and five
(e) Give me ten and five apples
(f) Give me ten and five

Exposed to data such as these, a child might reasonably hypothesize that *ten and five*, as in (e) and (f), is a phrase occurring in the same environment as the single word *ten*, as in examples (a) and (b). For examples such as (c) and (d), the hypothesis would be that this phrase can be interrupted by the modified Noun. A language-acquirer given such examples could reasonably hypothesize a structure such as (5.6.9) for example (e) above.

5.6.9



Such a structure would be innovative, but compatible with the surface data, as the actual string of words is not new. The conjunction of two Numeral constituents would in fact be a generalization of the rule for conjunction. Languages generally permit conjunction of like constituents, of almost any category. In this way, syntactically complex numerals involving addition could arise and become integrated into the structure of NPs. A child who knows the meanings of the single-word numerals knows each numeral word as a predicate satisfiable by any collection from a set of collections which can be put in a one-to-one correspondence with each other (that is, she knows the numeral words as denoting exact numbers). It would be reasonable and natural to generalize this fact about the interpretation of numerals to the case of syntactically complex numerals as well, so that complex numerals are interpreted as exact numbers.

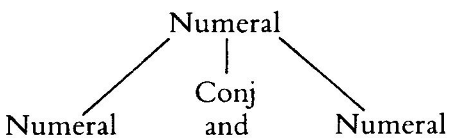
Beside the analogy with single-word numerals, there may be some pragmatic pressure for relatively low-valued numeral expressions to be interpreted with exact values. Compare two possible sets of mappings from expressions to meanings.

5.6.10 A *ten and five* means 15.
roughly ten and five means 10–15.
versus
B *exactly ten and five* means 15.
ten and five means 10–15.

In a society in which situations arose in which it was desirable to specify an exact number more frequently than an approximate number, we would expect the mapping A to be adopted; in such a case, explicit inclusion of a word such as *roughly* would be required for the less typical uses. If, on the other hand, it was more usual to wish to specify approximate numbers, mapping B would tend to be adopted. In fact, higher-valued numerals tend to have approximate interpretations more than lower-valued ones. Compare *There are ten people in this room* with *There are a billion people in China*, where the former but certainly not the latter would tend to have an exact interpretation.

Semantic interpretation of a co-ordinate Numeral as an exact number can be accounted for by rule (5.6.11), which is straightforwardly parallel with rule (5.6.6) for interpreting coordinate NPs.

5.6.11 The denotation of a Numeral of the form



is the set of all collections which are the natural union of a collection in the denotation of one constituent Numeral with a collection in the denotation of the other constituent Numeral.

The 'natural union' operation used in this rule requires comment. I define it as follows:

5.6.12 The natural union of A and B, where
A and B are disjoint collections, or
A and B are distinct individual objects, or
one of A and B is a collection and the other is an individual object which is not a member of it,
is the collection resulting from placing A with B.

There is no natural union of non-disjoint collections, or of objects with themselves, or of collections with any of their members.

The natural union operation corresponds exactly to a familiar physical operation, that of putting one collection or object with another. There is some psychological difficulty in conceiving of partially overlapping collections. The prototypical aggregate constructible as a collection is a spatio-temporally located grouping of physical objects with some kind of boundary of empty space round them. Given an array of objects on a surface, one can discern separate collections in different ways according to how wide a stretch of empty space one accepts as defining a collection boundary. If one sets the definition of collection boundary as a rather narrow space, there will be rather many collections; if one sets the definition as a rather wide space, there will be fewer collections. In this way it is possible to perceive subcollections included within collections. But the physical operation of adding a subcollection to a collection of which it is already a part is actually inconceivable. 'Add the pencils on the table to the objects on the table', construed as an instruction to carry out some physical operation, is impossible to obey literally, just like 'Close the door', if the door is already closed. And a conception of *partially overlapping* collections does not emerge at all naturally from such a scenario. Note that the traditional constellations of stars in the sky do not overlap with or include other constellations. Note also that the physical operation of adding a single object to a collection, which was appealed to in connection with the development of the basic sequence of numeral words [Chapter 3, (j) of (3.6.1.)] and used in the semantic interpretation rule (4.3.3) was assumed to involve an object not already in the collection; the naturalness of this assumed detail was taken to be unproblematic. [Wall (1972, p. 186) defines an operation on pairs of integers, which he calls 'cardinal addition', and defines in terms of the (ordinary) union of *disjoint* sets; he symbolizes this operation with ⊕, the same symbol as Link uses to designate the i-sum of objects.]

Expressions such as, *ten and five apples* arise, I claim, from ellipsis affecting expressions such as, *ten apples and five apples*. If the second noun were elided, we would have *ten apples and five*. In fact, many languages permit nouns to 'interrupt' conjoined numeral expressions exactly as in these last two strings. We have it in the archaic English *three score years and ten*. Biblical Welsh is especially fond of this construction: the following examples are from Hurford (1975).

5.6.13

saith	mlynedd	ac	wyth	gan	mlynedd
7	years	and	8	100	years
807	years				

bum	mlynedd	a	chan	mlynedd
5	years	and	100	years
105	years			

gant	a	phedwar	cufydd	a	deugain
100	and	4	cubits	and	2 20
144	cubits				

deng	wr	a	deugain	a	dau	cant
10	men	and	2 20	and	2	100
250	men					

In Kalabari (a dialect of Ijo) 'additive phrases containing *na* "and" require two occurrences of the non-numeral noun' (Jenewari, 1980, p. 77). Jenewari gives the following examples (my glosses, his translations).

5.6.14

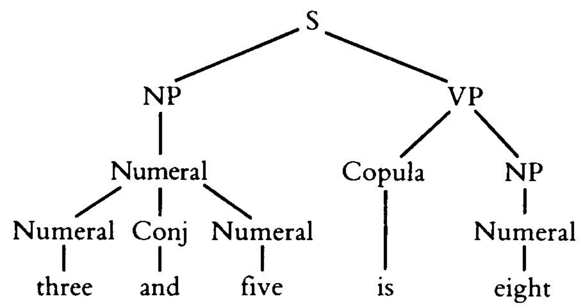
naira	a-tesi	oyi	finji	na	gboru	naira	na
naira	60	10	extra	and	1	naira	plus
seventy one naira							

naira	a-sona	si	na	tira	naira	na
naira	5	20	and	3	naira	plus
one hundred and three naira						

(A naira is a unit of currency.) The Arabic for *a thousand and one nights* is *ʔalf leel wa leela*, literally *thousand night and night*(emphatic singular), a slightly different, but related construction. Such conjoined constructions with a noun apparently interrupting a complex numeral are synonymous with constructions where the noun does not appear to wander into the middle of the numeral. Such structures seem very natural if conjoined numerals are taken to arise from (optional) ellipsis in conjoined NPs.

The syntactic and semantic machinery now given is adequate for the generation and interpretation of sentences expressing simple arithmetical truths, such as *Three and five is eight*. This sentence would have a structure as in (5.6.15).

5.6.15



Assuming that the semantics of such equative sentences stipulates that they are true if and only if the denotation of the subject is identical to the denotation of the NP after the copula, and that the denotation of an NP dominating just a numeral is the same as the denotation of the numeral itself, such a sentence will be assigned the value 'TRUE', as required. Here the denotation of the subject NP is the set of all collections which are the natural union of a collection of three objects and a collection of five objects, that is, the set of all collections of eight objects; and this is also the denotation of the NP after the copula. An arithmetically false sentence such as *Three and five is nine* will not be assigned the value 'TRUE'.

The account given shows a natural way in which syntactically complex numerals, expressing addition, could arise diachronically. It accounts for the naturalness of the coincidence of morphemes expressing addition and those expressing the formation of plural objects. It is not claimed that this is the *only* way in which complex numerals could arise; the means for expressing addition by complex numerals may have been invented via a somewhat different route in the history of some languages. In some languages addition is signalled by an allative preposition meaning 'on' or 'onto' (for example, Welsh *ar*), or by a comitative preposition meaning 'with'. In some languages the conjunction meaning 'and' is itself derived from such a comitative preposition. Such facts reinforce the view that constructions expressing arithmetical addition evolve out of constructions expressing the physical juxtaposition of objects or collections to form (larger) collections. The history of invention is not uniform. And in many languages the traces of this evolutionary path have been erased by subsequent historical change, so that an overt morpheme corresponding to *and* is no longer apparent, as for instance in the English numerals up to 99.

The passage of time tends to make complex numeral constructions more opaque. The overt connective (for example, *and*) expressing addition may get lost, the forms of the constituent numeral words may change, perhaps beyond easy recognition, and the possibility of using complex numerals as nominal modifiers to indicate the composition of collections by subcollections is also lost. Thus, where an expression such as *five and ten bricks* may originally have been able to convey that the collection in question was composed of two subcollections, one of five bricks and one of ten bricks, modern English *fifteen bricks* makes no comment on the composition of the collection. Of course, it is still possible in modern English to convey the composition of collections, but not within the structure of a complex numeral itself, so that we have to say, for example, *five bricks and ten (other) bricks*.

The account given predicts that complex numerals expressing addition may be formed out of any combination of the basic numeral words in a language. That is, not only will something like *ten and five* be available, but also expressions such as, *one and one*, *two and one*, *one and two*, *five and three*, and so on. The reasons for the normal non-occurrence of expressions such as these in mature numeral systems will be given in the next chapter, where the notion of the standardization will be developed.

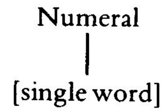
6

Standardization of Complex Numerals to a Fixed Base

6.1 Summary of Constructions So Far

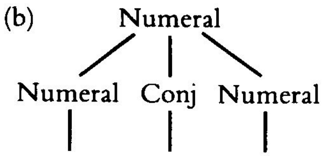
The picture so far developed includes the following numeral constructions.

6.1.1 (a)



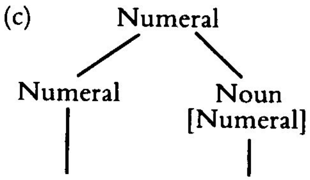
(e.g. five)

(b)



(e.g. five and seven, expressing addition)

(c)



(e.g. five sevens, expressing multiplication)

The notation 'Noun[Numeral]' in the last structure expresses the fact that in multiplicative constructions, the Noun has to be a 'Numeral Noun', i.e. one formed from a numeral. The structures in (6.1.1) would be generated by phrase structure rules as in (6.1.2).

6.1.2 Numeral → [numeral word, e.g. *one*, . . . , *five*, . . . , *ten*]
Numeral → Numeral Conj Numeral
Numeral → Numeral Noun[Numeral]

The naturalness of the use of recursion to embed one numeral construction inside another (inside another (inside another, and so on)) will be discussed in section 6.3 in connection with an experiment on children learning the English numeral system. The rest of this chapter will discuss the development of tight restrictions on the output of recursive production rules as in (6.1.2), under the topic of standardization mechanisms.

The rules in (6.1.2) will generate numeral expressions as in the following sample.

6.1.3 ADDITIVE
one and one, nine and one, one and nine, nine and nine, ten and one, ten and nine, ten and ten

MULTIPLICATIVE
two ones, nine ones, nine nines, ten nines, nine tens

COMPLEX ADDITIVE AND MULTIPLICATIVE
two ones and one, nine tens and nine, one and ten nines, seven eights and four sixes,
nine [one and one]s, three [five and four]s

These examples, which are grammatical in English, although not habitually used for counting, may be thought of as schematic patterns for the formation of complex numerals in languages generally. Routes by which the most frequently found types of additive and multiplicative constructions may emerge historically were described in the previous chapter. Once invented and adopted, particular expressions representing these constructions may be rote-learnt by successive generations. The once transparent analogy with other non-numeral constructions may be lost through the normal historical wear and tear affecting lexicalized forms. Thus numeral phrases may be collapsed into single words, morphemes may be phonologically modified beyond easy recognition, conjoining particles may be dropped, and so on. But usually, the constituent numeral morphemes can be assigned numerical values in a synchronic analysis on the basis of the meaning of the whole expression. For example, the English suffix *-ty* clearly means 10.

There are languages whose numeral systems make use of subtraction, – for example Latin, Ainu, Yoruba – but the use of subtraction is clearly unusual compared with the use of addition.

A few languages also make use of division in the sense that some whole numbers are expressed as fractions of others (for example Welsh *hanner cant* 'half hundred', 50), but this is also relatively rare. A very small number of languages make use of an unorthodox arithmetical operation called 'overcounting' by Hurford (1975, pp. 235–39, where it is discussed in more detail). An example of such overcounting from Ch'ol, a Mayan language, is the expression for 45, translated literally as '5 towards [3 20]' or '5 in the 3rd 20'. My concern in this book is to explain the central, most typical characteristics of numeral systems, and I will not attempt to account for constructions involving subtraction, division, and overcounting.

The numbers combined in additive and multiplicational constructions do not necessarily parallel those of familiar decimal-based systems. Thus:

6.1.4 ADDITIVE
mandu mandu [2 2] = 4, mandu mandu kunu [2 2 1] = 5,
parkulu parkulu [3 3] = 6 (Diyari)
parkuna nunara [2 1] = 3 (Ngamini)
parkulu kuna [2 1] = 3 (Yarluyandi)
(all from Austin, 1981, p. 56)

pymtheg [5 10] = 15, un ar bymtheg [1 on 5 10] (Welsh)
pram – 'bγy [5 3] = 8 (Khmer) (Jacob, 1965, p. 144)

MULTIPLICATIVE
deunaw [2 9] = 18 (Welsh)
tri chouech [3 6] = 18 (Breton)

In the account thus far there is nothing predicting the use of any particular number as a base for the formation of syntactically complex expressions for higher numbers. But all developed numeral systems use bases. Sometimes a system will be 'mixed' and use several bases, as with French, which uses both 10 and 20. The more developed systems also use higher bases, typically with values 100, 1000, which are powers of the bottom base number. On the other hand, there is widespread, if sporadic, evidence of deviation from the use of a standard base number, as in the Welsh and Breton expressions for 18, given above. (Apart from these expressions Welsh and Breton are mixed decimal/vigesimal systems.) The following sections discuss possible mechanisms by which particular numeral words become selected as bases for the formation of higher-valued expressions, usually to the exclusion of other numeral words.

6.2 The Packing Strategy – a Universal Constraint on Complex Numerals

The sets of syntactically complex numeral forms which languages adopt follow a definite pattern, so that a single, general, economical, and accurate descriptive account can be given of the synchronic patterns in any language. This account is, like any other non-trivial linguistic universal ever proposed, met by counter-examples, real and apparent, but is so generally successful in capturing a wide range of data that I dare to assume that it is, as a descriptive statement, substantially correct. In this section I present this descriptive statement, originally conceived in Hurford (1975, henceforth *LTN*) within the aims of generative grammar, and called the 'Packing Strategy'. The Packing Strategy is one of the devices developed at greatest length in *LTN*: for a complete exposition with examples from a variety of languages, see the passages indexed under 'Packing Strategy' there. I will give a simplified and less formal account of its workings here. In the next section, I give some space to arguing why this particular linguistic universal cannot be explained by appeal to innate mental structuring of the language acquirer, and, in subsequent sections, offer an explanation in terms of social interaction and the way numeral systems evolve diachronically.

When two numbers are added or multiplied to express a higher number, the resulting construction is usually markedly unbalanced, in the sense that one of the numbers is much greater than the other, and languages tend strongly to maximize this kind of imbalance. Nothing feels more natural than to express 34 as *thirty-four*, but, looking at the facts coldly, we have to explain why, for instance, **twenty fourteen* is also not acceptable. Shunning **septante* for idiosyncratic reasons of its own, Standard French chooses to express 70 as *soixante dix*, rather than as **cinquante vingt* or **quarante trente*. Classical Welsh expresses 16 as *un ar bymtheg* (1 + 15), rather than as **chwech ar ddeg* (6 + 10). And turning to multiplication, English *two thousand* is acceptable but **twenty hundred* is not. In Classical Welsh, 60 is neither **chwech deg* (6 10) nor **pedwar pymtheg* (4 15), but *tri ugain* (3 20).

The principle involved, which seems intuitively natural until one ponders it, could be expressed as 'When forming an expression for a high number, pick the highest-valued expression available as a starting point, and then build on that.' The label 'Packing Strategy' comes from the thought that when packing a trunk with books, it makes sense to pack the large volumes – encyclopaedias, atlases, and so on – first, and then to slip the slimmer volumes in afterwards. The precise working of this principle depends on a prior definition of the possible combinations of constituents. This is done in *LTN* by means of phrase structure rules and a lexicon, and the more technical details will not be repeated here. The question to be posed and answered here can take a fairly non-technical form: why do languages prefer to form numeral expressions by combining constituents whose arithmetical values are maximally far apart, within the constraints defined by the syntax of the system?

To illustrate what is meant by the 'constraints defined by the syntax', consider some of the examples above. In English, it seems fairly transparent that the words *hundred* and *thousand* belong to the same syntactic class (labelled 'M' in *LTN*), and words such as *two*, *nine*, *fifty* do not belong to this class. Thus *three hundred* and *three thousand* are well-formed numerals, but *three two*, *three nine*, and *three fifty* are not. Similarly, longer expressions, though they may be wellformed in isolation, cannot be substituted for words of the M class (*hundred*, *thousand*, *million*). Thus, 9000 could not be expressed as **two four-thousand-five-hundred*. Allowing a plural suffix on the second constituent still gives a non-standard, if still grammatical, expression: *two four-thousand-five-hundreds*. Within the constraints imposed by the syntax, then, both *nine thousand*, and *ninety hundred* are possible expressions for 9000, but only the former, in which the constituents are maximally far apart arithmetically, is a standard numeral expression. Why should this be so?

To give another example, in Classical Welsh, *deg* (10), *pymtheg* (15), and *ugain* (20) seem to belong to the same syntactic class. All three may occur in frames such as *un ar* . . . , *pedwar ar*, . . . , for example *pedwar ar deg* (= 14), *pedwar ar bymtheg* (= 19), *pedwar ar hugain* (= 24). Furthermore, the single-digit expressions, *un*, *dau*, *tri*, . . . , *wyth*, *naw* (1–9), all seem to belong to a class, as they are mutually substitutable in a wide range of environments.

An obvious simple syntactic rule would generate constructions in which single digit words combine freely with *deg*, *pymtheg*, and *ugain*. Some of the forms generated by this rule are well-formed, and some are not. Here the Packing Strategy comes to the rescue, rejecting, for example, *saith ar bymtheg* as not having the greatest syntactically permissible arithmetical imbalance between its constituents. The Packing Strategy, however, is merely a descriptive device, embedded within a generative grammar. It captures, perhaps even elegantly, the fact that some syntactic combinations are well-formed while others are not, but it is not clear that it explains that fact.

Sometimes, in a sense, the syntactic rules of a numeral system may appear to be already shaped by a principle similar to the Packing Strategy. Thus, in English the single digit words *one*, *two*, . . . , *nine* form a class and the word *ten* is *sui generis*. Allowing that one can recognize *ten* in the bound form *-teen*, and putting aside the idiosyncratic *eleven* and *twelve*, a syntactic rule of the form X → DIGIT *te(e)n* will generate *three teen*, *four teen*, . . . , *nine teen*, which, after some necessary phonological modification in some cases, are well-formed numeral expressions. (*Eleven* and *twelve* can be seen as suppletive versions of underlying *one teen* and *two teen*.) Now why, one may ask, is there a syntactic rule singling out the form *te(e)n* for special treatment? Why not also a rule allowing for additive constructions with a base of nine, a rule such as X → DIGIT *nine*? *A priori*, there seems nothing objectionable in expressing, say, 13 as **fournine*, or 14 as **fivenine*. Why is the syntax of the English numeral system this way?

