CITY UNIVERSITY OF HONG KONG

Self-organization and Language Evolution: System, Population and Individual

Submitted to
Department of Electronic Engineering
partial fulfillment
for the Degree of Doctor of Philosophy

by
Jinyun Ke

June 2004

Abstract

This thesis proposes a framework adopting the self-organization theory for the study of language evolution. Self-organization explains collective behaviors and evolution with the observation that the patterns at the global level in a complex system are often properties spontaneously emergent from the numerous local interactions among the individual components, and they cannot be understood by only examining the individual components.

Language can be viewed as such emergent properties instead of products from some innate blueprint in humans. We highlight the importance of recognizing language at two distinctive but inter-dependent levels of existence, i.e. in the idiolect and in the communal language, and a self-organizing process existing at each of the two levels. It is necessary to clarify what phenomena are properties of the idiolects, and what properties are the collective behaviors at the population level.

In linguistics, however, very often an abstract language system is taken as the object of analysis. This level of analysis disregards the distinction between idiolect and communal language, and neglects the heterogeneous nature of language at both levels. As a consequence, explanations for observed patterns based on this abstract level of analysis are often inadequate. However, this is a necessary step for linguists to identify interesting phenomena in the first place. At this abstract level of analysis, the self-organization framework can also be applied. It is assumed that the abstract language system self-organizes. A study on homophony in languages is taken as an example to illustrate the analysis at this level. It is shown that the existence of homophony reflects several self-organization characteristics in a dynamic process of language evolution, such as the predictable degree of homophony, the disyllabification in Chinese dialects, the differentiation of homophone pairs in grammatical class.

We are further interested in how the self-organization is implemented. To answer this question, we need to look into the idiolects in this self-organizing process, to know how the idiolects are formed and affect each other. Language change provides an informative window in addressing these issues. Language change is the result of the collective behaviors of idiolects, even as it affects the idiolects. The heterogeneity

among idiolects is exposed to the greatest extent in on-going changes.

An on-going sound change in Cantonese is taken as a case study to scrutinize the heterogeneity in the self-organizing processes. The fieldwork data reveal a large degree of variation both in the population (VT-I) and in the set of words (VT-II). Another type of variation (VT-III) is highlighted, that is, a word may also show variation within one single speaker. But this VT-III within speakers only exists in a proportion, but not all, of the words subject to the change. Also we find that if a speaker has some words consistently in the unchanged state and some words in the changed state, then this speaker must have some other words in the variation state. Most speakers show the existence of VT-III, but they vary in degree. The observed individual differences in the degree of VT-III suggest that the large heterogeneity may be not only accounted for by the variability of linguistic input, but also by individuals' different learning styles. We hypothesize two types of lexical learning styles, i.e. probabilistic and categorical learning. These differences in learning styles suggest that when we examine the agent's internal properties in the self-organization framework, it is not only necessary to examine the commonalities among agents, but also the differences among them.

In addition to empirical studies, this thesis employs computational modeling as a major tool for investigation, as modeling provides effective ways to test hypotheses beyond empirical studies, and suggests new questions. After a brief review of the modeling studies in the field, some models developed in this thesis for language origin and language change are reported.

The first model is to simulate the emergence of a consistent vocabulary from a set of random mappings between meanings and forms. It emphasizes the importance of implementing the actual process of interaction among agents, and the cumulative effect on agents' linguistic behaviors. The model suggests that the Saussurean sign with identical speaking and listening mappings may not be a biological predisposition from natural selection, but rather a result from the process of language learning and use. The process exhibits a phase transition from a long period of small oscillation to an abrupt convergence. Such phase transition is often observed in self-organizing systems.

The second model simulates language change as innovation diffusion, and examines the effects of various factors, including some concerning properties of agents and some affecting agents' interactions. By comparing the outcome under different conditions, the model illustrates the importance of incorporating realistic assumptions, such as finite population size, age-dependent propensity to change, different learning environment in a social network, etc. The model compares the dynamics of language change in different types of network structures and shows that in non-regular networks, the rate of innovation diffusion increases little as population size increases. The model also tests the effect of the two types of hypothesized learning styles, and shows that in a population with the presence of probabilistic learners, an innovation with a small advantage will easily spread into the population and lead to a change. This may explain why language changes are so frequent.

This thesis demonstrates that both empirical and modeling studies on language evolution can greatly benefit from adopting a self-organization framework. The convergence and interplay of the two lines of exploration, i.e. biological bases in agents and the long term effect of interactions among them, should bring us a deeper understanding of how language has evolved and is evolving.