These facts are captured in a generative theory of numerals by a combination, for each language, of a very permissive (that is overgenerating) set of phrase structure rules and a compensating set of deep structure constraints, or filters. The principal such filter, for any language, is the Packing Strategy, which selects just one of the many structures associated with a particular semantic value (a number) as the underlying structure of the numeral correctly expressing that number in that language. (The ultimate status of the Packing Strategy in the argument of this chapter is as a diachronic construct, rather than as the synchronic filter described first here. The aim is to show a device which, despite its apparent descriptive efficiency and neatness within a synchronic generative description, cannot plausibly be interpreted as psychologically real, although it lends itself well to a diachronic interpretation.) I give below a simplified version of the phrase

structure rules for English numerals to illustrate the workings of the synchronic generative account which incorporates the Packing Strategy. (The multiple recursivity and consequent generative power of these rules is justified in detail in *LTN*, pp. 19–28.)

6.2.1

$$\text{NUMBER} \rightarrow \left\{ \begin{array}{c} \text{DIGIT} \\ \text{PHRASE} \quad (\text{NUMBER}) \end{array} \right\}$$

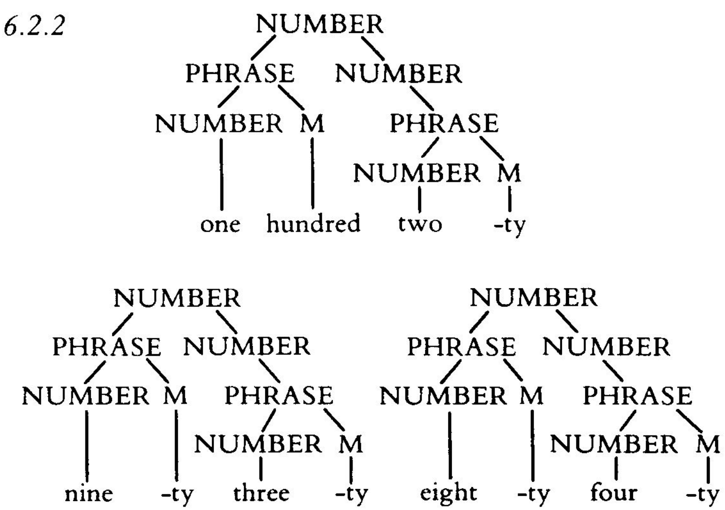
$$\text{PHRASE} \rightarrow \text{NUMBER} \quad \text{M}$$

$$\text{M} \rightarrow \left\{ \begin{array}{l} \text{-ty} \\ \text{hundred} \\ \text{thousand} \\ \text{million} \\ \text{billion} \end{array} \right\}$$

(In this simplified version, DIGIT expands to any of the words *one*, *two*, . . . , *eight*, *nine*; I shall concentrate here only on the clearly regular English numerals, ignoring the morphological vagaries of *ten*, *eleven*, *twelve*, and the *-teen* words, and assuming the obviously necessary idiosyncratic processes converting *two* + *ty* into *twenty*, and so on. Not to connect *ten* with *-teen* would miss generalizations. All such idiosyncrasies are discussed in *LTN*.)

Examples of some structures generated by these rules are as in (6.2.2).

6.2.2



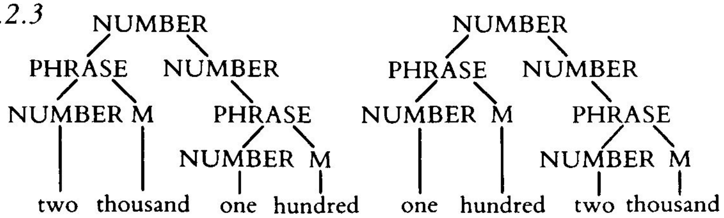
The semantic interpretation of such structures as these proceeds by adding together the values of the immediate constituents of any NUMBER, and by multiplying together the values of the immediate constituents of any PHRASE. (See *LTN*, pp. 28–36 for details.) Interpreting in this way, each of the structures in (6.2.2) has the value 120. But of course only the first of these structures underlies the standard expression for that number in English, *one hundred and twenty*. (*And* is assumed to be inserted by transformation.) The Packing Strategy captures this by filtering out the other structures, which correspond to the non-standardized expressions **ninety and thirty* and **eighty and forty*, along with a host of similar structures also generated by the rules of (6.2.1). The Packing Strategy states, essentially,

> that the sister constituent of a NUMBER must have the highest possible value, that is, the highest value that a constituent of its category can have less than or equal to the value of the immediately dominating node. (*LTN*, pp. 67–8)

Applied to the structures in (6.2.2), the constituent in question is in each case the left-hand PHRASE. This is the sister of a NUMBER. The value of the immediately dominating node is in each case 120, and the highest value that a PHRASE can have (in English) less than or equal to 120 is 100. Consequently the first structure in (6.2.2) is well-formed; the others are, correspondingly, ill-formed, and filtered out by the Packing Strategy.

To give some further examples, English *two thousand one hundred* is standard, but **one hundred two thousand* is not. The Packing Strategy accounts for this as illustrated in (6.2.3).

6.2.3

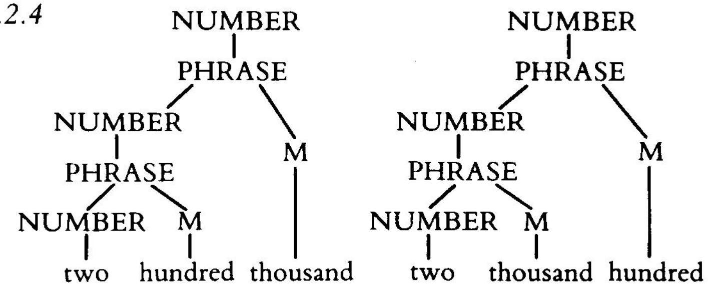


The pretheoretical intuitive account of the facts here would probably take the form 'in additive constructions, put the highest valued constituent first'. The Packing Strategy, which is not sensitive to the order of constituents, expresses it rather as, 'pack

the highest-valued constituent as near as possible to the root (top) of the tree'. (At least, this is an acceptable paraphrase of the Packing Strategy for the purposes of this example.) Consequently the first structure in (6.2.3) gives the wellformed expression for 2100 in English.

In the case of *two hundred thousand* as opposed to **two thousand hundred*, the pretheoretical intuitive account would probably run, 'in multiplicative constructions, put the highest-valued expression second'. But the apparent asymmetry between additive and multiplicative constructions is resolved by the Packing Strategy, which again, in effect, says, 'pack the highest-valued sister constituent of a NUMBER as high as possible in the tree'. This is illustrated in (6.2.4), where the constituent in question is an M.

6.2.4



The Packing Strategy selects the first of these as the underlying form of the English expression for 200,000.

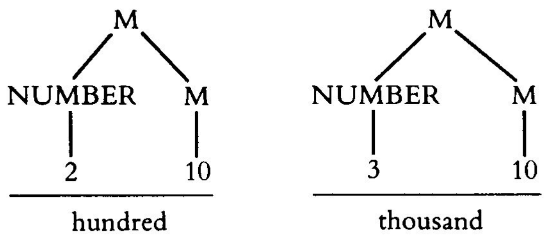
Turning briefly to another language, the ancient Hawaiian numeral system used base-words with values of 10, 20, 40, 400, 4000, 40,000, and 400,000 (for a fuller description and a list of sources, see *LTN*, p. 202). These base-words for 40 and above could appear in multiplicative constructions with a preceding numeral, for example *elua lau* (2 400 = 800), *eha kanaha* (4 40 = 160). But multiplicative constructions corresponding to (11 40 = 440) or (13 400 = 5200), for example, did not occur. If they had, this would have been in violation of the Packing Strategy, which insists on multiplication by the highest-valued base-word available in the language.

The examples given so far deal only with additive and multiplicative constructions. An argument is also made in *LTN* that the Packing Strategy captures a generalization involving the more complex operation of exponentiation (raising to a power).

The words *hundred, thousand, million, billion* and *trillion* are (in British English – in American English the facts are a little different but a parallel case can be made) respectively the 2nd, 3rd, 6th, 12th, and 18th powers of the base number, 10. Note that there are no words for the 4th, 5th, 7th, and 8th powers of 10, and these gaps are quite natural. Similar gaps occur in the higher numeral vocabulary of many languages. The appropriate generalization is that there are words for every successive power of the base number, up to some arbitrary limit (here the 3rd power), and then for each successive power of that, again to some arbitrary limit, and so on. Thus *hundred* and *thousand* express the 2nd and 3rd powers of 10, and (in British English usage, now actually declining in this respect) *billion* and *trillion* the 2nd and 3rd powers of 1,000,000. Every word which expresses a power of some lower number expresses a power of the highest available lower number for which there is a word in the language. The Packing Strategy accounts for this in the following way.

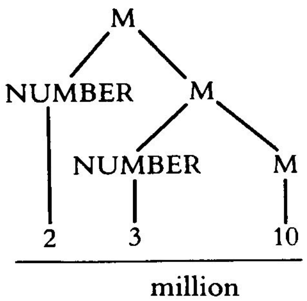
In *LTN* it is proposed that words such as *hundred, thousand,* and so on have a complex underlying structure as in (6.2.5).

6.2.5



These structures are interpreted by exponentiation, that is they correspond to the conventional arithmetic expressions 10^2 and 10^3. The structure for *million* is (6.2.6).

6.2.6



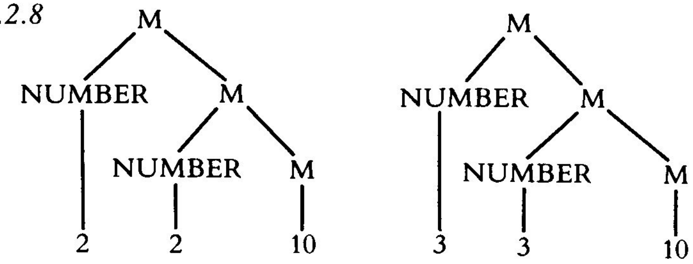
This corresponds to $(10^3)^2$. Structures as in (6.2.5) and (6.2.6) are generated by the rule (6.2.7).

6.2.7

$$M \rightarrow \left\{ \begin{array}{ll} 10 & \\ \text{NUMBER} & M \end{array} \right\}$$

(This highly recursive rule is justified in *LTN*, pp. 26–8). Rule (6.2.7) will generate structures as in (6.2.8), corresponding to the arithmetic expressions $(10^2)^2$ and $(10^3)^3$.

6.2.8



There are no single words in (British) English for the values assigned to these structures, that is for 10,000 and 1,000,000,000 even though there are words for values assigned to the embedded M constructions in each case. This is attributed by the Packing Strategy to the fact that in each case the embedded M construction (in each case the sister constituent of a NUMBER) is not the highest-valued construction of that category made available by the other rules of the system. In the first structure of (6.2.8), the embedded M evaluates to 100, which is lower than 1000, also provided as an M by the system, and hence the whole structure is ill-formed. Similarly in the second structure, the embedded M evaluates to 1000, which is lower than 1,000,000, also provided by the system, and hence the whole structure is ill-formed. This example is parallel to the case of **twenty hundred* which is ill-formed because of the availability of *two thousand*, in which the sister constituent of the NUMBER has the highest available value. The only difference between the cases is that the structures (6.2.5), (6.2.6), and (6.2.8) are more abstract, since they decompose the senses of (putative) words.

Chinese has a series of words with the values 10, 10^2, 10^3, 10^4, $(10^4)^2$, and $(10^4)^3$. Note that there is no word corresponding to English *million*, since this could only have structure corresponding to $(10^3)^2$, or even less plausibly $(10^2)^3$, and in these structures the embedded expression is not the highest one made available by the system, due to the existence of a word for 10^4.

It may provide some further clarification here if I now mention some actual counter-examples to the Packing Strategy. Some varieties of Classical Welsh express 18, quite idiosyncratically, as *deunaw* (2 9). In the same dialects, 17 and 19 are expressed by addition to a base of 15, as *dau ar bymtheg* (2 on 15) and *pedwar ar bymtheg* (4 on 15). Given the existence in the system of a base word with the value 15, expressing 18 as *deunaw* clearly does not use the highest available base word, thus violating the Packing Strategy. A similar Celtic vagary is found in the Breton for 18, *tri-chouech* (Menninger, 1969, p. 97). This also violates the Packing Strategy. Examples such as these are found sporadically in languages. It is usually clear, even pretheoretically, that there is something odd about them, and their existence barely touches the general validity of the Packing Strategy. The Indian English words *lakh* and *crore* and the corresponding source words in the Indian languages, standing for 100,000 and 10,000,000, respectively, are also counter-examples to the Packing Strategy, as these languages also have words for 1,000, but no word for 10,000. Although this counter-example is very widespread, it is clear that it stems from a single historical source, and so can be seen as an isolated case.

Another class of counter-examples to the Packing Strategy involves cases where a language expresses a number in more than one way, as, for instance, English expresses 1100 both as *one thousand one hundred* and as *eleven hundred*. The nature of the Packing Strategy is to select just one expression above all rivals as the way of expressing a particular number in a particular language, and cases such as this are clear counter-examples. But such examples are relatively rare: in all languages the great majority of numbers are standardly expressed as just one out of the vast range of eligible binary arithmetic combinations. (The case of different stylistic variants of the same arithmetic combination, for example *twenty-five* versus *five and twenty*, is a different matter, not affected by the Packing Strategy.)

A final class of counter-examples is not so rare. I quote from Seidenberg:

> Any system in which we find (with minor variations)
> (Type I): $6 = 2 \times 3,\ 7 = 6 + 1,\ 8 = 2 \times 4,\ 9 = 8 + 1$
> or $6 = 2 \times 3,\ 7 = 8 - 1,\ 8 = 2 \times 4,\ 9 - 10 - 1$
> or (Type II): $6 = 3 + 3,\ 7 = 4 + 3,\ 8 = 4 + 4,\ 9 = 5 + 4$
> ... we will refer to as a neo-2 system. (1960, p. 227)

Neo-2 systems, or at least traces of them, are not uncommon. Seidenberg (p. 227) lists a large number. Hymes (1955) uses the term 'pairing' for these systems and mentions 13 Athapaskan languages which exhibit characteristics of such pairing systems. Such systems violate the Packing Strategy in so far as they use a base word with the value 4 in the expressions for 8 (and sometimes also 7), while not making use of this base in the expression for 6, even though it is available and higher in value than the base word actually used in the expression for 6. In partial defence of the Packing Strategy in the face of such examples it can be said that they all involve numbers at the very lowest end of the counting scale, that is below 10, and counter-examples of this sort above 10, like the very eccentric Welsh *deunaw* mentioned above, are quite rare. It was for reasons such as this that the epigraph chosen for *LTN* was Wittgenstein's

> Our language can be seen as an ancient city: a maze of little streets and squares, of old and new houses, and of houses with additions from various periods; and this surrounded by a multitude of new boroughs with straight regular streets and uniform houses. (1972, p. 8e)

The Packing Strategy is a rule accounting well for the straight streets and uniform houses of the new boroughs, while not fitting the maze of little streets and squares in the ancient city quite so well.

Notwithstanding various counter-examples, the Packing Strategy states a strong universal tendency affecting natural language numeral systems. What can explain this universal? I argue in the next section that it would be extremely implausible to attempt to explain it by appeal to some inherent structure built into the language-acquisition device, predisposing the child to internalize a system conforming to the Packing Strategy, in preference to other systems equally compatible with the set of numerals he hears around him. In all of what follows it must be kept in mind that the fact that a child learns a system conforming to the Packing Strategy in no way necessitates the view that he mentally possesses the Packing Strategy, either innately or as a principle induced from data. It could be that the examples to which he is exposed all happen to conform to the strategy and he simply learns to reproduce these and similar examples by means of idiosyncratic and *ad hoc* rules.

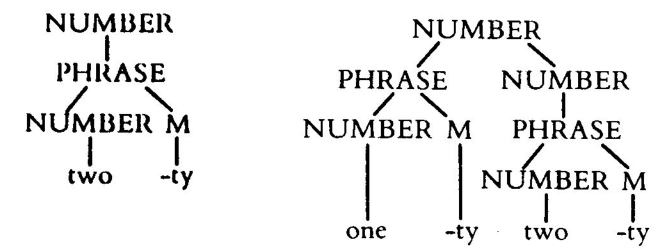
6.3 The Packing Strategy is Not Innate

The Packing Strategy is a *comparative* (technically a *transderivational*) constraint on the well-formedness of numeral expressions. Its operation in accounting for any particular numeral depends crucially on a comparison with other numerals in the language. In accounting for any one numeral it requires the availability of various pieces of information relating to other numerals. It is conceived as a filter constraining the output of highly recursive phrase structure rules which massively overgenerate. These rules typically generate a large class of structures for a given arithmetic value, and the application of the Packing Strategy systematically eliminates all but one of these as that of the appropriate expression for that number in the language. The successful structure is selected because it is, in a well-defined sense, the best available; it is not well-formed in any absolute sense, regardless of the availability of other structures.

Thus the introduction into a language of a new numeral word can cause previously well-formed expressions to become ill-formed. An example from *LTN* (pp. 95–98) is the case of modern Mixtec in which the Spanish loanword *sientu* (= 100) has been introduced. Where before the introduction of this word numbers over 100 had been expressed simply as multiples of 20 (for example 120 = 6 20, 200 = 10 20, 300 = 15 20, 380 = 19 20), all these expressions fell out of use on the introduction of a higher-valued base word than that for 20. [Mixtec data are from Merrifield (1968). It is hard to imagine that there was not a transitional period during which both old and modern systems were used.]

To take the simplest case, the phrase structure rules (6.2.1) in the previous section generate two distinct structures semantically evaluated as 20. These are given in (6.3.1).

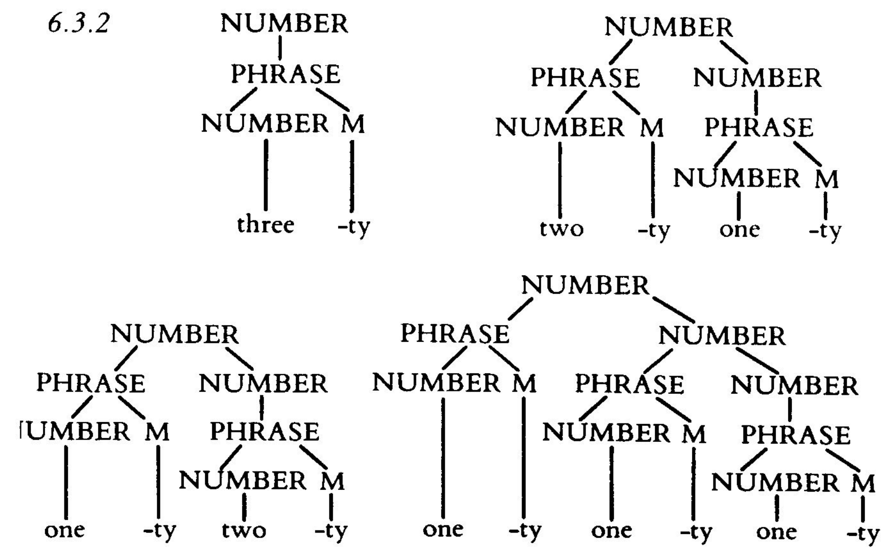
6.3.1



(The first structure would be realized as *twenti* and the second as *ten ten*. Treating *ten* as a suppletion for **onety*, just as *went* is a suppletion for **goed* allows the capture of a generalization in the phrase structure rules.) Clearly, *twenty* is well-formed, while *ten ten* is not; the Packing Strategy predicts this by eliminating the second structure. In this simplest case, application of the Packing Strategy requires comparison of just two rival structures, corresponding to the arithmetic formulae (2 × 10) and [(1 × 10) + (1 × 10)]. [Note that it not just a case of selecting the shortest, or simplest, expression for a number; a preferred form can be longer or more complex than one eliminated by the Packing Strategy, for example Welsh *deg a tri ugain* (10 + 3 20 = 70), which is longer than the theoretically possible *saith deg* (7 10)].

Take now the next simplest case, of structures evaluating to 30. The rules of (6.2.1) generate four distinct structures with this value. They are given in (6.3.2)

6.3.2



These correspond respectively to the formulae (3 × 10), ((2 × 10) + (1 × 10)), ((1 × 10) + (2 × 10)), and ((1 × 10) + ((1 × 10) + (1 × 10))). The Packing Strategy characterizes all but the first here as ill-formed because of the availability within the system of the PHRASE in the first structure, which has a higher value than any of the topmost PHRASEs in the other structures. In this case, application of the Packing Strategy requires comparison of four structures, with subsequent elimination of all but one. We do not express 30 as *twenty ten*, *ten twenty*, or *ten ten ten*, even though it is clear that if these expressions were well-formed they would (or at least could) signify 30.

To generalize, for an arbitrary arithmetic value 10*n* the rules of (6.2.1) generate at least 2^n distinct structures with that value. I say 'at least' because the availability of the words *hundred*, *thousand*, and so on beside *-ty* makes for further possibilities, whose extent I will not trouble to compute. It is clear that for a number as modest as, say, 210, the rules of (6.2.1) will generate *well over a million distinct structures, of which all but one are eliminated* by the Packing Strategy.

Psychological reality for the Packing Strategy was explicitly disclaimed in *LTN* (p. 105); indeed it seems inherently ridiculous to claim that over a million structures, corresponding to crazy expressions such as *one hundred and ninety twenty*, *one hundred one hundred and ten*, and others far, far worse are all in some sense present to the mind and systematically eliminated when a speaker conceives of the standardized way of expressing 210 in English.

From the extreme implausibility of the Packing Strategy as a psychologically real component or determinant of the linguistic competence of a speaker of some particular language one can infer, *a fortiori*, its implausibility as a component of the innate language-acquisition device. Presumably language acquisition does not involve actually forgetting (or in some sense losing) anything known innately, but rather involves amplifying and filling out the innate knowledge. So, if the Packing Strategy is not present in the mind of the adult speaker, it is not present in the mind of the newborn child.

The argument presented here is so straightforward that its point is likely to be missed unless the two component propositions that it brings together are clearly grasped. These are:

6.3.3 Treating the numerals of a language as an unordered set to be generated by means of rules which maximize the capture of significant generalizations (that is adopting for numerals the orthodox methodology of generative grammar) leads one to a formulation of the Packing Strategy for a significantly large number of the world's languages.

6.3.4 The Packing Strategy is extremely implausible as a component of adult linguistic competence, and hence also as a component of the innate mental *faculté de langage*.

Both of these propositions are open to objections of a standard nature, which I will try to refute.

Proposition (6.3.3) is open to the objection that there could be some way, within the standard methodology of generative grammar, of accounting for the generalizations apparent in numeral systems without resorting to a device like the Packing Strategy. To this objection I can only reply that no such alternative account has been proposed, and the *LTN* account, which features the Packing Strategy, remains the principal account of numerals within the generative framework. This claim is more modest than it may seem, since the area of numerals has not been a battleground of linguistic controversy. The *LTN* formulations may be wrong, but until this is shown, we have no reason not to accept them. Note that it is not a valid argument against proposition (6.3.3) to say that the Packing Strategy is implausible as a component of human linguistic competence. To assert this would be to prejudge the very issue I am addressing, namely that in certain instances the fruits of generative methodology can be psychologically implausible. [Such an assertion would, however, actually support proposition (6.3.4) above.]

Proposition (6.3.4) is open to the objection that implausibility can be predicated only of models of performance and not of models of competence. That is, the objection goes, the fact that an enormous amount of computation is involved in working out the standardized expression for a given arithmetic value relates only to the question of what speakers actually do on particular occasions of language use, and not to the tacit knowledge of the linguistic system which is contained in their minds at all times, whether language is being used or not. The issue thus raised is at the heart of much general debate about the empirical status of generative grammar, and cannot be done full justice to here, but I believe that a straightforward reply to the above objection can be given along the lines sketched below.

The standard view of the relationship between competence and performance – as outlined, for instance, in Chomsky (1965) – is that a theory of competence forms one subpart of a theory of performance. In performance, a variety of factors dealt with by different theories are involved, for example social constraints on politeness, general limitations on memory, and so on. An individual's linguistic competence, his knowledge of his language, is one of several factors determining his linguistic performance. The appropriate division of labour between the various components of a theory of performance has not been sorted out, and if pursued too far, such questions can become merely terminological. For example, the question of whether some particular statement counts as a perceptual strategy or a rule of grammar may in some cases not be resolvable except by some arbitrary fiat. (See Hankamer, 1973, p. 36n. for a succinct statement of this position.) But note that it is not usually envisaged that any part of a theory of competence lies outside a theory of performance. That is, there is no part of an individual's knowledge of his language that cannot in principle contribute to some aspect of his linguistic performance (even if only in the performance of paper and pencil exercises).

It follows from this view of the relation between competence and performance that if any statement is implausible as a part of a theory of performance, it is, *a fortiori*, implausible as a part of a theory of competence. To deny psychological reality to the Packing Strategy as a factor in performance automatically denies it psychological reality as a factor in competence, since competence is (and is only) a component of performance. More informally, to say of the Packing Strategy, 'surely all that computation couldn't be involved when an English speaker conceives of the meaning of 210 and says *two hundred and ten*' is to say also that surely all that computation couldn't be involved when an English speaker stores the knowledge that 210 is expressed as *two hundred and ten*, since the latter knowledge is the basis for his ability to perform the former feat.

Power and Longuet-Higgins (1978) have shown that a device for acquiring competence in natural language numeral systems need not incorporate anything corresponding to the Packing Strategy, despite its (near) universality. Power and Longuet-Higgins describe a computer program capable of learning numeral systems. After a period of instruction in which the program is given a sample of meaning-form pairs (for example 210, *two hundred and ten*), the program is able to supply, for any numeral

meaning given to it, the appropriate form for that meaning in the language concerned. The program will also decode arbitrary numeral forms given to it. The program manages to learn some quite complex numeral systems, for example French and Biblical Welsh. But nowhere in the program is there any statement, routine, or whatever corresponding in any way to the Packing Strategy. Although Power and Longuet-Higgins' program does not exactly simulate the circumstances in which real children learn numerals, there is a strong suggestion in their work that numerals can be learned without the Packing Strategy as a guide and that there is no need to suppose that children possess an innate version of the Packing Strategy in order to explain their ability to learn this part of the grammar of their language.

There is also empirical evidence from children themselves against the Packing Strategy as part of an innate schema brought to the language-acquisition task. Where children generalize beyond the linguistic data they have heard, the pattern of the generalization can be taken as evidence of some internal (and presumably innate) specification of what a natural linguistic generalization is. In some cases, no doubt, such generalizations actually mould the form of the language as each successive child acquiring the language makes the natural generalization. In this sense, parts of the language are re-created by each new language acquirer. I do not doubt that this Chomskyan account is correct of some universal regularities found in languages. But in the case of the Packing Strategy (a generalization handling a range of such regularities) the generalizations actually made by children often violate the strategy. Fuson, et al. (1982, p. 56–7) list 54 different invented expressions (types) produced by 96 3- to 6-year-olds. The examples include *ten-eighty*, *twenty-fourteen*, *thirty-nineteen*, *sixty-twenty-seven*. Several such expressions occurred quite commonly, for example *twenty-ten*. All but two tokens of such invented expressions are violations of the Packing Strategy. (The two invented forms which are not violations of the Packing Strategy involve failure to shorten and shift the first vowel in **fiveteen* and **fivety*.) Genie, the linguistically deprived child, produced *thirty ten* (Curtiss, 1977, p. 165). Gwen Awbery (personal communication) tells me that in her son's learning of the traditional Welsh numerals 'efforts like *saith ar bymtheg* for 22 are quite common!' [*Saith ar bymtheg* (7 on 15): *dau ar hugain* is the correct form, predicted by the Packing Strategy.] A somewhat more systematic sampling by the author of such errors for higher-valued numerals is reported below.

In a very simple experiment, five children from the 2nd year of primary school (aged between 5.5 and 6.5) were sat around a tape-recorder and asked to 'think of a very big number'. When they suggested numbers, I encouraged them to suggest more by such remarks as, 'Yes, that's a really big one!', and 'Can you think of an even bigger one?' In about 5 minutes, over 60 utterances were recorded. These were transcribed, attributed to the individual children (by voice and accent), and analysed. Of the complex numeral expressions used, 33 conformed to the usual arithmetic relations between numeral words observed in adult English, and 35 did not. Examples are given in (6.3.5).

6.3.5

Conforming to adult rules	Not conforming to adult rules
seven thousand billions	million hundred
seven hundred gillions	sixty hundred
a trillion and a billion	sixty hundred and million ten hundred
eleven hundred	a hundred and a million and a hundred
a billion billions	

(Fictitious numeral words such as *gillion* and *pillion* were counted as correct if used in what would have been an acceptable context for *billion* or *trillion*. There were three such cases, all in examples classed as correct.) The children varied in the overall correctness of their performance, as shown in (6.3.6).

6.3.6

Child	'Correct' expressions	'Incorrect' Expressions	Percentage Correct	MLU Counted in Numeral Words
J	9	0	100	3.33
K	7	5	58	2.08
C	10	8	56	2.67
G	3	4	43	2.29
T	4	18	18	2.50

The most advanced child, J, seemed to control the adult system completely. His longest example was *seven thousand eight hundred and fifty trillion*. The other children still made many mistakes in the ordering of numeral words. The worst performer was the hardest trier, T, attempting long expressions and giving more examples than the other children. One of her incorrect examples was *seventy eight and a hundred and a million*.

Several of these children seemed to realize the potential for stringing numeral phrases together *ad lib*. They had apparently grasped the recursiveness of the syntax of complex numerals, but without mastering the concomitant constraints on ordering. In this respect, another child, E, aged 6 years, tested separately, showed a delight in exploiting the recursive potential, producing expressions such as the following:

6.3.7 two trillions three thousand nine hundred two billion one trillion and one thousand
nine hundred two thousand one trillion one hundred two billion and three thousand
three thousand two million one trillion a thousand billions and ten hundred
hundred billion thousand million hundred
a hundred thousand billion trillion million thousand hundred

These data suggest that the child is innately receptive to intimations of recursive syntax, and spontaneously produces expressions with highly recursive structures such as she is unlikely to have heard from adults. Adults would in fact be inhibited by the lack of high-valued M-words, so that after *hundred thousand million billion trillion* there is nowhere to go. But this does not inhibit the child, who appears to string these words and phrases together in a more or less random order. The child possesses nothing resembling the Packing Strategy. In all, E produced 22 such expressions during a test session when asked to think of big numbers. Of these, only six were fully correct.

Despite the incorrectness of the examples, it is clear which adult structures are being imitated. In each of the first two examples above, for instance, a succession of six PHRASEs with the structure [DIGIT M] is strung out just as in an adult additive construction. In each of the last two examples above, a series of Ms (*hundred*, *billion*, and so on) forms what would be a deeply nested structure of PHRASEs within PHRASEs in adult usage, interpreted by multiplication. Thus it is possible to see (imitations of) additive relations between successive PHRASEs and multiplicative relations between successive Ms in these examples. In the adult language, added PHRASEs are ordered highest first, as in *two thousand three hundred*, and multiplied Ms are, conversely, ordered lowest first, as in *thousand million*. The child E got these pairwise transitions between PHRASEs and between Ms in the right order with roughly chance frequency, as shown in (6.3.8).

6.3.8 Ordering in pairwise transitions

	Correct order	Incorrect order
Additive Structures	e.g. seven thousand nine hundred	e.g. three thousand two billion
	19	17
Multiplicative structures	e.g. hundred thousand	e.g. thousand hundred
	13	10

The data from E do in fact show one very significant constraint on the ordering of PHRASEs and Ms. Out of 59 pairwise transitions, either between PHRASEs (in imitated additive relationship) or between Ms (in imitated multiplicative relationship), only two transitions were between (expressions containing) the same M-word, or PHRASEs containing the same M. The two examples were *thrillion thrillion* (using a fictitious word) and *one trillion two thousand and a hundred and a hundred and three*, in which two PHRASEs with *hundred* are juxtaposed. Given only six Ms to choose from (*hundred*, *thousand*, *million*, *billion*, *trillion*, *thrillion*) and 59 transitions between (PHRASEs containing) members of this set, the figure of only two transitions between similar items is very significant. Despite what otherwise appears to be random ordering of PHRASEs and Ms in these recursive structures, the child is clearly avoiding juxtaposing similar expressions. 'Each M-word must be different from the preceding one' seems to be a rule that the child observes.

The avoidance of juxtaposing similar items is discussed by Menn and MacWhinney. Although their discussion concentrates on morphology, their conclusion may be generalizable to syntax. They summarize: 'Strong grounds exist for claiming that there is a general output constraint which tends to prohibit sequences of phonologically identical morphs. Since violations of the constraint certainly exist, the proposed constraint is properly referred to as a weak morphological universal' (1984, p. 529). As part of an explanation for this constraint, they argue that 'accidental morph repetition creates some inconvenience for language processing' (p. 519).

This section has argued that the highly recursive syntax of complex numerals can be attributed to an innate disposition to internalize recursive rules, which generate longer examples than a language-acquirer is likely to have heard. But the Packing Strategy, the arithmetically stated universal principle which severely constrains the output of such highly recursive rules, cannot plausibly be attributed to any innate psychological disposition of language-acquirers. Thus an alternative explanation is needed for the severely constrained set of arithmetical combinations found in numeral systems. The next section moves towards this, by developing a social conventional concept of 'standardized expression', as opposed to the mental or psychological notion so far adopted of 'well-formed expression'. Of course, if speakers conform to social conventions, then they probably have (sets of) mental representations corresponding to the conventions, but speakers do not necessarily represent the conventions to which they conform as single, economical generative rules. The mental representations may actually be lists of apparently unrelated facts with no common motivation save that they are all conventions which the speaker has learned. The fact that an economical statement *can* be made summarizing the essence of these separate facts does not mean that the mental representation is unified. The unifying causal factor that one naturally seeks could lie outside the synchronic psychological domain, which only contains its separate individual effects. In the subsequent sections a diachronic social mechanism will be proposed to account for the arithmetical constraints on standard numeral systems. At that stage, the psychological inconvenience in processing repeated forms, for which numeral evidence is given above, will be reinvoked.

6.4 Standardization of Numerals – Semantic/Pragmatic Motivation

Compare the expressions in the two columns in (6.4.1).

6.4.1

A	B
fourteen	two sevens
	twice seven
	eight and six
	a dozen plus two
	one thousand less nine hundred and eighty-six
	(etc.)
forty-nine	fifty minus one
	two twenties and nine
	seven squared
	(etc.)

Although all these are well-formed English numeral expressions, there is a clear difference between those in column A and those in column B. Expressions like those in column A I shall call 'standardized numeral expressions', and expressions like those in column B I shall call 'non-standardized'. (This terminology implies no distinction between the two types of numerals related to the social status of their users.) The phenomenon of the existence of this distinction I will call the 'standardization' of numerals. Languages, where they have a numeral system at all, generally exhibit a distinction between standardized and non-standardized expressions; this is a general characteristic of languages and stands in need of explanation. A number of criteria allow the distinction to be discerned in languages generally.

Typically, the standardized numeral expressions in a language are described in a special section or chapter of a traditional or pedagogical grammar of a language, usually headed 'Numerals', whereas non-standardized expressions are typically not mentioned, or only mentioned briefly and in passing, in such grammars.

In answers to 'How many?' type questions, non-standardized numeral expressions feel definitely marked in nature, and seem unduly coy, indirect, or uncooperative, whereas standardized numeral expressions are unmarked, direct, and not unduly coy in answer to such questions. For example

6.4.2 Question: How old is your sister?
Answer (a): Sixteen.
(b): Four squared.
Nine and five.
Twice eight.
Twenty minus four.

Only the latter expressions here seem to violate the Gricean Maxim of Manner ('Be perspicuous').

The syntactic distribution of non-standardized numerals is for some speakers more restricted than that of standardized numerals. For example in English the 'b' expressions below are felt by some to be ungrammatical.

6.4.3 (a) There are forty nine Democrats.
(b) ?There are seven squared Democrats.
?There are fifty minus one Democrats.
?There are thirty plus nineteen Democrats.

Standardized numeral expressions occur in the standard counting sequence of a language, whereas non-standardized expressions do not. For example

6.4.4 ..., ..., eighty, eighty-one, eighty-two, ...
* ..., ..., fourscore, nine squared, two forty-ones, ...

In mental arithmetic reported aloud non-standardized expressions are used freely, but the purpose of mental calculations of this sort is usually to reduce, or equate non-standardized expressions to standardized ones. The multiplication tables we all learn by heart do this, for example *six eights* (non-standardized) are *forty eight* (standardized). A multiplication table relating nonstandardized expressions to other nonstandardized expressions would be less useful, for example *six eights are four twelves*.

Sometimes standardized numeral forms follow the same phonological and syntactic rules as non-standardized ones. For instance, the very general English listing construction with *and* is used in both of the following:

6.4.5 three thousand, four hundred and eight (standardized, = 3408)
three, four and eight (non-standardized, = 15)

But in many instances standardized numeral expressions follow rules which are peculiar and restricted to just this class of expressions. Thus suffixation of *-ty* to one of the single digit numerals *two*, . . . , *nine*, with concomitant idiosyncratic phonological changes, is the standardized way of expressing multiplication by 10 in English. But this construction cannot be extended. **eleventy* and **twelvety* are not merely non-standardized; they are downright illformed in modern English even though they can be semantically interpreted. The simplest non-standardized ways of expressing what one would mean by **eleventy* and **twelvety* are *eleven times ten* and *twelve times ten*, or *ten elevens* and *ten twelves*. Standardized complex numeral expressions have clearly often become lexicalized to some extent in previous stages of a language and undergone the more or less idiosyncratic phonological and morphological changes affecting single lexical items.

The standardization phenomenon normally involves a single expression being the standardized form for a particular number, but there are exceptions to this very general tendency. One kind of exception occurs when a language combines expressions for the same constituent numbers, using the same arithmetical operation, but there is a possibility of some kind of stylistic reordering. In some dialects of nineteenth-century British English, for example, both *twenty-five* and *five and twenty* were found. Stylistic variation of this sort seems particularly rife in some Classical Welsh dialects, where the same number, 182, might be expressed as any of the following:

6.4.6

cant	a	phedwar	ugain	a	dau
100	+	4	20	+	2

cant	a	dau	a	phedwar	ugain
100	+	2	+	4	20

dau	a	phedwar	ugain	a	chant
2	+	4	20	+	100

(See *LTN*, Chapter 6, for more details of this, and other kinds of variability in Welsh.)

Another kind of exception to the generalization that standardization involves a single standardized form for each number occurs when there are synonymous standardized expressions formed from

different constituents and perhaps involving different arithmetical operations. Examples would be English *one thousand, one hundred* versus *eleven hundred* and *nine thousand nine hundred* versus *ninety-nine hundred*. Note in such examples, however, that still there are only two standardized forms, as opposed to the vast range of non-standardized ways in which a number can be expressed (for example *ten thousand minus a hundred, eight thousand plus nineteen hundred*, and so on). Even though both *one thousand, one hundred* and *eleven hundred* are well attested in English, speakers will often judge expressions of the former type to be in some sense more proper or less slangy. So a high degree of standardization exists, even where more than one standardized form remains. There is a very general tendency for the degree of standardization to decrease, though perhaps only slightly, the higher and the less round the number being expressed. Thus in Yoruba, where there is unusual variety in the standardized numerals, low-valued multiples of a base number are typically expressed by only a single standardized expression, whereas higher and more awkward numbers may often be expressed by several standardized expressions.

The phenomenon of standardization could arise, broadly speaking, in two different ways. One possibility is that standardization is a diachronic process whereby, in a stable language community, the range of preferred expressions for some particular number becomes gradually narrower, the process eventually culminating in there being just one standardized expression for each number. In the exceptional cases where there is actually more than one standardized expression, one would postulate simply that the process has not yet reached its culmination. Where the language community is not stable, one may expect the standardization process to be arrested, or even temporarily reversed, by interference between dialects and language varieties. [Hetzron (1977) argues that the modern Central Semitic 'Digit-Teen' word order replaced an original 'Teen-Digit' order through cultural interference from Akkadian.] I cite below some evidence, from French, of the diachronically progressive nature of the standardization process, in a situation where rival decimal and vigesimal systems have been thrown together by external historical events. Brunot writes thus of the state of French in the sixteenth century.