Acknowledgements

I am indebted to my supervisor Prof. William Shi-Yuan WANG. I would not have had the chance to start a PhD study in linguistics without Prof. Wang. He led me into linguistics from a student of engineering, and have guided me with his special wisdom to learn how to appreciate and cherish the beauty and joy in doing research in basic science from a multi-disciplinary perspective. I am deeply grateful for his unfailing patience and continuous encouragement all along these years. He has helped me to learn how to be disciplined and persistent. He is not only a mentor of research and scholarship, but also a mentor of life.

I would like to extend my gratitude to Prof. Ron CHEN who has stimulated my interest in complex network studies, to Prof. CHENG Chin-chuan who has been very helpful in my study of Chinese dialects and whose windows-version DOC program is so useful in the study of homophony, to Prof. John HOLLAND for his stimulation and encouragement in the last period of the study and valuable comments on the thesis draft, and to Prof. Thomas Hun-tak LEE from whom I have learnt a lot about language acquisition and about devotion and enthusiasm for linguistics. Also I would like to thank my PhD panel members Prof. Felipe CUCKER and Prof. YAN Hong in CityU. I am thankful for Prof. Kim Man who first gave me the chance to pursuit my graduate study in Hong Kong. In addition, I would like to thank the following professors from whom I benefit in various ways: CHEN Bao-ya, Morten CHRISTIANSEN, DENG Xiao-hua, HE Jian-hua, Simon KIRBY, Partha NIYOGI, Godfrey LIU, Mieko OGURA, Merritt RUHLEN, Laurent SAGART, Tom SCHOENEMANN, Luc STEELS, Benjamin TSOU and Eric ZEE.

I have benefited a lot from a number of visits to other institutions and I would like to thank the directors and professors who gave me these opportunities. They include Prof. MU Ji-hong in Yunnan University in China, Prof. HO Dah-an in the Institute of Linguistics, Academia Sinica in Taiwan, Prof. Jean-Marie HOMBERT in the Laboratoire Dynamique de Langage, Institut des Sciences de l'Homme in Lyon, France, Prof. Erica JEN in Santa Fe Institute in US. Also I would like to thank the organizers of the 2001 Complex Systems Summer School in Santa Fe Institute, where

I first got to know studies on complex networks.

The Language Engineering Laboratory of CityU has been my home in the last five years. I am grateful for all members of this great family, including AU Chin-pong, Marco CABOARA, Charles CHEN, Joyce CHEUNG, GONG Tao, KONG Jiang-ping, LIANG Yuan, LIU Juan, James MINETT, PENG Gang, WANG Feng, YAO Yao, ZHANG Bo and ZHENG Hong-ying. And I thank several visiting scholars to the lab for their stimulating interactions, including Mahe BEN HAMED and Christophe COUPÉ from France, and Charles Chien-jer LIN from Taiwan. I am specially grateful for James and Christophe for their invaluable collaboration and help in the studies presented in this thesis, Christophe again for his thorough comments on the draft of several parts of the thesis, Peng Gang for his help in solving all sorts of technical problems with computers, and Wang Feng and Yao Yao for their help in data processing for the disyllabification in Chinese dialects and the Cantonese fieldwork data, respectively. Also I thank Dinoj SURENDRAN of Chicago University in helping me in use of CELEX database.

I would like to thank QIU Dong-yuan, my sister KE Jin-xia, brother-in-law LI Jin-yan and my angelic niece Kaiwen for the happy moments they have given me outside of research. I would like to express my utmost gratitude to my parents, KE Ming and YANG Bing-qing for their unparalleled care and support. Lastly, I thank my husband YAO Yuan. Without him, I would not have made up my mind to start this PhD study, nor have been able to find the courage to complete this journey. I dedicate this thesis to the loves I have been immersed in all these years.

Table of Contents

List of Figures

List of Tables

Chapter 1

Theoretical framework: Self-organization and language evolution

The most characteristic trait of humans is curiosity about ourselves. Among the many questions in the curiosity pool, "Why do humans have language?", "When did humans begin to have language?" and "How did language come about?" may be some of the most intriguing ones. Language is generally considered as one defining trait of human species. Its origin and evolution along the historical scale (phylogeny) and the individual scale (ontogeny) are foci of interest of scientific investigation of the human nature. There may be no other research area which arouses such wide interests across different disciplines. Various fields have joined in the investigation, including paleoanthropology, archaeology and genetics which work on reconstructing human history from its origin; primatology and animal communication which analyze the similarities and differences between humans and other animals in physiology, cognition and social structure; neurosciences which reveal how the brain functions for language processing; linguistic typology which collects the universal characteristics in the thousands of extant languages; sociolinguistics which informs how language is influenced by social factors and how it is used for building and asserting social identities; developmental psychology and psycholinguistics which tackle the puzzle how children can learn language within a few years with apparently little effort. A number of anthologies provide a glimpse of such convergent efforts in this exciting area in recent years. To name a few in the list: Hawkins and Gell-Mann (1992); Hurford et al. (1998b); Knight et al. (2000); Wray (2002), among many others.