> La lutte continue entre les nombres hérités du latin pour les dizaines, et les formes faites par addition: *soixante dix, quatre*

vingt dix. Presque tous les grammairiens donnent encore *septante* et *nonante*. Cependant Palsgrave reconnaissait que si cette manière de compter était celle des gens instruits, le peuple tenait pour *soixante dix*, et Meigret dit formellement que la manière nouvelle est plus reçue et plus approuvée. Ces témoignages sont confirmés par celui de Fabri, qui se plaint 'de cet erreur incorrigible de dire *quatre vingt douze* pour *nonante deux*'.

... En second lieu, il faut signaler la continuation de la lutte entre le système latin de numération par dix et le système rival de numération par *vingt*.

Quatre vingts s'impose peu à peu au dépens de *octante* ou *huitante*. Non que *octante* soit proscrit; il est au contraire recommandé par plusieurs grammairiens et donné par tous. Il se rencontre de même chez les auteurs.

Mais Meigret considère déja *qatre vins* comme plus reçu.

En revanche, les autres multiples de *vingt*, quoiqu'usités jusqu'à 400, ne sont pas également en usage. *Sis vins* l'emporte sur *cent vins*, mais *cent soessante* est aussi bien dit que *huyt vins*, et *quinze vins*, sauf dans le nom de l'hospice, est à peu pres abandonné. (1906, pp. 309–10)

If and when standardization comes about by diachronic elimination of rival expressions, the result may be either uniform standardization across the whole original language community or a splitting into several subdialects in which the standardized forms differ. In both cases, an individual speaker in a community before standardization has more freedom of choice, so to speak, than a speaker in that community after standardization. The splitting into subdialects is illustrated by the French case, where modern Belgian, Swiss, and French dialects differ in detail in the forms which have survived the contest between Latin decimal and Celtic vigesimal forms.

6.4.7	French	Belgian	Swiss
70	soixante dix	septante	septante
80	quatre vingts	quatre vingts	octante
90	quatre vingt dix	nonante	nonante

The other possible way, broadly speaking, in which the phenomenon of standardization could arise is by a kind of 'prior arrival pre-empts survival' mechanism. That is, present standardized numerals may be the direct descendants of the first numeral forms invented, and other forms invented later may never have been candidates for standardized status. In this case, there is no kind of elimination of rival forms. Quite possibly, both the pre-emption-by-prior-arrival and the elimination-of-rivals explanations could be true for the standardized forms in a given language, with, say, standardization of lower-valued numerals explainable by one means and standardization of higher-valued numerals explainable by the other. The elimination-of-rivals process is historically attested in cases such as French. And it is hard to see how the pre-emption-by-prior-arrival model could account for cases of a plurality of standardized forms, such as is found with the higher-valued numerals in Yoruba.

With respect to the phenomenon of standardization, one can pose two questions, a general and a specific: (1) Why does it come about at all? that is, why are some expressions from paraphrase sets accorded a special status? And (2) given that standardization exists, why does it take the specific form that it does? Why, for example is *seven sevens* not the standardized English expression for 49, as opposed to *forty-nine*? The most interesting subpart of this second, more specific question is the question of why standardized numerals involve uniform *counting on a base*, for example 10 or 20. The previous section argued against a psychological answer to this specific question, an answer which would postulate psychological reality in individual minds for a principle such as the Packing Strategy. In the next section, a social diachronic mechanism will be shown giving rise to the standardization of numeral systems to a specific base.

The straightforward answer to the first, more general, question of why standardization should occur at all can be illustrated by an everyday example. A mother wants to know how many eggs she has in the refrigerator, so she sends two of her children to find out for her. She sends them separately, in order to double-check the reports she receives. One child reports that there are *six times eight* eggs in the fridge; the other child reports that there are *four times twelve*. How does the mother know that both children are in effect telling her the same thing? A tedious method would be to construct, with sticks or something, a model of *six times eight*, for example by making six piles of eight sticks, and then to see that the same model (the collection of sticks) can be arranged into four piles of twelve. A preferable method is to have remembered, or to be able to generate, a linguistic equation *six times eight equals four times twelve*. But even better would be for the mother to teach her children to use canonical standardized expressions, eliminating the need for any calculation at all.

The communicative advantages of standardized canonical expressions are obvious. The expression *forty eight* is not intrinsically or *a priori* more transparent than the non-standardized *six times eight* or *four times twelve*; the apparent greater directness with which standardized forms speak their meanings to the mind (putting it vaguely) arises simply from the fact that they are canonical, socially inculcated in children as the standard forms by drilling in the counting sequence and rote memorization of multiplication tables. The counting sequence, in which children are routinely drilled, is an instrument by which the standardization of numeral forms is reinforced. The fact that the counting sequence is a series of expressions in a one-to-one correspondence with positive integers (up to some finite limit) ensures that each number is expressed by one and only one expression. To have branchings in the standard counting sequence, to cater for alternative expressions for the same number, would confusingly disrupt its natural structure and the rhythm with which it can be put to practical use in counting off objects in a collection. It is to be expected that in communities where there is less formal drilling of the numeral forms, there will be a less marked difference between standardized and nonstandardized expressions.

The active propagation and dissemination of standardized forms of expression is familiar in scientific, technical, and trading circles, where the advantages of generally accepted standard notations are appreciated, and the rather closed social structures involved permit and even encourage adherence to the discipline of a standard usage. But the everyday non-technical language of an open mobile society is well known to be less susceptible to any externally imposed discipline. It is virtually impossible to legislate for everyday language. The use of numerals is not restricted to specialized technical groups, but is widespread through whole communities. How, in this apparently isolated case, has a high degree of standardization managed to manifest itself in open and loosely structured language communities? Well, if it happens by diachronic elimination of rival expressions, it does not necessarily happen quickly, as the French evidence quoted above shows. That the phenomenon should arise at all, whether by elimination-of-rivals or by pre-emption-by-prior-arrival, can be explained by appeal to the special nature of numeral meanings. Numerals are a special case, and, unlike other areas of language, preserve or gain no communicative advantage by resisting standardization, as I argue in detail below.

Take first the case of variability in the order of constituents, as in *five and twenty* versus *twenty-five*. Alternative orderings of constituents are generally much more common, across languages, in higher-level structures, such as sentences and clauses, than in lower-level structures, such as noun phrases, adjectival phrases, and prepositional phrases. English examples are given in (6.4.8) and (6.4.9).

6.4.8 If it rains, I'll go.
I'll go if it rains.
John turned to the vicar impetuously.
John turned impetuously to the vicar.
John impetuously turned to the vicar.
Impetuously, John turned to the vicar.
I adore Brahms.
Brahms, I adore.

6.4.9 The red table
*The table red
*Red the table, etc.
On the table
*The table on
Very nice
*Nice very

Even the order of adjectives in NPs is fairly tightly constrained: *scruffy old black cat* versus **scruffy black old cat* and *?old scruffy black cat*. Variability in high-level structures has a communicative function. It allows a speaker to relate his utterance to preceding discourse, and to adumbrate directions for the ensuing discourse, in ways which apparently affect the overall nature of conversational interactions. But there is no such function for variability in lower-level structures which (therefore?) tend to be less variable in the ordering of their constituents. Numerals are low-level structures, being embeddable as nominal modifiers in noun phrases. The difference between *twenty-five* and *five and twenty* has no communicative function in the way that the difference between *I like Brahms* and *Brahms, I like* has. This lack of communicative function for variable ordering does not cause, but it does permit, standardization of the order of constituents in numerals.

It might be argued that the case of *five and twenty* versus *twenty-five* is a case of the alternative ordering of co-ordinated constituents, and that there are, in general, no constraints on the ordering of co-ordinated constituents, at any level of structure. Thus, for instance, both *John and the King of Siam* and *The King of Siam and John* are equally acceptable in principle and the difference in order may even correlate with differences in the context of use. The ordering of co-ordinated constituents in actual use does seem to have something to do with the relative familiarity of speakers with the referents of the constituents. For instance, my wife and I are usually referred to by my family and my work colleagues with the expression *Jim and Sue*, while my in-laws and my wife's close friends typically refer to the same couple using the expression *Sue and Jim*. This is a fact. While the difference in order may not exactly have a communicative function, it does seem to have some correlation with the salience in the mind of the speaker of the referents involved. Presumably the artificially imposed preference for NP *and I*, as opposed to *I and* NP or *me and* NP, reflects a feeling that polite people should consider other people more important (salient) than themselves. This kind of consideration, however, plays no part in the ordering of co-ordinated constituents within numerals, due to the peculiar way in which we know the meanings of numerals, which is unlike the way in which we know the referents of referring expressions like *Jim* and *Sue*.

Although speakers may believe that each numeral stands for some abstract entity called a 'number', they do not have any language-independent stored representations of the denotations of the individual numerals (except in the case of the very lowest-valued ones). In the case of physical objects for which a speaker has names or linguistic descriptions, he may usually be said to have concepts of both the signifier and the signified, of the name – or description(s) – and of the object itself. In the cases of numbers/numerals, a non-linguistic concept of the signified is, oddly, lacking, although it is convenient to maintain that there is indeed a signified entity, an abstraction referred to as a 'number'. While a speaker who uses the expression *Jim and Sue* may know Jim and/or Sue as individuals in the real world, and indeed have known them before knowing their names, it cannot be similarly argued that a user of the expression *twenty-five* is acquainted with the individual numbers 20 and 5 in any way independent of their

expression in some linguistic code, for example the English words *twenty* and *five*. Thus, while different real-world contexts may give rise respectively to both *Jim and Sue* and *Sue and Jim*, it is much less plausible that different contexts should in the same way give rise respectively to *twenty-five* and to *five and twenty*. The perfect reversibility of many coordinated pairs of expressions reflects the fact that in actual life the referents of the constituents may come at us in either order, but numbers, except the very lowest, only come to us through the ritual of counting, or already clothed in their linguistic forms, and there is less reason for any variability in the ordering of constituents to survive.

Numerals do not provide the only examples of standardized invariant ordering of co-ordinated constituents. Stereotyped phrases such as *fish and chips*, *bread and butter*, *gin and tonic*, *well and truly*, and so on are standardized in a similar sense, as compared to *chips and fish*, *butter and bread*, *tonic and gin*, *truly and well*, and so on. These latter phrases are certainly not ungrammatical in English. What seems to be going on here is that the combinations referred to (for example the fish plus the chips) are very familiar, and are thought of as whole complex unities. The ensemble is more salient than either of its ingredients. With numerals, the situation is similar, though clearly not identical. In expressing 25, what is important is to get the right expression for the whole number; marshalling the constituent forms from component numbers is entirely subservient to the goal of expressing the ensemble and the question of their relative salience *vis-à-vis* each other hardly arises.

Turning now to a different kind of variability, in which a number is expressed by arithmetical combinations of different constituent numbers, for example *four score* versus *eighty*, or French *quinze vingts* versus *trois cents*, a similar lack of motivation for variability can be discerned. I will use an everyday example again to illustrate.

Consider my neighbour. Depending on who is talking to whom, and when and where, this single individual may be referred to by any of the following expressions and more:

6.4.10 Gertrude's husband
My neighbour
The man who walks his dog in the park
The supervisor
(etc.)

Loosely speaking, predicates such as *husband*, *neighbour*, *man*, *walk*, *dog*, and so on can be used to construct referring expressions which are maximally transparent to a potential hearer, taking into account what entities in the world are actually known to him. Thus, I would hardly refer to my neighbour as *Gertrude's husband* to an addressee who does not know who Gertrude is; and *the man who walks his dog in the park* is appropriate when addressed to someone who knows which park I am talking about and knows that there is a man who walks his dog there. Hearers gain access to the referent of a complex expression addressed to them through knowledge of the referents of its constituents. The basic prior knowledge presupposed here is non-linguistic. The entity referred to as *Gertrude* is knowable independently of any linguistic expression used to refer to her: you may have seen her, for example. In short, the availability of a plurality of expressions referring to the same individual is a most useful characteristic of language, allowing us to refer successfully in different contexts and for interlocutors with different knowledge. But in the case of numbers/numerals, our knowledge of the domain is uniform. We cannot find people who know the numbers 6 and 8 better than the numbers 4 and 12, so that to them we express 48 as *six eights* rather than as *four twelves*. Thus there is nothing in the ways in which different speakers know the number domain which might tend to foster variability in the components chosen for the purpose of referring to some particular number.

A kind of standardization of referring expressions is not unknown outside the numeral domain. It occurs, as one would expect from the argument presented above, when knowledge of a referent is relatively uniform across all speakers. Compare the following forms:

	A	B
6.4.11	the sun	the star at the centre of our solar system
		Phoebus
		the great fireball 93 million miles away
	the moon	Earth's natural satellite
		Cynthia
		the destination of the last Apollo missions

Someone choosing the B expressions here would be doing so for some unusual calculated effect, and would, in normal

circumstances, be violating the Gricean maxim of manner. The A expressions are, in a somewhat weaker sense than in the case of numerals, standardized.

In summary, standardization of numerals arises because of the obvious communicative advantages of standardized, canonical forms, and because no communicative advantage is lost by such standardization, due to the rather special nature of numerals/numbers. This is a matter of (linguistic) evolution, and it is to be noted that the idea of natural selection (*mutatis mutandis* for the case of linguistic forms) is emerging in the argument. The argument for a kind of linguistic natural selection needs next to focus on detailed questions, and to show how it comes about that specific forms with certain properties tend, universally, to be the standardized expressions.

Mill comments on the standardization of numeral expressions (without using this terminology)

> The modes of formation of [i.e. the various expressions expressing] any number are innumerable; but when we know one mode of formation of each, all the rest may be determined deductively. . . .
>
> It is sufficient, therefore, to select one of the various modes of formation of each number, as a means of ascertaining all the rest. And since things which are uniform, and therefore simple, are most easily received and retained by the understanding, there is an obvious advantage in selecting a mode of formation which shall be alike for all; in fixing the connotation of names of number on one uniform principle. The mode in which our existing numerical nomenclature is contrived possesses this advantage . . . and this mode of its formation is expressed by its spoken name and by its numerical character. (1906, p. 401)

Mill is surely right about the advantage of 'fixing the connotation of names of number on one uniform principle', and, as indicated by the reference to the 'spoken name', he includes natural language numerals in his generalization. I have here established the broad fact of the standardization phenomenon, which is related to the peculiar nature of numeral meanings. We will not be concerned further with non-standardized numeral expressions. Mill suggests a vague explanation for which specific numeral expressions get standardized (omitted in the ellipsis in the above quotation). In the next section, I propose a specific diachronic social explanation for the particular ways in which languages choose to standardize numeral expressions.

6.5 Evolution of a Standardized Base by Social Negotiation

Numeral systems evolve by gradual stages. Grosso modo, the history of a numeral system is a series of static periods, during which the set of rules for arithmetically combining simple expressions to form more complex expressions do not change. Connecting these static periods are periods of innovation or invention, in which one or more new rules or new lexical items are added to the system.

The rules I am chiefly concerned with are those which specify the constituency of syntactic constructions, and which associate arithmetical operations, such as addition or multiplication, with them. The proposal is that as numeral systems evolve, such rules may be added to the system, but existing rules are not specifically deleted or modified, although the effect of adding new rules may eventually be to modify or even nullify the output of pre-existing rules. I am not concerned here with diachronic change in other types of rule, such as phonological rules, or stylistic rules determining the linear order of constituents within a construction. Thus, as far as the present proposal is concerned, what is important about English *fifteen* is just that it has two constituents, with arithmetic values 5 and 10, and is interpreted by addition: the fact that *five* is phonologically modified to *fif-*, and that the lower-valued morpheme precedes, rather than follows, the higher-valued one, is, for my purposes here, immaterial.

The introduction of new rules or lexical items into a system comes about by borrowing and/or nonce-formation by inventive individual speakers. I shall not be concerned with the social mechanisms by which new rules and/or words are adopted by the whole of a community, but will assume, as an idealization, that new rules become immediately available to whole communities. But a social mechanism *is* proposed whereby certain outputs of the new rules of a system become standardized, leaving the rest non-standardized. This standardization, according to the model to be proposed, is a slow process spreading its results through the language-community by means of repeated acts of linguistic intercourse between individuals.

Syntactic rules may make available to a speaker a range of synonymous expressions denoting a single number. In the absence of any principled reason for preferring one expression over another, the speaker selects an expression at random from this range. This is the situation immediately after the adoption by a community of a new rule. But as individuals continue to express numbers to each other, it will happen that certain types of expression occur more frequently than others. This provides speakers with a basis for applying a pragmatic principle when choosing from a range of synonymous expressions. The principle is: choose an expression of a type with which your interlocutor is likely to be familiar. This could be seen as a component of Grice's Cooperative Principle, falling perhaps under 'Be perspicuous'.

Now a speaker cannot know for certain what types of expression his interlocutor is familiar with. Familiarity is a function of prior experience. A speaker does not know what experience his addressee has had; all he knows about is his own experience. He makes the necessary assumption that his interlocutor's experience is likely to be similar to his own and bases his choice of expression on the relative frequencies in his own experience of the various expressions in the range generated by the rules of the system. In short, a speaker chooses an expression of a type which, in his own experience, has been used most frequently. This proposal does not require that a speaker be able to maintain exact numerical records of the frequencies of types of expressions he has heard. All that is required is that a speaker have some awareness, possibly temporary, inexact, and represented in quite crude terms, of the relative rareness or commonness of expressions of various types. As more exchanges take place between individuals it happens, as will be demonstrated shortly, that expressions of certain types are chosen with increasing frequency, leading finally to the exclusive use of expressions of this type. This happens because of certain elementary arithmetical facts, as will be explained.

At this stage, we can say that expressions of the type now used exclusively have become standardized expressions. The remaining, non-standardized, expressions may be used in circumstances when the normal pragmatic principles are, for some reason, waived, in particular when the principle 'Choose an expression of a familiar type' is not observed. This fits in well with one of the criteria given early in the previous section, which characterize non-standardized numeral expressions as in some sense less direct and helpful than their standardized counterparts.

An example should make all of this clearer.

Imagine a stage in the history of a language when it has words for the numbers 1-10 (for convenience, call these words *one*, *two*, . . ., *nine*, *ten*), but as yet no syntactic rules for combining these numeral words to form structures with arithmetical values greater than 10. Then, somehow, a rule combining two number words together into a single construction, to be interpreted by the operation of addition, is invented and built on to this primitive numeral lexicon. The rule can be stated as:

6.5.1 Expression → word + word

Immediately, expressions such as the following, with the given interpretations, become available.

6.5.2

(a)	one one	2
	two one	3
	five four	9
	five five	10
(b)	six five	11
	seven four	11
	eight three	11
	nine two	11
	eight seven	15
	nine six	15
(c)	ten one	11
	ten five	15
	ten nine	19

(Remember that the linear order of constituents is not relevant to the discussion here. I adopt the convention of giving higher-valued words before lower-valued words. For present purposes, *seven four*, for example, is entirely equivalent to *four seven*.)

All of the expressions in (6.5.2) are made newly available by the adoption of the new rule. The expressions in (6.5.2a) are synonymous with previously existing one-word expressions. I assume that conservatism and a preference for simpler expressions will guarantee that these new expressions do not become standardized or in any way oust the previously existing one-word expressions. The expressions in (6.5.2b) and (6.5.2c) express

numbers which were not previously expressible in the system. Let us label the highest-valued word in a numeral construction the 'base-word'. The simple model proposed here predicts that expressions with *ten* as base-word, as in (6.5.2c), become standardized, whereas other expressions in this simple example do not.

There is a sense in which *ten* is a more useful base-word than the other words in the system. Using *ten* as base-word, one can form expressions for numbers up to 20, whereas using *nine*, for example, one can only form expressions for numbers up to 18. *Eight* is even more limited as a base-word, allowing the formation of expressions for numbers only as far as 16. And so on. This is a straightforward consequence of the simple arithmetical properties of the numbers involved.

These simple arithmetical properties generate another factor leading to the greater 'usefulness' of the highest available base-number. This factor involves the process of counting a collection of objects. Sometimes the objects in a collection present themselves in convenient subgroupings. Imagine a person with a simple numeral lexicon and a rule forming additive constructions calculating the number of people comprising his own and his neighbour's households. He might reasonably attach a number word to each household and then form an additive expression with them, thus: 'One, two, three, four. That's four here. One, two, three. That's three next door. The answer's four-and-three.' But say he has to count a large flock of sheep as they stream through a gate, not conveniently supgrouped. It would seem natural that he would use up all the words in his lexicon before resorting to the additive construction. Thus he would use the highest-valued available word as a base. A factor such as this could well increase the frequency in use of expressions with the highest-valued base word, although it will not be taken into account in subsequent discussion.

If speakers express numbers from 11 to 20 to each other, using the new additive construction (assuming all numbers from 11 to 20 are expressed with equal frequency), and if they choose their expressions at random from among those made available by the new rule, inevitably more expressions with *ten* as base-word will be used than any other type of expression. If hearers classify expressions which they have heard uttered to them in terms of base-words, the class with base-word *ten* will tend to be the most frequent in the experience of the typical individual. And if, in choosing an expression for a particular number (where there is a choice, as between *eight seven*, *nine six*, and *ten five* for 15) an individual chooses an expression of the class he has heard most frequently, then expressions with base-word *ten* will tend to become acceleratingly more frequent. There will be a snowball effect.

A computer simulation of the above process will be described in the next section, showing that under reasonable, if idealized, social and psychological assumptions, the suggested process does in fact work.

Consider now a more advanced situation, such as might develop after the hypothetical language-community has standardized on a set of numeral expressions for numbers up to 20, using one-word expressions up to 10 and a simple binary additive construction on a base of 10 from 11 to 20. This limited numeral system may persist for some time, defining a normal limit to numeral competence in this society. The individual additive expressions may get phonologically modified and eroded somewhat, so that children learning the system learn them by rote rather than learning a productive rule, even though they may later become aware of the underlying regular arithmetical basis of these expressions. How, historically, might such a system develop next, to extend the range of expressible numbers?

I assume that the 'repeated morph constraint' of Menn and MacWhinney (1984), mentioned in Section 6.3, discriminates to some extent against an expression such as **ten ten*, even though it the only way of expressing 20 in the system as described so far. This would make expressions such as *ten ten one*, *ten ten two*, and so on for 21, 22, ... feel somewhat awkward. In fact no numeral system that I am aware of expresses 21, 22, etc. as [10 + 10 + 1], [10 + 10 + 2], and so on. The awkwardness of something like *ten ten* is a barrier to expressing higher numbers. There are two obvious strategies for extending the expressive power of a language: either invent new words or invent new syntactic constructions.

Overcoming the particular awkwardness of [10 10] by inventing a new word would lead to the invention of a word for 20, presumably introduced initially as a synonym of *ten ten*. The new word could be used in new expressions generated by the existing additive rule to give [20 + 1], [20 + 2], and so on. This strategy gives rise to familiar vigesimal systems such as Celtic, Basque, Mixtec, and Yoruba. The alternative strategy of inventing a new

construction would plausibly lead to a construction interpreted by multiplication, so that 20 can be expressed as [2 × 10] or [4 × 5]. One need not assume that the prior adoption of a base of 10 in additive constructions automatically selects a base of 10 for the novel multiplicative construction. Since not all numbers are expressible as multiples of two integers, it would be natural at or soon after this stage to invent complex multiplicative/additive constructions along the lines of [[2 × 10] + 3], [[4 × 5] + 3], [[6 × 3] + 5], and so on for 23. This would give rise, as I will show, to the familiar pure decimal systems of Germanic, Romance, Semitic, and Chinese. I investigate the consequences of these two strategies, below, in particular concentrating on mechanisms by which single standardized expressions emerge from the range of arithmetically expressions.

Take first the vigesimal route, arising from the invention of a word for 20. With this invention alone, and using the existing additive rule (6.5.2), the system provides expressions for 21–30, and 40. The expressions for 21–30 could only be formed using the word for 20, given only a binary additive rule combining words. So from the time of its invention, the word for 20 has no competitor as a base-word in numerals above 20. If, by the time the word for 20 is invented, the decimal expressions for 11–19 have become lexicalized, that is can be taken as words, then it is also possible to express 31–39 using the new base-word and the old additive rule. This is in fact the usual pattern in vigesimal systems. Presumably there is the same awkwardness about [20 20] as there is for [10 10], due to the repeated morph constraint.

Now look at the alternative to the word-invention strategy, in which a multiplicative construction involving existing words is invented. In fact no language has multiplicative constructions without the possibility of embedding them in additive constructions. Presumably the desire to express a continuous sequence of numbers exerts some pressure to allow addition to a multiplicative expression, so let us assume the invention of a rule producing this combination. At this point the constructions so far developed could be expressed by the following rules:

6.5.4

$$\text{Expression} \rightarrow \left(\begin{Bmatrix} 10 \\ \text{multexpression} \end{Bmatrix} + \right) \text{word}$$

$$\text{Multexpression} \rightarrow \text{word} \times \text{word}$$

It is consistent with what diachronic evidence there is on the historical growth of numeral systems, and with the synchronic language-internal evidence of growth marks, to suppose that developing numeral systems do not run before they can walk (nor begin simultaneously to be able to run and walk). Perhaps individual speakers may be aware of the creative recursive possibilities, but it takes time for successively more complex constructions to gain acceptance in the language community at large. Thus the rules of (6.5.4) summarize a (hypothetical) social situation. If an individual speaker has internalized rules which go beyond the possibilities generated by (6.5.4), he is a potential inventor of new constructions. He may attempt to use these further possibilities from time to time, but at the risk of inconvenience or even failure in communication. A speaker who plays safe will remain within the bounds set by the rules of (6.5.4).

Expressions such as the following, with the given interpretations, now become available.

6.5.5

(a)	two three	6
	three six	18
	five four	20
(b)	[two six] one	13
	[three five] two	17
(c)	three seven	21
	seven four	28
	eight seven	56
	nine ten	90
(d)	[six three] five	23
	[five eight] eight	48
	[nine ten] nine	99

The expressions in the first two groups (6.5.5a, b) would serve no use not already served by existing standardized expressions. One would expect them not to become standardized. But the expressions in the last two groups (6.5.5c, d) express previously inexpressible numbers. The new constructions often make available more than one way of expressing a particular number, for example 24, which could be [3 × 8], [4 × 6], [[2 × 10] + 4], [[4 × 5] + 4], [[3 × 6] + 6], [[3 × 5] + 9], [[2 × 7] + 10]. One must not take the easy option and claim that it is in some sense 'obvious' that most of these are inappropriate. The problem is to find some principle which selects just the expression which becomes standardized. Clearly it is not a matter of relative simplicity of the expressions themselves, since [3 × 8] and [4 × 6] are simpler than [[2 × 10] + 4], and therefore might be expected to be preferred on grounds of economy. The question, in short, is: how does a base of 10 become standardized, with expressions using a base other than 10 being relegated to non-standardized status? Analogical pressure from the fact that 10 is at this stage already the standardized base in simple additive constructions is perhaps not sufficient to establish it as the standardized base in the new multiplicative constructions.

The explanation I propose is exactly similar to that suggested for the standardization to a base of 10 in the first simple additive constructions. When the new constructions are first invented, speakers use number words freely in them, producing a great variety of arithmetically different expressions for the same number. Speakers are, however, influenced in their choice of expression by the wish to be perspicuous, or to use an expression of a type with which the hearer is likely to be familiar. If a particular number word happens to be used more frequently than another as the higher word in a multiplicative construction, the former is preferred on grounds of probable greater familiarity to the hearer. Just as in the earlier case, for arithmetical reasons, 10 is a more useful number than lower numbers in forming expressions for higher numbers. With the rules of (6.5.4), there are more numbers expressible using a base-word of value 10 than there are expressible using any other base-word.

As with the simpler case of the standardization of 10 as a base in simple additive constructions, a computer simulation has been carried out of the hypothesized social process whereby 10 emerges as the standardized base to be used in the constructions generated by the rules of (6.5.4). This simulation is described in the next section.

To end this section, I comment on the social diachronic nature proposed for the standardization process. Recall that in Section 6.3 I rejected an explanation in terms of a psychological rule or structure common to all individual acquirers of numeral systems disposing them to acquire a system in accordance with the Packing Strategy (which essentially guarantees that the highest available word is chosen as a base). The theoretical argument there was that it is psychologically implausible to attribute to the individual the amount of computational work involved in comparing large sets of rival expressions and eliminating all but one. It may be suspected that what is now being offered as a 'social diachronic' explanation falls to a similar theoretical objection, in that the individual speaker is still attributed with the ability to carry out implausible computations, now involving frequency, in his head. But the objection in fact fails, due to the crucial difference between the two proposed explanations in terms of time-scale.

The time-scale of language acquisition is relatively rapid, the major part of it being accomplished in a few years. But the historical standardization process could conceivably take centuries to complete, and involve several generations. All that is required of the typical individual (and there may be non-conforming individuals) is that she be capable of registering, however dimly, an awareness of the relative frequency of use of number words used as bases, and of modifying her own choice of expression, however fitfully, in the direction of the perceived general preference. A particular individual may not achieve uniform usage in her lifetime, but the claim is that she probably will *tend* to move in the direction of uniform usage. Over successive generations, the standard usage of the language community will thus tend to become more uniform. It must be remembered that numerals are a special case in that no obvious communicative purpose is served by expressing a number by different arithmetical combinations, as argued in the previous section. It is this lack of motivation to preserve communicatively important distinctions which allows the convergence on uniform usage.

A connection has been proposed between frequency, which is a scalar property of items in corpora, and standardization, which is a binary, all-or-nothing, property of items regardless of the corpus in which they occur. Standardization is formally similar to grammatical well-formedness (on a common view) in being a binary property. It may be argued that this formal difference between frequency and standardization in principle prevents one being used to explain the other. Something more is needed before a strong statistical tendency in usage can become an absolute rule.

In the model proposed, individual speakers respond in a discrete all-or-nothing way to overwhelming frequency facts. Speakers do not merely adapt their own usage to mimic the frequencies

in the data they experience. Rather, they 'make a decision' to use only certain types of expressions once the frequency of those types of expression goes beyond some threshold. At a certain point there is a last straw which breaks the camel's back and speakers 'click' discretely to a decision about what for them constitutes preferred usage. What I have in mind is similar to Bally and Sechehaye's suggestion about Saussure's view of language change. 'It is only when an innovation becomes engraved in the memory through frequent repetition and enters the system that it effects a shift in the equilibrium of values and that language [langue] changes, spontaneously and *ipso facto*' (Saussure, 1966:143n). Bever and Langendoen (1971, p. 433) make the same point nicely by quoting *Hamlet*: 'For use almost can change the form of nature.' This psychological process is left completely mysterious in the present proposal. However, in the computer simulations described in the next section, various assumptions about the required threshold are made. Several of the possibilities simulated are quite strict in that, for example, they require an item to be at least twice as frequent as some other item before it is preferred to it. If an item is not at least twice as frequent as some rival item, the choice between them is taken to be random.

Clearly there can be a connection between frequency and binary linguistic properties, such as well-formedness. Corbett (1983) reviews a number of agreement patterns in the Slavic languages and proposes that two universal linguistic hierarchies, the Agreement Hierarchy and the Predicate Hierarchy, account for the relation between syntactically determined agreement and semantically determined agreement.

> The agreement hierarchy
> attributive – predicate – relative pronoun – personal pronoun
>
> In absolute terms, if semantic agreement is possible in a given position in the hierarchy, it will also be possible in all positions to the right. In relative terms, if alternative agreement forms are available in two positions, the likelihood of semantic agreement will be as great or greater in the position to the right than in that to the left. (1983, pp 10–11)

Corbett's Agreement Hierarchy – and likewise the Predicate Hierarchy, taken from Comrie (1975) – successfully predict an impressive range of facts both about absolute well-formedness and about relative frequency. Some Slavic languages reflect a hierarchy in absolute terms and others reflect it in relative terms. But it is clear that there is a single generalization concerned with a unitary phenomenon. Corbett expresses this as the hierarchies being able to apply either 'at the sentence level' (absolutely) or 'at the corpus level' (reflected in frequency). I expect that some psychological process such as I have proposed above in terms of individuals being pushed to respond discretely to frequency data mediates between the operation of the constraint at corpus level and at sentence level.

The explanation I have proposed for the phenomenon of standardization to a fixed base-number is 'social-diachronic'. One might therefore expect some historical evidence for the process I have described whereby usage becomes progressively more fixed. That is, one should be able to see cases of historical development from a less standardized to a more standardized situation. By their nature, written records tend to reflect some kind of already standardized usage, and it is not easy to find documentary evidence of variation in earlier stages of the histories of languages which at present lack such variation. Such evidence can however be found here and there. In Old English, there were expressions ***hundteontig***, ***hundendlefontig***, and ***hundtwelftig*** (Brook, 1955, p. 50) corresponding to modern playful or childish expressions **tenty* (100), **eleventy* (110) and **twelvety* (120). These forms must have competed with the forerunners of the modern standardized expressions. The following expressions from Shakespeare plays are also suggestive.

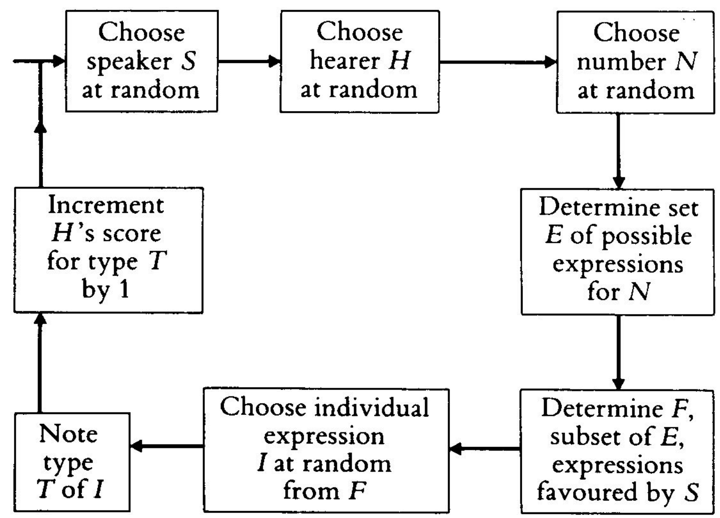
6.5.6 about the world have times twelve thirties been (*Hamlet*, III. ii.158)
one and twenty fifteens (*Henry VI, part* II, IV.vii.221)
battles thrice six (*Coriolanus*, III.iii.153)
twice six moons (*Pericles*, III.iv.31)

Obviously, Shakespeare was not necessarily concerned with accurate recording of usage, but with creating an appropriate literary impression. But these examples must bear some relation to usages which would have been familiar to Shakespeare's audiences. Such usages would not be familiar (except perhaps through Shakespeare) to modern English speakers, and we can therefore perceive some evidence of the diachronic narrowing of the range of accepted expressions.

6.6 Computer Simulations

Computer programs have been written which simulate the above processes. A simulated population of some finite number of individuals is set up. In the case of a simple lexicon {*one*, ..., *ten*} being augmented by a single rule generating simple binary additive structures, these individuals 'know' single words for each of the numbers 1-10, and they 'know' a syntactic construction combining a pair of these words to form an expression interpreted by adding the values of the constituent words. The individuals in the simulated population spend their time uttering numeral expressions to each other, and hearers keep count of the frequencies of various classes of expression addressed to them, defined in terms of base. Here 'base' simply means the highest-valued member in an additive pair. A single conversational encounter, involving one speaker uttering one numeral expression to one hearer, takes the form shown in (6.6.1) below.

6.6.1 Structure of a single encounter in the simulation



The simulation program runs repeatedly around the circuit in (6.6.1). With each encounter, the 'experience' of one individual (the 'hearer') is changed, so that his count for the expression-type (in terms of base) just addressed to him is incremented by one. This becomes relevant on a subsequent encounter when this individual happens to be chosen as 'speaker'. At intervals during the simulation, a 'census' is taken, inspecting the counts for all bases of all individuals in the population, and determining how many individuals 'favour' (see definitions below) each of the possible bases 6-10. When all the individuals in the population favour only a base of 10, or when it becomes clear that this situation is unlikely to be reached, the simulation stops.

In writing the program, care was taken to avoid building in, either explicitly or implicitly, any clause identifying a base of 10 as special in any way. The arbitrariness of the various 'random' choices made by the program was produced by a couple of random number generating procedures, one generating a list of numbers from a 'seed' taken from the time shown on the computer's clock at the time of running the program. The numbers in this first list were then used as seed numbers by a second random number generating procedure. The choices made by the program were in this sense doubly arbitrary.

Within the general structure defined by diagram (6.6.1), there are many different specific possibilities. Some of these are listed in (6.6.2).

6.6.2

(a) Size of population.
Simulations have been carried out with simulated populations of between 10 and 100 individuals. These numbers are unrealistically small for a real speech community, but larger populations make the simulation computationally unwieldy. Even with these small numbers, however, it is possible to demonstrate some quite significant results.

(b) Range of numbers expressible.
This obviously depends on the system of rules and lexical items 'known' by the community. Simulations have been carried out with the numbers 11-20 expressible by means of binary combinations of single words with values in the range 1–10. Further, more complex, sets of simulations, in which speakers know different constructions, interpreted by various combinations of addition and multiplication, with the capacity for expressing (some of the) numbers in the range 11–1000, have also been carried out.

(c) Lexicon and rules in the numeral system. (See (b) above.)

(d) Criteria by which individuals favour expressions of certain types.
A simple criterion, used in one simulation, is as follows:
If E is an expression with base B, and there is no other base B' ($B \neq B'$) more frequent than B in speaker S's experience, then S favours E.
A slightly more complex, and stronger criterion, also used in a simulation, is as follows:
If E is an expression with base B, and there is no other base-word B' ($B \neq B'$) at least twice as frequent as B in speaker S's experience, then S favours E.
Obviously, further, more or less plausible, more or less baroque, criteria can be defined.

As implied above, a number of trial simulations have been run along the lines indicated. A number of suggestive results emerge. I give a sample of these below.

In a simulation involving a population of 100 individuals, who know lexical items for the numbers 1–10, and have a rule forming binary combinations of these, interpreted by addition, a record of the first few encounters in the simulated interaction between these individuals looks as follows:

6.6.3

Speaker 23	to	Hearer 73	19	expressed as	[10 + 9]	(encounter 1)
Speaker 29	to	Hearer 59	14	expressed as	[10 + 4]	(encounter 2)
Speaker 16	to	Hearer 41	18	expressed as	[9 + 9]	(encounter 3)
Speaker 31	to	Hearer 73	14	expressed as	[9 + 5]	(encounter 4)
Speaker 34	to	Hearer 29	16	expressed as	[8 + 8]	(encounter 5)
Speaker 67	to	Hearer 97	19	expressed as	[10 + 9]	(encounter 6)
Speaker 71	to	Hearer 5	19	expressed as	[10 + 9]	(encounter 7)
Speaker 77	to	Hearer 98	11	expressed as	[8 + 3]	(encounter 8)
Speaker 82	to	Hearer 14	20	expressed as	[10 + 10]	(encounter 9)
Speaker 61	to	Hearer 87	12	expressed as	[10 + 2]	(encounter 10)

(Individuals in the community are represented by numbers 1–100.) At this beginning stage, varied usage is evident; there are two uses of 8 as a base (encounters 5, 8), two uses of 9 as a base (encounters 3, 4) and six uses of 10 as a base. Of the six uses of base-10, four were arithmetically necessary, in that speaker had in these encounters to express either 19 or 20. One thousand nine hundred and fifty encounters later, in this particular simulation, a record of the exchanges looked as follows:

6.6.4

Speaker 72	to	Hearer 81	12	expressed	as	[10+2]	(encounter 1952)
Speaker 95	to	Hearer 5	11	expressed	as	[10+5]	(encounter 1952)
Speaker 81	to	Hearer 43	15	expressed	as	[10+5]	(encounter 1953)
Speaker 93	to	Hearer 73	15	expressed	as	[10+5]	(encounter 1954)
Speaker 27	to	Hearer 34	16	expressed	as	[10+6]	(encounter 1955)
Speaker 26	to	Hearer 58	13	expressed	as	[10+3]	(encounter 1956)
Speaker 86	to	Hearer 58	13	expressed	as	[10+3]	(encounter 1957)
Speaker 3	to	Hearer 53	20	expressed	as	[10+10]	(encounter 1958)
Speaker 8	to	Hearer 37	13	expressed	as	[9+4]	(encounter 1959)
Speaker 64	to	Hearer 1	19	expressed	as	[10+9]	(encounter 1960)

It is evident that the situation has firmed up considerably, with an almost uniform usage of a base of 10. In the one case where a base other than 10 is used, this must be due to speaker 8 favouring, on the basis of expressions addressed to him, a base of 9. The definition of 'favour' used for this simulation allows several (indeed all) possible bases to be favoured, if no single base is sufficiently more frequent than the others. The definition used is: a base *B* is favoured by an individual *I* unless *I*'s count for some other base is at least twice as great as his count for *B*. For example, given the following counts by individual *I*:

6.6.5

Base	Frequency
6	3
7	5
8	7
9	11
10	13

this individual favours the bases 8, 9, and 10. If the frequency of base-10 in this example were to be increased to 22, leaving all the other frequencies the same, then individual *I* would favour base-10 alone, according to the particular criterion adopted in this simulation. Many simulations in fact terminate with all individuals in the population actually favouring base-10 alone (see details below).