1.1 Three perspectives on language evolution – biological, cultural, or both?

The study of language evolution from an evolutionary viewpoint can be divided into the following perspectives. One is to view language as the result of a biological evolution process governed by natural selection mechanism. The second is to emphasize the nature of language as a cultural phenomenon and study its evolution as a cultural selection process. Very often these two perspectives are contradictory to each other when scholars try to explain a certain linguistic phenomenon as exclusively either biological or cultural. However, these two perspectives should be considered as complementary, in their contribution to constructing a full picture of language evolution, especially regarding language emergence. In addition to the above two directions of exploration, a more recent one is to view biological and cultural aspects of language as co-evolving. In the following, I will give a brief review on these views.

1.1.1 Biological evolution of language

There have been mainly two types of observations which have led to the view that language should be studied as a biological phenomenon. One is that different human languages exhibit similarities of various structural features as if "cut from the same pattern" (Greenberg, 1963, p255). The other is that children normally learn language rapidly and with little effort uniformly across different cultures and societies. To account for these two facts, it has been suggested that in humans there exist either a specific language faculty (Chomsky, 1972), a bioprogram (Bickerton, 1984), or a language instinct (Pinker, 1994), which are assumed highly modular and task-specific devoted to language. And the development of such a device is assumed to be genetically determined.

There are two types of proposals to explain the emergence of language as an innate language faculty from an evolutionary perspective, namely the exaptationist and the adaptationist ones (Christiansen, 1994). The first explanation suggests that language may not be the direct product of natural selection, but a by-product of other evolutionary consequences such as an increase of the brain size and restructuring of the brain (Chomsky, 1986; Piattelli-Palmarini, 1989). The term "exaptation" was first proposed by Gould and Vrba (1982), which says new uses are made of parts that were originally adapted to some other functions, or made of parts that had no function at all but were present for other reasons, like the spandrels in the dome of the San Marco basilica in Venice (Gould and Lewontin, 1979)[1]. On the other hand, adaptationists propose natural selection as the only explanation for a complex language faculty. As stated in Pinker and Bloom (1990),

> "Evolutionary theory offers clear criteria for when a trait should be attributed to natural selection: complex design for some function, and the absence of alternative processes capable of explaining such complexity. Human language meets this criterion: grammar is a complex mechanism tailored to the transmission of propositional structures through a serial interface."

But what exactly is the structure or function of the innate language faculty? The generative school posits it is a Universal Grammar (UG) which is wired into the brain and determines the nature of grammatical structures which are learnable by children. In an earlier paradigm, a set of "principles and parameters" was proposed as the basis of UG, such as head parameter (i.e., the parameter which determines the relative positioning of heads with respect to their complements), wh-parameter, null subject parameter, and so on. In the latest minimalist program (Chomsky, 1995), many of the elements in UG proposed earlier have been reformulated as some basic operations, such as merge and move. However, the UG school retains the position that these innate mechanisms are specific to language.

[1]"Spandrels - the tapering triangular spaces formed by the intersection of two rounded arches at right angles ... are necessary architectural by-products of mounting a dome on rounded arches" (Gould and Lewontin, 1979). The space is filled with mosaics as decorations. But it should not be mistaken that the spandrels are designed on purpose for decoration.

When it comes to how the language faculty as defined by UG evolved under natural selection, scholars within the nativist school differ in their proposed scenarios. Pinker and Bloom (1990) and Newmeyer (1991) propose that each element in UG may have its own adaptive function and therefore UG may have evolved gradually and in a piecemeal manner as the result of natural selection. To explain how natural selection is at work, pressures from learnability, expressiveness, or parsing with memory constraints are considered as selection forces to account for the emergence of UG. Lightfoot, on the contrary, suggests that "...some effects of UG are dysfunctional" and proposes UG is like "spandrels evolving as a by-product of something else and not the result of adaptive change favouring survival to the reproductive age" (Lightfoot, 2000, p245-246).