At any point before the termination of a simulation, individuals will vary in what bases they favour, depending on their 'experience'. Given below are some static snapshots of the situation at various stages in a simulation. Here, a distinction is made between arithmetically possible expressions, which are the same for the whole community, since they all know the same lexical items and syntactic rule, and a particular individual's favoured expressions, which depend on which base(s) that individual favours at a given stage in the simulation. In the table below the arithmetically possible expressions for various numbers are listed, and those which are not members of an individual's favoured subset of these are marked with an asterisk.

6.6.6

(a) Starting situation. Same for whole population.

Number	Possible expressions
11	[10 + 1], [9 + 2], [8 + 3], [7 + 4], [6 + 5]
15	[10 + 5], [9 + 6], [8 + 7]
18	[10 + 8], [9 + 9]
20	[10 + 10]

(b) Mid-simulation situation. Some favour base-10 only, some base-9 only, and some base-10 and base-9. (Nobody favours base-8, base-7, base-6.)

(i) Those individuals favouring base-10 only

Number	Possible expressions
11	[10 + 1], *[9 + 2], *[8 + 3], *[7 + 4], *[6 + 5]
15	[10 + 5], *[9 + 6], *[8 + 7]
18	[10 + 8], *[9 + 9]
20	[10 + 10]

(ii) Those individuals favouring base-10 and base-9 equally, and no other base number

Number	Possible expressions
11	[10 + 1], [9 + 2], *[8 + 3], *[7 + 4], *[6 + 5]
15	[10 + 5], [9 + 6], *[8 + 7]
18	[10 + 8], [9 + 9]
20	[10 + 10]

(iii) Those individuals favouring base-9 only

Number	Possible expressions
11	*[10 + 1], [9 + 2], *[8 + 3], *[7 + 4], *[6 + 5]
15	*[10 + 5], [9 + 6], *[8 + 7]
18	*[10 + 8], [9 + 9]
20	[10 + 10]

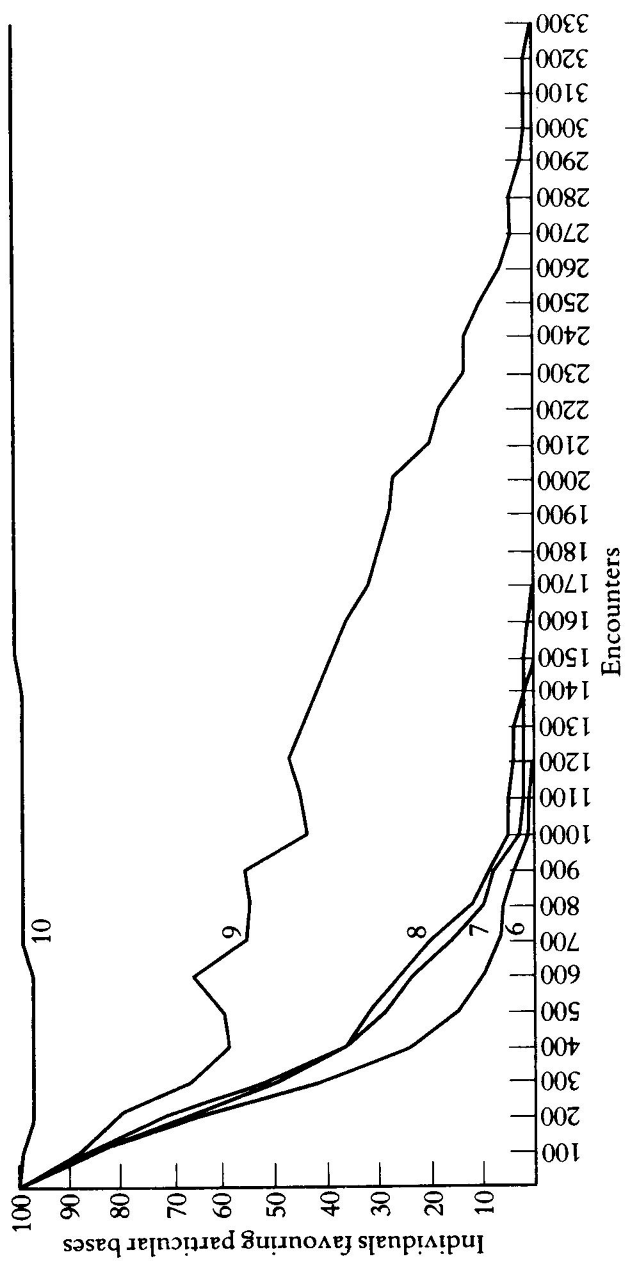
Note that, even if an individual favours base-9, he must still use base-10 in expressing the numbers 19 and 20, since the numeral system possessed by the community at this stage in its evolution provides no expressions for these numbers with base-9.

For this condition, with a simple lexicon *one*, . . . , *ten* augmented by a rule generating simple binary additive constructions, five separate series of test simulations were carried out, under slightly varying conditions. These are summarized in (6.6.7), on the next page. A comparison of the second, third, and fourth columns in this table shows that in each series of tests a single parameter from a preceding series was altered. In series B, for instance, the criterion by which speakers favour particular bases was made stricter than in series A. In series C, all individuals in the population were initially 'credited' with the experience of having heard two instances of each base, rather than starting from zero; this was an attempt to dampen down unnaturally volatile change early in a simulation, due to the fact that any positive integer is at least twice zero. In series D, the conditions of series C were repeated, but with a slightly modified random number generating procedure, to test whether a certain odd/even subregularity in the numbers generated was prejudicing the results; theoretically, the slightly modified random procedure generates series of numbers which are 'more random', in that they show no obvious patterning or repetition. In series E, the population was expanded from 25 to 100 individuals, to make the simulations approximate more closely to reality; the unmodified random number generating procedure was used.

6.6.7

Series label	Population	Starting situation	'Favour' criterion	Tests done	Tests converging
A	25	All scores zero	*I* favours most frequent *B*(s)	9	9
B	25	All scores zero	*I* favours *B* unless some other *B* at least twice as frequent	26	22
C	25	All scores 2	Ditto	24	19
D	25	All scores 2	Ditto	20	20 (slightly modified random number generator)
E	100	All scores 2	Ditto	18	18

The right-hand column in (6.6.7) indicates the numbers of tests in each series which converged on base-10, that is tests which terminated with all individuals favouring a base of 10 and no other base. As can be seen, in three series (A, D, E) all tests converged in this way, while in two series (B, C) roughly one-sixth of the tests did not so converge. The essential randomness of the simulations ensures that there is always a possibility that a simulation will not converge on any particular base, or will converge on a low-valued base, such as 6. The simulations indicate that such possibilities have low probability, and that by far the most probable situation to emerge from the conditions set up is one in which all individuals exclusively favour expressions with base-10.



6.6.8

Figure (6.6.8) is a graph showing how in a typical test from Series E the number of individuals favouring particular bases changed during the simulation. At the outset of the simulation, all individuals favour all bases. Thereafter, the numbers of individuals favouring bases 6, 7, 8, and 9 decline fairly steeply, with only a slight decrease in the number favouring base-10 (down to a nadir of 97%). Soon, however, the number favouring base-10, is restored to 100%, while bases 6, 7, 8, 9 decline steadily in favour, with the lower-valued declining faster than the higher-valued ones. From about halfway through the simulation, there is a long 'tail', in which a dwindling number of individuals still favour base-9 as well as base-10. By the 3300th encounter, the last such individual has been bombarded with so many expressions with base-10, and so few with base-9, that finally he gives up his allegiance to base-9, and the simulation terminates. For the record, summary details are given in (6.6.9) of the results in all test series for the condition with a simple lexicon {*one*, ..., *ten*} augmented by a rule generating simple binary additive structures.

6.6.9

Series	Done	Conver-ging	Mean: encounters to con-vergence	Fastest converger	Slowest Converger	Standard deviation: encounters to con-vergence
A	9	9(100%)	166	80	330	75
B	26	22(85%)	617	150	1750	469
C	24	19(79%)	621	250	1400	361
D	20	20(100%)	670	400	1400	285
E	18	18(100%)	6744	1400	48000	10802

These statistical details are not actually of any great importance as far as the central point of this proposal is concerned. The central point is that, given the conditions set up in the simulations, in the vast majority of cases the simulated individuals end up all exclusively favouring expressions with base-10. This result is not preprogrammed into the simulations, but arises from the interaction of arithmetical properties of the numbers involved and some very simple assumptions about social interaction and the way speakers choose familiar forms of expression. The general tenor of the results obtained here indicates that there may be a plausible explanation in terms of social interaction for the general tendency of standardized numeral expressions to involve binary combinations of maximally different values. Now that the principle is established for relatively simple expressions in the range 11–20, the door is open for more complex simulations, extending the proposal to higher numbers. Some further work of this kind is presented below.

A further batch of simulations was carried out, simulating developments after the stage reached in the previous experiments. In this new batch, the individual members of the population are credited with knowing a lexicon of number words from *one* to *ten*, and binary additive expressions on a base of *ten* up to 20. This is the standardized situation inherited from the hypothesized previous stage in development; there is a single standardized expression for each number up to 20. Further, speakers now know a binary multiplicative rule combining any two number words, and a more complex additive construction in which a single number word is placed in an additive relationship with a multiplicative expression. In other words, speakers now know rules as in (6.5.4) of the previous section. This gives the capability of expressing numbers up to 110 ([[10 x 10] + 10]), but, so far, there are no preferences for any particular arithmetical combinations in expressions for numbers above 20. Thus the possibilities include, but are by no means exhausted by, the following:

6.6.10

21 [3 × 7], [[2 × 10] + 1], [[2 × 9] + 3], [[3 × 4] + 9]
45 [5 × 9], [[4 × 10] + 5], [[6 × 7] + 3], [[6 × 6] + 9]
71 [[7 × 10] + 1], [[8 × 8] + 7], [[7 × 9] + 8]
99 [[9 × 10] + 9]

At the beginning of a simulation, all such expressions are equally possible for all speakers, except that a simpler expression is preferred over a more complex one, all other things being equal. So, for instance, [3 × 7] would actually be preferred over [[2 × 10] + 1] on grounds of simplicity, since at this stage no preference for any particular base elevates the latter in preference to the former. The base of a binary expression is defined as its highest-valued member, and the base of a complex expression as the base of the embedded binary expression. Thus 7 would be the base in both [3 × 7] and [[4 × 7] + 3].

At the beginning, all other things are indeed equal, in that all speakers prefer all bases equally (for numbers above 20). As the simulation proceeds, speakers keep track, as before, of the frequency with which they have heard particular bases, and when the frequency of one base drops below some specified threshold in relation to other bases, that base ceases to be preferred by that speaker. The situation is in principle reversible, in that a run of expressions with the 'demoted' base addressed to that speaker could restore it to the preferred set.

In simulating a speaker's choice of an expression for a number, the program first finds the set of expressions for that number using any of that speaker's preferred bases. For example, if a speaker's preferred bases are 10, 9, 8, and 7, then the set from which he chooses an expression for 21 is [3 × 7], [[2 × 10] + 1], [[2 × 9] + 3], [[2 × 8] + 5], [[2 × 7] + 7]. Other expressions, such as [[3 × 6] + 3] and [[4 × 5] + 1] are not among the possibilities, since their bases (6 and 5) are not in the preferred set of bases. The program next selects the structurally simplest subset of possible expressions, in this case the single expression [3 × 7]. If several expressions remain after this selection for simplicity, for example [4 × 6] and [3 × 8] as (equally simple) expressions for 24, an expression is chosen at random from this set. If the speaker in the above example were to cease to prefer base-7, then the set from which he chooses an expression for 21 would be reduced to [[2 × 10] + 1], [[2 × 9] +3], [[2 × 8] + 5]. One of these would be chosen at random. Note that the structurally simpler [3 × 7] would be eliminated in this way.

As before, various series of simulations were carried out, systematically varying the initial conditions and the criterion for preferring bases. The conditions for these series are summarized in (6.6.11).

6.6.11

Series label	Population	Starting situation	'Favour' criterion	Tests done	Tests converging
F	25	All scores zero	*I* favours most frequent *B*(s)	26	26
G	25	All scores zero	*I* favours *B* unless some other *B* more than twice as frequent	26	26
H	25	All scores 2	Ditto	26	26
I	100	All scores 2	Ditto	26	26

The conditions in each series are successively more demanding tests of the hypothesis that convergence on the highest available base number arises from the simulated social interaction. Thus in series F, the criterion for preference of a particular base is simply that it be used more frequently than others: a base is preferred unless some other base is more frequent in the speaker's experience. In series G the required criterion is stronger: a base is preferred unless some other base is more than twice as frequent. This has the effect of keeping bases in the preferred set for longer. In series H, speakers are initially 'credited' with having heard two instances of each base. This is an attempt to damp down any early advantage that any particular base might gain in the very first stages of a simulation. Finally, in series I, the simulated population is increased from 25 to 100 individuals, bringing the simulation closer to a real social situation. All simulations converged on base-10. Some statistical details are given for information in (6.6.12).

6.6.12

Series	Done	Converging	Mean: encounters to convergence	Fastest converger	Slowest Converger	Standard deviation: encounters to convergence
F	26	26(100%)	202	125	350	60
G	26	26(100%)	327	175	650	126
H	25	25(100%)	721	425	975	156
I	26	26(100%)	4123	2800	6600	942

A sample printout of results from a typical simulation in series G is given in (6.6.13). For this simulation the population was 25 and convergence on base-10 was achieved by the 325th encounter.

6.6.13

	Individuals favouring base-									
Encounters	1	2	3	4	5	6	7	8	9	10
25	11	11	11	11	11	11	13	13	14	23
50	5	5	5	5	7	5	8	8	11	21
75	3	3	3	3	4	3	7	5	7	23
100	2	2	2	2	3	2	7	4	5	24
125	0	0	0	0	1	0	3	2	2	25
150	0	0	0	0	1	0	5	2	3	25
175	0	0	0	0	1	0	3	1	3	25
200	0	0	0	0	0	0	3	1	1	25
225	0	0	0	0	0	0	3	1	1	25
250	0	0	0	0	0	0	1	0	0	25
275	0	0	0	0	0	0	1	0	0	25
300	0	0	0	0	0	0	1	0	0	25
325	0	0	0	0	0	0	0	0	0	25

A record of a sequence of encounters from the early stages of a simulation in series I is given below in (6.6.14). Note that at this stage, the situation has not become standardized to a single base, and hence a variety of types of expression is found.

6.6.14

Speaker 70 to Hearer 11 33 expressed as [[4 × 8] + 1
Speaker 38 to Hearer 77 20 expressed as [2 × 10]
Speaker 49 to Hearer 80 52 expressed as [[5 × 9] + 7]
Speaker 43 to Hearer 93 77 expressed as [[8 × 9] + 5]
Speaker 88 to Hearer 1 26 expressed as [[2 × 9] + 8]
Speaker 90 to Hearer 2 21 expressed as [3 × 7]
Speaker 46 to Hearer 58 38 expressed as [[4 × 8] + 6]
Speaker 87 to Hearer 56 17 expressed as [10 + 7]
Speaker 1 to Hearer 58 32 expressed as [[3 × 10] + 2]
Speaker 49 to Hearer 41 91 expressed as [[9 × 10] + 1]

The simulations so far reported correspond to the first two stages in the standardization of a decimal system hypothesized in the previous section. Clearly one could carry on, running further simulations of more advanced and more complex situations. I believe the plausibility in principle of the social/diachronic hypothesis on the standardization of complex numerals to the highest available base number is established by the simulations described above, and I have not explored further with as much thoroughness. But one further set of simulations is worth reporting on.

In the starting situation for this last set of simulations, individuals know all the same numeral words and constructions as in Series F-I above, and in addition, they permit constructions in which a binary expression, either additive or multiplicative is in either an additive or a multiplicative relationship with a number word. The available rules may be summarized as in (6.6.15)

6.6.15

$$\text{Expression} \rightarrow \begin{Bmatrix} \text{binexpression } (+ \text{ word}) \\ \text{word } (\times \text{ binexpression}) \end{Bmatrix}$$

$$\text{Binexpression} \rightarrow \begin{Bmatrix} \text{word} \times \text{word} \\ 10 + \text{word} \end{Bmatrix}$$

This simulates a situation in which a wider range of new constructions than so far considered 'comes on the market' simultaneously, making a greater range of expressions possible. Expressions now possible for 48, for example, include [[3 × [10 + 6]], [4 × [10 + 2]], [4 × [2 × 6]], [3 × [4 × 4]]. In these expressions, a complex numeral is in a multiplicative relationship with a lower-valued numeral word; the complex numeral thus in some sense serves as a multiplicative base. Such situations are in fact very rare in languages, which almost always use single words as multiplicative bases (for example English *hundred*, *thousand*, *million*). (See *LTN* pp. 225–32, 239–43 for discussion of rare counterinstances from Yoruba, and a related phenomenon in Ainu.)

To simulate standardization to one expression per number in a situation such as this, one needs to consider not only preferences of base number, but also preferences regarding type of embedded expression, i.e. whether additive or multiplicative. Merely adopting a preferred base of 10 will not discriminate between [[4 × 10] + 8] and [3 × [10 + 6]], since both of these expressions have a base of 10, as 'base' has been defined here. But if one can have a preference for multiplicative expressions in embedded position over additive expressions, then [3 x [10 + 6]] will be eliminated in favour of [[4 × 10] + 8], which is indeed the standard way of expressing 48 in typical decimal systems.

Series J of simulations worked in essentially the same way as the others described, but speakers also recorded the frequency of the types (that is additive or multiplicative) of the embedded expressions in complex expressions. The type (additive or multiplicative) of a binary expression embedded in a complex expression was labelled the 'headtype' of the complex expression. Speakers counted the bases and the headtypes of expressions addressed to them, and preferred particular bases and headtype on the basis of their relative frequency. The details of this series of simulations are given below.

6.6.16 Series J
Population: 25
Starting situation: all scores zero
Preference criterion for bases: individual favours most frequent base(s)
Preference criterion for headtypes: individual favours most frequent headtype(s)
Tests done: 26
Tests converging on base-10: 26 (= 100%)
Tests converging on multiplicative headtype: 26 (= 100%)
Mean encounters to convergence: 654
Fastest converger: 200
Slowest converger: 2325
Standard deviation – encounters to convergence: 494

Here one sees the emergence, from an initial wide range of possibilities, of a standardized system for expressions up to 100 in which the multiplicative base is always a single word, rather than a more complex expression. Just as, in a clear arithmetical sense, the highest-valued numeral word available in a system is also the most useful, and gets standardized as the only permitted base-word, similarly, for expressions up to 100, with rules as in (6.6.15) a binary multiplicative construction is more useful than an additive one as a basis for more complex expressions. This is simply because multiplication allows one to reach further up the number series, given limited lexical resources. Adding to a binary multiplicative expression allows one to express prime numbers, which could not be expressed by multiplication alone. But there is no arithmetical advantage, up to a certain limit, in embedding a multiplicative expression inside another multiplicative construction, since (up to the limit) a simpler multiplicative expression will always be available. Thus, for example, [[3 × 10] × 3], though it has the preferred multiplicative headtype and the preferred base of 10, is less preferable than [9 × 10] which also has the preferred base and is simpler. The limit mentioned here, after which multiplicative expressions embedded as higher-valued members of a multiplicative pair suffer no competition from simpler binary multiplicative expressions, is 100 (given a lexicon with words up to 10).

Up to 100, the preferences for base-10 and multiplicative headtype, together with the criterion (*ceteris paribus*) of simplicity, interact to produce just one standardized expression for each number. The expression for 20 remains [10 + 10], and 30, 40, . . . , 90 are expressed as multiples of 10. The expression for 20 is thus out of line with those for other multiples of 10. This in fact happens in some decimal systems, for example Arabic, where the relevant expressions are:

6.6.17 20 ʕiʃriin [10 *-iin*]
30 talatiin [3 *-iin*]
40 arbaʕiin [4 *-iin*]
⋮
90 tisaʕiin [9 *-iin*]

If the morpheme *-iin* is to be given a constant meaning, it must be 10. Thus 20 is [10 + 10], but 30, . . . , 90 are [3 × 10], . . . , [9 × 10].

In view of the awkwardness of binary combinations of the same word, due to the repeated morph constraint, mentioned earlier, there would be some pressure to abandon [10 + 10] as the expression for 20 and to adopt [2 × 10] on the analogy of the expressions for 30 , . . . , 90. This revision of the expression for 20 could happen simultaneously with the slow social process by which other numbers up to 100 acquire single standardized expressions. In many decimal systems the expression for 20 is quite analogous to those for 30, . . ., 90. But many other systems show signs of stress just at the word for 20. Thus in Germanic languages (for example English, German) the value 2 shows through in the expressions for 20 (*twenty*, *zwanzig*) much less transparently than the values 3, . . ., 9 in the expressions for 30, . . ., 90. In Romance languages (for example Italian, French) the value 2 is not synchronically transparent at all in the expressions for 20 (for example *venti*, *vingt*), whereas the values 3, . . ., 9 are (just about) transparent in the expressions for 30, . . ., 90 (30, . . ., 60 in French).

From 100 to 110, the preferences for base-10 and multiplicative headtype do yield single standardized expressions, for example [[10 × 10] + 5], but these all contain the awkward [10 × 10], and so might be expected to be somewhat unstable. And after 110, the rules in (6.6.15) fail to provide any expression at all for many numbers, for example 111, 113, and only expressions with non-preferred bases and/or non-preferred headtypes for many others, for example 112 (expressible as [[2 × 7] × 8], [[4 x 4] × 7], [[10 + 4] × 8], or [[10 + 6] × 7]). Thus 100 provides a very natural pausing point in the development of a standardized system. The usual response to the difficulties after this point is of course to invent a new word, for 100. Then the whole process of negotiating standardized expressions can begin again for numbers above 100, presumably along the same kind of lines as the earlier processes. It seems likely that at this point more socially formal processes will begin to play a part, such as perhaps some kind of explicit agreement within sections of the community who frequently deal with numbers (for example traders, scribes) and deliberate instruction. Any such socially formal interference can be expected to regard the characteristics of the naturally evolved system up to 100 as providing a natural pattern for further developments.

7

Denouement and Prospect

To our language may be with great justness applied the observation of Quintilian, that speech was not formed by an analogy sent from heaven. It did not descend to us in a state of uniformity and perfection, but was produced by necessity and enlarged by accident, and is therefore composed of dissimilar parts, thrown together by negligence, by affectation, by learning, or by ignorance. (Dr Johnson, 1747, p. 17)

The study of nature used to be called 'Natural History' (and sometimes still is). The word 'history', alluding to diachrony in the name of a subject whose focus is on synchronic states of nature, may seem puzzling. But perhaps it springs from an awareness that synchronic states of nature are what they are as a result of historical processes. Linguistics is a field within the study of nature. The Saussurean dichotomy into diachronic and synchronic has no doubt had its usefulness, but the dogmatic insistence on its fundamental status has had the effect of compartmentalizing the subject so that a theoretical framework for expressing connections between diachrony and synchrony is largely missing. Langacker expresses very acutely the lack of such a framework, which he notes in connection with a particular French construction.

the actual or potential problems facing the derivation of *qu'est-ce que* etc. from cleft sentences are of two sorts. First, *qu'est-ce que*-type questions may not have quite the same semantic value as cleft sentences. Second, these interrogative formulas lack syntactic flexibility; they are restricted to

present tense and cannot be negated, but declarative cleft sentences are not so restricted. Given the usual assumptions and forms of argumentation characteristic of generative syntax, these properties count as evidence against the analysis, however weak the argument may be. But I am not sure that the argument has any force to it at all, primarily because I believe we may be dealing with a widespread and perhaps crucially important phenomenon that contemporary linguistic theory does not adequately handle, or even clearly recognize.

Semantic specialization and lack of syntactic flexibility are both characteristic of complex lexical items, such as nominalizations and idioms. To me this suggests that *qu'est-ce que* and the other interrogative formulas have frozen to some degree into fixed 'lexical' patterns. (1971, p. 29)

Since 1971, theories of the lexicon have made more room than Langacker envisaged for the kind of semi-productivity he was concerned with. But in so far as such developments have been expressed within a purely synchronic framework, I believe the heart of the problem has been overlooked. Note that Langacker resorts to the natural assumption that some kind of partially completed diachronic process ('freezing') is the cause of the synchronist's dilemma. But unless a single overarching synchronic/diachronic theoretical framework for language is adopted, there is no room in theory for diachronic explanations of synchronic *états de langue*. Langacker continues:

I would like to call attention to an important but hitherto hardly recognized phenomenon that might be called 'syntactic metaphors' or 'syntactic idioms'. Syntactic idioms are fully productive syntactic constructions that become 'lexicalized' in the sense that they take on special semantic significance, sometimes with concomitant syntactic peculiarities. (p. 31)

Numeral constructions in all languages tend to be such syntactic idioms. They are of a slightly different kind from those discussed by Langacker, in that the semantics of numeral constructions, typically involving addition and multiplication, is straightforward and in no way idiosyncratic; the idiosyncrasies of numeral constructions, requiring (in a synchronic description) the postulation of *ad hoc* rules which apply only to numerals, are chiefly syntactic. The syntax of numeral constructions is partly frozen, or fossilized. The analytic linguist can see the original regularities through the haze of distortions produced by historical change. But to children learning numeral systems, and to adults who have mastered them, the regularities are probably often not present to the mind. Bright considers the very irregular Hindi numeral series up to 100. He asks:

> Is memorization the only factor involved in the learning and production of the paradigm up to 'one hundred'? If so, should a grammar, for the sake of psychological realism, simply list these hundred forms (as, in fact, practical grammars do), with no attempt to state general rules governing their phonological shapes? To put the matter in other terms, should we regard all the forms from '11' to '99' as suppletive? (1969, p. 30–1)

Bright goes to the trouble of providing a complete set of rules for the Hindi numerals up to 100 which results in

> 184 items which represent phonological shapes or specific environments. If, on the other hand we give a simple list of the phonological shapes of forms – in effect, an *ad hoc* rule for each form – then, of course, there would be just 100 items, with a clear advantage in economy. But we have no guarantee that economy in rules is a simple or unique reflection of psychological reality. (p. 40)

Bright considers no psychological evidence to answer his question. Fuson, et al. (1982) give ample evidence that children learn the syntactically complex numerals first of all as unanalysed units, only later becoming aware of their internal structure and the systematicity of their semantic interpretations.

A developmental sequence has been traced in the preceding chapters, from the complete lack of numerals, through a system expressing cardinalities up to 2 or 3, perhaps by grammatical number, through a system with a short counting sequence of single words, typically up to 5 or 10, and finally to a system with recursive layerings of syntactic structure, expressing addition and multiplication. At each stage in this development, the

synchronic system is such that it can be acquired by new speakers without too much difficulty. But in the case of numerals, salient aspects of the shape of the system are not wholly reimposed by each fresh generation of acquirers on the basis of innate dispositions to internalize rules and representations of a narrowly constrained type. Clearly, there *are* innate contributions, most importantly:

7.1 The concepts of collection and individual object, and the relations between them.

7.2 The ability to represent arbitrary links between signified and signifier (the Saussurean Sign).

7.3 The disposition to make the sizeable inductive leap from a memorized sequence of words to the use of these words expressing the cardinality of collections (the Cardinality Principle).

7.4 The ability to acquire and control syntactic rules forming longer expressions out of the simple vocabulary, together with associated semantic interpretation rules.

7.5 The ability to assemble such rules into highly recursive rule sets.

Of these capacities, only the third, the Cardinality Principle, is special to numeral systems; the rest are very familiar in human language more generally. At least some of these capacities are probably unique to humans.

With respect to number and its expression in language (that is numerals), I claim to have shown that these innate capacities are sufficient to determine the number faculty in Man, but insufficient to determine the universal morphosyntactic peculiarities found in the human linguistic systems that express number. Man has the capacity for language and for number, capacities which his ancestors at some stage lacked. Children, born with the capacity to acquire language and number, acquire them simultaneously, and this simultaneity is significant. Language is the mental tool by which we exercise control over numbers. Without language, no numeracy. This is presumably not a logically necessary fact, as we can imagine languageless creatures who calculate proficiently with largish numbers, say to keep track of their numerous young, and perhaps using tallies, but never in any sense expressing or communicating publicly the objects of their calculations (numbers). But the dependence of developed numeracy on the possession of language, a public signalling system, is universally observed in humans. The resources of the public system are used even for completely private calculations, as when people are doing sums in their heads. Sums beyond the most elementary cannot normally be done without some mental linguistic rehearsal of the propositions involved. The capacity to reason about particular numbers, above about 3, comes to humans only with language.

If the human number faculty itself is largely a by-product of innate linguistic capacities, the linguistic subsystems dealing with number are shaped by further principles, which are not innate in individuals. The two main such principles are:

> *7.6* Languages and their subsystems grow gradually over time. Their structures exhibit traces of this growth in the form of discontinuities and irregularities.

> 7.7 Pragmatic factors make certain forms favoured for communication and such pragmatic preferences become grammaticalized, that is regarded by new acquirers as having the status of grammatical rules.

These principles affect the shape of all languages. What follows from them is a research programme to determine in detail the precise ways in which they work, a programme complementary to that which attempts to determine the contribution made to the shape of languages from innate characteristics of individuals.

> linguistic structure and evolution are a joint function of the various systems for the use of language. Attempts to explain language universals as a formal function of just one of these systems are doomed to incompleteness whether the system considered is that of speech perception, production, or the grammatical prediction of new sentences. (Bever and Langendoen, 1971, p. 455)

Languages are artefacts resulting from the interplay of many factors.

References

Aitchison, J. (in press) Other Keyholes: Language Universals from a Pidgin-creole Viewpoint. In Modgil and Modgil (in press).

Allan, K. 1977: Classifiers. *Language*, 53, 258–311.

Altham, J. E. J. 1971: *The Logic of Plurality*. Methuen, London.

Andersen, H. 1973: Abductive and Deductive Change. *Language*, 40, 765–93.

Antell, S. E. and Keating, D. P. 1983: Perception of Numerical Invariance in Neonates. *Child Development*, 54, 695–701.

Armstrong, D. M. 1978: *A Theory of Universals: Universals and Scientific Realism*, Vol. II. Cambridge University Press, Cambridge.

Austin, P. 1981: *A Grammar of Diyari, South Australia*. Cambridge University Press, Cambridge.

Bagge, L. 1906: The Early Numerals. *The Classical Review*. 20, 259–67.

Barker, M. A. R. 1964: *Klamath Grammar*. University of California Press, Berkeley.

Barker, S. F. 1964: *Philosophy of Mathematics*. Prentice-Hall, Englewood Cliffs, New Jersey.

Bartsch, R. 1973: The Semantics and Syntax of Number and Numbers, In Kimball, *Syntax and Semantics*, Vol. 2, pp. 51–93.

Barwise, J. and Cooper, R. 1981: Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*, 4, 159–219.

Bäuerle, R., Schwarze, C. and von Stechow, A. (eds) 1983: *Meaning, Use, and Interpretation of Language*. Walter de Gruyter, Berlin.

Beeston, A. F. L. 1970: *The Arabic Language Today*. Hutchinson, London.

Benacerraf, P. 1965: What numbers could not be. *Philosophical Review*, 74, 47–73.

Bever, T. G. and Langendoen, D. T. 1971: A Dynamic Model of the Evolution of Language. *Linguistic Inquiry*, 2, 433–463.

Bickerton, D. 1981: *Roots of Language*. Karoma Publishers, Ann Arbor, Mich.

Blackburn, S. 1984: *Spreading the Word*. Oxford University Press, Oxford.
Bloomfield, L. 1933: *Language*. Holt, Rinehart and Winston, New York.
Boas, F. 1939: *The Mind of Primitive Man*. Macmillan, New York.
Bostock, D. 1974: *Logic and Arithmetic*. Oxford University Press, Oxford.
Bower, T. G. R. 1974: *Development in Infancy*. Freeman, San Francisco.
Brainerd, C. J. 1973: Mathematical and Behavioral Foundations of Number. *Journal of General Psychology*, 88, 221–281.
(ed.) 1982: *Children's Logical and Mathematical Cognition*. Springer, New York.
Bright, W. 1969: Hindi Numerals. *Working Papers in Linguistics*, Department of Linguistics, University of Hawaii, Honolulu, No. 9, pp. 29–47.
Brook, G. L. 1955: *An Introduction to Old English*. Manchester University Press, Manchester.
Brouwer, L. E. J. 1913–14: Intuitionism and Formalism. *Bulletin of the American Mathematical Society* 20, 81–96.
Brown, R. 1958: *Words and Things*. The Free Press, New York.
Brunot, F. 1906: *Histoire de la Langue Française des Origines à 1900, Tome II: Le Seizième Siècle*, Librairie Armand Colin, Paris.
Bunt, H. C. 1985: *Mass Terms and Model-theoretic Semantics*. Cambridge University Press, Cambridge.
Burling, R. 1965: How to Choose a Burmese Numeral Classifier. In Spiro, *Context and Meaning*, pp. 243–264.
Butterworth, B. Comrie, B. and Dahl, O. (eds) 1984: *Explanations for Language Universals*. Mouton, Berlin.
Carlson, G.N. 1977: *Reference to Kinds in English*. Graduate Linguistic Student Association, Department of Linguistics, University of Massachusetts, Amherst, Mass.
Chomsky, N. 1957: *Syntactic Structures*. Mouton, The Hague.
1965: *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.
1968: *Language and Mind*. Harcourt, Brace, and World, New York.
1976: *Reflections on Language*. Fontana/Collins.
1980a: *Rules and Representations*. Basil Blackwell, Oxford.
1980b: Contributions to Piattelli–Palmarini (ed.) (1980).
1981: *Lectures on Government and Binding*. Foris Publications, Dordrecht, Holland.
1982: *The Generative Enterprise: a Discussion with Riny Huybregts and Henk van Riemsdijk*. Foris Publications, Dordrecht, Holland.
1986: *Knowledge of Language: its Nature, Origin, and Use*. Praeger Publishers, New York.
Chomsky, N. and Halle, M. 1968: *The Sound Pattern of English*. Harper and Row, New York.

Chomsky, N. and Lasnik, H. 1977: Filters and Control. *Linguistic Inquiry*, 8, 425–504.
Clark, E. 1979: Building a Vocabulary: words for objects, actions and relations. In Fletcher and Garman (1979), pp. 149–160.
Clark, H. and Clark, E. 1977: *Psychology and Language: an Introduction to Psycholinguistics*. Harcourt, Brace, Jovanovich, New York.
Clocksin, W. and Mellish, C. 1981: *Programming in Prolog*. Springer, Heidelberg.
Cohen, R.S. and Wartofsky, M.W. (eds) 1983: *Language, Logic, and Method. Boston Studies in the Philosophy of Science*, Vol. 31. D. Reidel, Dordrecht.
Comrie, B. 1975: Polite Plurals and Predicate Agreement. *Language*, 51, 406–18.
1984: Form and Function in Explaining Language Universals. In Butterworth, et al. (1984), pp. 87–103.
Conant, L. L. 1923: *The Number Concept: its Origin and Development*. Macmillan, New York.
Corbett, G. 1978a: Universals in the Syntax of Cardinal Numerals. *Lingua*, 46, 355–68.
1978b: Numerous Squishes and Squishy Numerals in Slavonic. *International Review of Slavic Linguistics*, 3, 43–73.
1983: *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. Croom Helm, London.
Corder, S. P. 1981: *Error Analysis and Interlanguage*. Oxford University Press, Oxford.
Corstius, B. (ed.) 1968: *Grammars for Number Names, Foundations of Language Supplementary Series*, Vol. 7, Reidel, Dordrecht.
Craig, C. (ed.) 1986: *Noun Classes and Categorization*. John Benjamins, Amsterdam.
Curtiss, S. 1977: *Genie: A Psycholinguistic Study of a Modern-day Wild Child*. Academic Press, New York.
Dantzig, T. 1940: *Number: the Language of Science*. George Allen and Unwin, London.
Davidson, D. and Harman, G. (eds) 1972: *Semantics of Natural Language*. Reidel, Dordrecht.
Davis, H. and Memmott, J. 1982: Counting Behavior in Animals: A critical Evaluation. *Psychological Bulletin*, 92, 547–71.
de Villiers, M., 1923: *The Numeral Words: their Origin, Meaning, History and Lesson*. H. F. and G. Witherby, London.
Diehl, R. L. and Kolodzey, K. F. 1981: Spaka: a private language. *Language*. 57, 406–24.
Dixon, R. M. W. 1977: Where Have All the Adjectives Gone? *Studies in Language*. 1, 19–80. (Reprinted in Dixon, 1982, pp. 1–62.)
1980: *The Languages of Australia*. Cambridge University Press, Cambridge.

1982: *Where Have All the Adjectives Gone? and other Essays in Semantics and Syntax*. Mouton, Berlin.
forthcoming, *A Grammar of Boumaa Fijian*.
Eddington, A. S. 1928: *The Nature of the Physical World*. Cambridge University Press, Cambridge.
Epstein, S. 1978: Review of Hurford, 1975. *Journal of Linguistics*, 14, 123–24
Feldin, H. Goldin-Meadow, S. and Gleitman, L. 1978: Beyond Herodotus: the Creation of Language by Linguistically Deprived Deaf Children. In Lock (1978), pp. 351–414.
Fletcher, P. and Garman, M. (eds) 1979: *Language Acquisition*. Cambridge University Press, Cambridge.
Field, H. 1980: *Science without Numbers*. Basil Blackwell, Oxford.
Frege, G. 1950: *The Foundations of Arithmetic*: (English Translation by J. L. Austin). Basil Blackwell, Oxford.
Fodor, J. D., 1984: Constraints on Gaps: is the Parser a Significant Influence? In Butterworth et al. (1984), pp. 9–34.
Fodor, J. A., 1976: *The Language of Thought*. The Harvester Press, Hassocks, Sussex.
Fuson, K. C., and Mierkiewicz, D. 1980: A Detailed Analysis of the Act of Counting. Paper presented at the Annual Meeting of the American Educational Research Association, Boston, Mass.
Fuson, K. C. and Hall, J. W., 1983: The Acquisition of Early Number Word Meanings: a Conceptual Analysis and Review. In Ginsburg (1983), pp. 49–107.
Fuson, K. C., Richards, J. and Briars., D. J. 1982: The Acquisition and Elaboration of the Number Word Sequence. In Brainerd (1982), pp. 33–92.
Gelman, R. and Gallistel, C. R. 1978: *The Child's Understanding of Number*. Harvard University Press, Cambridge, Mass.
Ginsburg, H. P. (ed.) 1983: *The Development of Mathematical Thinking*. Academic Press, New York.
Ginsburg, H. P. and Russell, R. L. 1981: Social-Class and Racial Influences on Early Mathematical Thinking. *Monographs of the Society for Research in Child Development*, 46.
Givón, T. 1972: Studies in ChiBemba and Bantu Grammar. *Studies in African Linguistics*, 3, Supplement 3.
1979: *On Understanding Grammar*. Academic Press, New York.
Godel, R. 1970: F. De Saussure's Theory of Language. In Sebeok, *Current Trends in Linguistics*, pp. 479–493.
Goodglass, H. 1983: Disorders of Lexical Production and Comprehension. In Bäuerle et al. (1983), 134–46.
Goodman, N. 1952: On Likeness of Meaning. In Linsky (1952), Reprinted in Olshewsky (1969), pp. 537–42.

Goodwin, W. W. 1890: *An Elementary Greek Grammar*. Ginn and Co., Boston.
Goody, J. 1977: *The Domestication of the Savage Mind*. Cambridge University Press, Cambridge.
Gratzer, G. 1971: *Lattice Theory: First Concepts and Distributive Lattices*. W. H. Freeman, San Francisco.
Greenberg, J. H. 1972: Numeral Classifiers and Substantival Number: Problems in the Genesis of a Linguistic Type. *Proceedings of the 11th Congress of Linguists, Bologna*, also in *Working Papers in Language Universals* (Stanford University), Vol. 9, pp. 1–39.
1963a: Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In Greenberg, (1963(b), pp. 73–113.
(ed.) 1963b: *Universals of Language*. MIT Press, Cambridge, Mass.
1975: Dynamic Aspects of Word Order in the Numeral Classifier. In Li (1975), pp. 27–45.
Griffiths, P. D. 1977: Review of Hurford, 1975. *York Papers in Linguistics*, 7, 219–23.
Groenendijk, J. A. G, Janssen, T. M. V. and Stokhof, M. B. J., (eds) *Formal Methods in the Study of Language*, University of Amsterdam Mathematical Centre Tract 136.
Gunderson, K. (ed.) 1975: *Language, Mind, and Knowledge*. University of Minnesota Press, Minneapolis.
Gupta, A. 1980: *The Logic of Common Nouns: An Investigation into Quantified Modal Logic*. Yale University Press, New Haven, Conn.
Haddon, A. C. 1889: Ethnography Western Tribes Torres Strait. *Journal of the Anthropological Institute*.
Haiman, J. 1980: The Iconicity of Grammar. *Language*, 55, 515–40.
Hankamer, J. 1973: Unacceptable Ambiguity. *Linguistic Inquiry*, 4, 17–68.
Hayes, B. J. and Masom, W. F. 1928: *The Tutorial Latin Grammar*. University Tutorial Press, London.
Hellan, L. 1980: *Toward an Integrated Theory of Noun Phrases*. University of Trondheim.
Hermes, H., Kambartel, F. and Kaulback, F. (eds) 1979: *Frege: Posthumous Writings*. Chicago University Press, Chicago, Ill.
Hetzron, R. 1977: Innovations in the Semitic Numeral System. *Journal of Semitic Studies*, 22, 167–201.
Hintikka, J. and Kulas, J. 1983: *The Game of Language*. Reidel, Dordrecht.
Hodes, H. 1984: Logicism and the Ontological Commitments of Arithmetic. *Journal of Philosophy* 81, 123–49.
Hoepelman, J. 1983: Adjectives and Nouns: a New Calculus. In Bäuerle et al. (1983), pp. 190–220.
Hopper, P. J. and Thompson, S. A. 1984: The Discourse Basis for Lexical Categories in Universal Grammar. *Language*, 60, 703–52.

Hudson, R. A. 1980: *Sociolinguistics*. Cambridge University Press, Cambridge.
Hughes, M. 1984: Learning about Number. *ESRC Newsletter*, 52, 9–11.
1986: *Children and Number: Difficulties in Learning Mathematics*. Basil Blackwell, Oxford.
Humboldt, W. von 1832–9: Über die Kawi-Sprache der Insel Java. In *Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*.
Hurford, J. R. 1975: *The Linguistic Theory of Numerals*. Cambridge University Press, Cambridge.
1977: The Significance of Linguistic Generalizations. *Language* 53, 574–620.
1979a: Numerals and the Homogeneity of Description and Explanation. *Lingua*, 48, 35–42.
1979b: Review article on Postal 1974, *Journal of Linguistics*, 15, 111–20.
1980: Generative Growing Pains. *Lingua*, 50, 117–53.
Hyman, L. 1984: Form and Substance in Language Universals. In Butterworth et al. (1984), pp. 105–22.
Hymes, V. D. 1955: Athapaskan Numeral Systems. *International Journal of American Linguistics*, 21, 26–45.
Jacob, J. 1965: Notes on the Numerals and Numeral Coefficients in Old, Middle and Modern Khmer. *Lingua*, 15, 143–162.
Jakobson, R. 1968: *Child Language, Aphasia and Phonological Universals*. Mouton, The Hague.
Jenewari, C. E. W. 1980: *A Linguistic Guide to Spoken Kalabari*. School of Humanities, University of Port Harcourt, Port Harcourt, Nigeria.
Jespersen, O. 1909, reprinted 1965: *A Modern English Grammar on Historical Principles*, Part I, *Sounds and Spellings*. George Allen and Unwin, London.
1964: *Language: its Nature, Development and Origin*. W. W. Norton, New York.
1965: *The Philosophy of Grammar*. W. W. Norton, New York.
1969: *Analytic Syntax*. Holt, Rinehart and Winston, New York.
Johnson, S. 1747: *The Plan of a Dictionary*. (Facsimile edition, 1970), The Scolar Press, Menston.
Johnson-Laird, P. 1983: *Mental Models*. Cambridge University Press, Cambridge.
Kamp, J. A. W. 1975: Two Theories about Adjectives. In Keenan (1975), pp. 123–55.
Katz, J. 1981: *Language and Other Abstract Objects*. Basil Blackwell, Oxford.
Kautzsch, E. 1910: *Gesenius' Hebrew Grammar*, trans. A. E. Cowely. Clarendon Press, Oxford.
Keenan, E. L. (ed.) 1975: *Formal Semantics of Natural Language*. Cambridge University Press, Cambridge.
Keenan, E. L. and Faltz, L. M. 1985: *Boolean Semantics for Natural Language*. Reidel, Dordrecht.

Kempson, R. M. 1975: *Presupposition and the Delimitation of Semantics*. Cambridge University Press, Cambridge.
Kempson, R. M. and Cormack, A. 1981: Ambiguity and Quantification. *Linguistics and Philosophy*, 4, 259–309.
Killingley, S.-Y. 1982: *A Short Glossary of Cantonese Classifiers*. Grevatt and Grevatt, Newcastle upon Tyne.
1983: *Cantonese Classifiers: Syntax and Semantics*. Grevatt and Grevatt, Newcastle upon Tyne.
Kimball, J. P. (ed.) 1973: *Syntax and Semantics*, Vol 2. Seminar Press, New York.
Kitcher, P. 1978: The Nativist's Dilemma. *Philosophical Quarterly*, 28, 1–16.
1984: *The Nature of Mathematical Knowledge*. Oxford University Press, Oxford.
Klahr, D. and Wallace, J. G. 1973: The Role of Quantification Operators in the Development of Conservation of Quantity. *Cognitive Psychology*, 4, 301–27.
Klahr, D. and Wallace, J. G. 1976: *Cognitive Development: an Information Processing View*. Erlbaum, New York.
Kleene, S. C. 1967: *Mathematical Logic*. John Wiley, New York.
Klein, E. 1981: The Syntax and Semantics of Nominal Comparatives. In Moneglia (1981), pp. 223–53.
1981: The Interpretation of Adjectival, Nominal and Adverbial Comparatives. In Groenendijk et al., *Formal Methods*, pp. 381–98.
Kluge, T. 1937–42: I. *Die Zahlenbegriffe der Sudansprachen*; II. *Die Zahlenbegriffe der Australier, Papua und Bantuneger*; III. *Die Zahlenbegriffe der Voelker Americas, Nordeurasiens, der Munda und der Palaioafricaner*; IV. *Die Zahlenbegriffe der Dravida, der Hamiten, der Semiten und der Kaukasier*; V. *Die Zahlenbegriffe der Sprachen Central–und Suedostasiens, Indonesiens, Micronesiens und Polynesiens*. Published by the author, Berlin.
Ladefoged, P. 1980: What are Linguistic Sounds Made of? *Language*, 56, 485–502.
Langacker, R. W. 1971: French Interrogatives Revisited. Unpublished paper, University of California, San Diego.
Lass, R. 1984: *Language and Time: a Historian's View*, Inaugural Lecture, University of Capetown, New Series No. 90, Capetown.
Lean, G. A. 1985–86: *Counting Systems of Papua New Guinea*, Vols 1–11 and Research Bibliography (draft editions). Department of Mathematics, Papua New Guinea University of Technology.
Leech, G. N. 1969: *Towards a Semantic Description of English*. Longman, London.
Leonard, H. S. and Goodman, N. 1940: The Calculus of Individuals and its Uses. *Journal of Symbolic Logic*, 5, 45–55.
Lepsius, R. A. M. 1865: Die altägyptische Elle and ihre Einteilung. *Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*.

Lewis, D. 1972: General Semantics. In Davidson and Harman (1972), pp. 169–218.
1975: Languages and Language. In Gunderson, (1975), pp. 3–35.
Li, C. (ed.) 1975: *Word Order and Word Order Change*. University of Texas Press, Austin.
Lightfoot, D. W. 1979: *Principles of Diachronic Syntax*. Cambridge University Press, Cambridge.
Lindblom, B., MacNeilage, P. and Studdert-Kennedy, M. 1984: Self-organizing Processes and the Explanation of Phonological Universals. In Butterworth et al. (1984), pp. 181–203.
Link, G. 1983: The Logical Analysis of Plurals and Mass Terms: a Lattice-Theoretical Approach. In Bäuerle, et al. (1983), pp. 302–23.
1985: Generalized Quantifiers and Plurals. Unpublished rough draft, from the author at University of Munich, or CSLI, Stanford.
(forthcoming): Plural. In Wunderlich and von Stechow (forthcoming).
Linsky, L. (ed.) 1952: *Semantics and the Philosophy of Language*. University of Illinois Press, Urbana.
Lock, A. 1978: *Action, Gesture and Symbol: the Emergence of Language*. Academic Press, London.
Locke, J. 1975: *An Essay Concerning Human Understanding*. Oxford University Press, Oxford.
Lyons, J. 1977: *Semantics*, 2 vols. Cambridge University Press. Cambridge.
Mandler, G. and Shebo, B. J. 1982: Subitizing: an Analysis of its Component Processes. *Journal of Experimental Psychology: General*, 111, 1–22.
McCawley, J. D. 1981: *Everything that Linguists have always Wanted to Know about Logic*. Basil Blackwell, Oxford.
Menn, L. and MacWhinney, B. 1984: The Repeated Morph Constraint. *Language*, 60, 519–41.
Menninger, K. 1969: *Number Words and Number Symbols*. MIT Press, Cambridge, Mass.
Merrifield, W.R., 1968: Number Names in Four Languages of Mexico. In Corstius (1968), 91–102.
Mill, J. S. 1906: *A System of Logic Ratiocinative and Inductive*, 8th edn. Longmans Green, London.
Miller, G. 1951: *Language and Communication*. McGraw-Hill, New York.
Mitchell, T. F. 1962: *Colloquial Arabic*, (Teach Yourself Books). English Universities Press, London.
Modgil, S. and Modgil, C. (in press): *Noam Chomsky: Consensus and Controversy*. Falmer Press, Lewes, Sussex.
Moneglia, M. (ed.) 1981: *Tempo Verbale Strutture Quantificate in Forma Logica: Atti del Seminario, Accademia della Crusca*. Presso l'Accademia della Crusca, Florence.
Moravcsik, J. M. 1983: Natural Languages and Formal Languages: a Tenable Dualism. In Cohen and Wartofsky, (1983), pp. 225–239.

Ogden, C. K. and Richards, I. A. 1949: *The Meaning of Meaning*, 10th edn. Routledge and Kegan Paul, London.
Olshewsky, T. M. (ed.) 1969: *Problems in the Philosophy of Language*. Holt, Rinehart and Winston, New York.
Opie, I. and Opie, P. 1959: *The Lore and Language of Schoolchildren*. Oxford University Press, Oxford.
Palayer, P. 1970: *Elements de Grammaire Sar (Tchad)*. Afrique et Langage, and College Charles Lwanga, Fort-Archambault, Chad.
Partee, B. and Rooth, M. 1983: Generalized Conjunction and Type Ambiguity. In Bäuerle et al. (1983), pp. 361–83.
Pateman, T. 1983: What is a Language? *Language and Communication*, 3, 101–27.
1985: From Nativism to Sociolinguistics: Integrating a Theory of Language Growth with a Theory of Speech practices. *Journal for the Theory of Social Behaviour*, 15, 38–59.
1987: *Language in Mind and Language in Society*. Oxford University Press, Oxford.
Piaget, J. 1952: *The Child's Conception of Number*. Routledge and Kegan Paul, London.
Piattelli-Palmarini, M. (ed.) 1980: *Language and Learning: the Debate between Jean Piaget and Noam Chomsky*. Routledge and Kegan Paul, London.
Popper, K. 1972: *Objective Knowledge*. Oxford University Press, London.
Popper, K. 1973: Indeterminism is not Enough. *Encounter*, April, 1973: 20–6.
Postal, P. M. 1974: *On Raising: one rule of English Grammar and its theoretical implications*. MIT Press, Cambridge, Mass.
Pott, A. F. 1847: *Die Quinare und Vigesimale Zählmethode bei Völkern aller Welttheile*. Dr Martin Sandig, Wiesbaden.
Power, R. J. D. and Longuet-Higgins, H. C. 1978: Learning to Count: a Computational Model of Language Acquisition. *Proceedings of the Royal Society of London B*, 200, 391–417.
Prior, M. H. 1985: Syntactic Universals and Semantic Constraints. Ph.D. thesis, University of London.
Pukui, M. K. Elbert, S. H. and Mookini, E. T. 1975: *The Pocket Hawaiian Dictionary: with a Concise Hawaiian Grammar*. The University Press of Hawaii, Honolulu.
Pulman, S. G. 1983: *Word Meaning and Belief*. Croom Helm, London.
Putnam, H. 1975a: Mathematics without Foundations. In Putnam (1975b), pp. 43–59.
1975b: *Mathematics, Matter and Method: Philosophical Papers*, Vol. I. Cambridge University Press, Cambridge.
1980: What is Innate and Why: Comments on the Debate. In Piattelli-Palmarini (1980), pp. 287–309.
Quirk, R., Greenbaum, S., Leech, G. N. and Svartvik, J. 1972: *A Grammar of Contemporary English*. Longman, London.

Rajaobelina, P. 1966: *Parler Malgache*. Imprimerie Lutherienne, Antsahamanitra, Tananarive, Madagascar.
Resnik, M. D. 1980: *Frege and the Philosophy of Mathematics*. Cornell University Press, Ithaca, New York.
Russac, R. J. 1983: Early Discrimination among Small Object Collections. *Journal of Experimental Child Psychology*, 36, 124–38.
Russell, B. 1919: *Introduction to Mathematical Philosophy*. George Allen and Unwin, London.
Rutherford, W. G. 1930: *First Greek Grammar*. Macmillan, London.
Salman, D. H. 1943: Note on the Number Conception in Animal Psychology. *British Journal of Psychology*, 33, 209–19.
Sapir, E. 1949: *Language*. Harcourt, Brace and World, New York.
Saussure, F. de 1959: *Course in General Linguistics*, trans. Wade Baskin. McGraw–Hill, New York.
Saxe, G. B. 1981: Body Parts as Numerals: a Developmental Analysis of Numeration among Remote Oksapmin Village Populations in Papua New Guinea. *Child Development*, 52, 306–16.
1982a: Culture and the Development of Numerical Cognition: Studies among the Oksapmin of Papua New Guinea. In Brainerd (1982), pp. 157–76.
1982b: Developing Forms of Arithmetical Thought among the Oksapmin of Papua New Guinea. *Developmental Psychology*, 18, 583–94.
Saxe G. B. and Moylan, T. 1982: The Development of Measure Operations among the Oksapmin of Papua New Guinea. *Child Development*, 53, 1242–8.
Schaeffer, B., Eggleston, V. H. and Scott, J. L. 1974: Number Development in Young Children. *Cognitive Psychology*, 6, 357–79.
Schon, D. A. 1963: *Invention and the Evolution of Ideas*. Tavistock, London.
Sebeok, T., (ed.) 1970: *Current Trends in Linguistics*, Vol. 3. Mouton, The Hague.
Seidenberg, A. 1960: *The Diffusion of Counting Practices, University of California Publications in Mathematics*, Vol. 3, 215–300. University of California Press, Berkeley.
Siegel, L. S. 1982: The Development of Quantity Concepts: Perceptual and Linguistic Factors. In Brainerd (1982), pp. 123–55.
Siegel, M. E. A. 1980: *Capturing the Adjective*. Garland, New York.
Sigurd, B. 1977: Review of Hurford, 1975, *Studia Linguistica*. 31, 192–4.
Siromoney, R. 1968: Grammars of Number Names of Certain Dravidian Languages. In Corstius (1968), pp. 82–90.
Sperber, D. and Wilson, D. 1986: *Relevance*. Basil Blackwell, Oxford.
Spiro, M. (ed.) 1965: *Context and Meaning in Cultural Anthropology (Festschrift Hallowell)*. New York.

Starkey, P. and Cooper, R. G. 1980: Perception of Numbers by Human Infants. *Science*, 210, 1033–5.
Strawson, P. F. 1959: *Individuals: an Essay in Descriptive Metaphysics*. Methuen, London.
Tennant, N. 1981: Formal Games and Forms for Games. *Linguistics and Philosophy*, 4, 311–20.
Thomas, R. K., Fowlkes, D. and Vickery, J. D. 1980: Conceptual Numerousness Judgments by Squirrel Monkeys. *American Journal of Psychology*, 93, 247–57.
Thorndike, E. and Lorge, I. 1944: *The Teacher's Word Book of 30,000 Words*. Columbia University Teachers' College, New York.
Tylor, E. B. 1891: *Primitive Culture*. Vol. 1. John Murray, London.
Vesper, E. 1969: Phrase Structures of Kusaien. MA thesis, University of California at Davis.
Wall, R. 1972: *Introduction to Mathematical Linguistics*. Prentice–Hall, Englewood Cliffs, NJ.
Weisenberg, T. and McBride, K. E. 1935: *Aphasia*. Commonwealth Fund, New York.
Wesley, F. 1961: The Number Concept: a Phylogenetic Review, *Psychological Bulletin*, 58, 420–8.
Whitehead, A. N. and Russell, B. 1962: *Principia Mathematica*. Cambridge University Press, Cambridge.
Wilder, R. L. 1968: *Evolution of Mathematical Concepts*. John Wiley and Sons, Inc., New York.
Winston, M. E. 1982: *Explanation in Linguistics: a Critique of Generative Grammar*. Indiana University Linguistics Club, Bloomington, Indiana.
Wittgenstein, L. 1972: *Philosophical Investigations*. Basil Blackwell, Oxford.
1974: *Philosophical Grammar*. Basil Blackwell, Oxford.
Wright, C. 1983: *Frege's Conception of Numbers as Objects*. Aberdeen University Press, Aberdeen.
Wunderlich, D. and von Stechow, A. (eds) (forthcoming): *Handbook of Semantics*.

Index