In the following I will take "subjacency" as an example to illustrate the controversy between adaptationist and exaptationist. "Subjacency", a principle included in the early UG framework, refers to prohibition "against dependencies between a gap and its antecedent that spans certain combinations of phrasal nodes" (Pinker and Bloom, 1990). Putting it informally, subjacency "keeps rules from relating elements that are 'too far apart from each other', where distance apart is defined in terms of the number of designated nodes that there are between them" (Hurford, 1999a). It accounts for the violation of grammaticality such as in the sentence

(1) *What do you wonder where John put? (Newmeyer, 1991)

The ungrammatical sentence (1) can be analyzed as follows: [What $_{Comp1}$ [do you wonder] $_{S1}$] $_{S1'}$ [where $_{Comp2}$ [John put _] $_{S2}$] $_{S2'}$

In the framework of generative grammar, the position of the wh-items can be explained by movement, and the subjacency principle states that only one single bounding node may be crossed during any move. In (1), there are two bounding nodes (i.e., $_{S2}$ and $_{S1}$), and the movement directly from the gap (the object of "put") to the position of Comp1 is not allowed. In contrast, the following sentence (1') is grammatical, as there could be two moves, and only one bounding node is crossed during each move, including the first move from the gap position to Comp2 ("that"), and the second move from Comp2 to Comp1.

(1)' [What $_{Comp1}$ [do you wonder]$_{S1}$] $_{S1'}$ [that$_{Comp2}$ [John put _] $_{S2}$] $_{S2'}$

Adaptationists believe that the Subjacency Principle provides a helpful constraint on the assignment of an understood grammatical role for displaced elements such as question words and relative pronouns, and therefore, this principle would confer an advantage for effective communication and therefore became selected through natural selection. Pinker and Bloom (1990) consider this principle provides a compromise solution to the conflicting constraints of processing from both speaker and listener, while Newmeyer (1991) views these two constraints as asymmetrical and places importance on the listener's part. Such explanations have received various criticisms (see the special issue in Language and Communication, 1991, 11(1)). One of the most common criticisms is that it is not convincing that those who solved the parsing problem by evolving a subjacency constraint were more likely to reach puberty and a higher reproduction ability (Lightfoot, 1991). Fouts (1991) presents an extreme version: "it seems critical to me that he [Newmeyer] demonstrates how a human male or female who uses Chomskian perfect grammar has a better chance of breeding than one who failed English 101 and is noted for ungrammatical monosyllabic utterances yet has bedroom eyes and drives a BMW" (p42) (quoted in Hurford (1999a)). Hurford also points out another problem concerning the fitness of a mutant with the subjacency principle. He comments that a mutation of having a subjacency constraint would have put a child at a disadvantage as he would actually suffer from being unable to understand the utterances produced by his parents who do not have the subjacency principle. He suggests that it is be hard for constraints, which limit the sets of structures that grammars can generate, to evolve.

While adaptationist or exaptationist in the nativist school argue for how the innate subjacency principle has evolved, non-nativists totally disregard the existence of such a principle. Tomasello (2003) states that "in the current view, the principles and structures whose existence it[sic] is difficult to explain without universal grammar (such as Chomskian things such as subjacency constraint, ...) are theory-internal affairs and simply do not exist in usage-based theories of language" (p7). Ellefson and Christiansen (2000) use evidence from artificial language learning and neural network modeling to argue that "rather than having an innate UG principle to rule out subjacency violations, ... they may have been eliminated altogether through an evolutionary process of linguistic adaptation constrained by prior cognitive limitations on sequential learning and processing".

It is beyond doubt that there are some biological bases for language, and some of them are possibly specific to humans. However, we disbelieve that natural selection produce a "language faculty" with very refined grammatical constraints and structures, such as the UG advocated by the generative linguists. Contrary to the UG proposal, language is better regarded as "a kind of 'interface' among a variety of more basic abilities" (Wang, 1982a, p116), which makes the postulation of "a language faculty" unnecessary. "These abilities underlie nonlinguistic processes as well. ... Many of these abilities are present to different degrees in other animals. Most of them probably emerged much earlier than language in hominid evolution. Gradually and piece by piece, these abilities were increasingly made accessible for use in the elaborations of language, much as adding pieces to a mosaic" (ibid). In recent years studies on primate cognition and animal communication have provided more and more support to this view (e.g., Hauser, 1996; Tomasello, 2000). We will elaborate this further in Section 1.3.5 about some properties of a language-ready body.

One of the strong arguments put forward to support an innate language faculty or UG comes from the "logical problem of language learning"[2]. However, many empirical and computational studies on language acquisition have closely examined the actual process of children's phonological, lexical and grammatical development, and tend to

[2] A version of "the logical problem of language learning" can be represented as follows: "suppose we find that a particular language has the property P, ... P is sufficiently abstract and evidence bearing on it sufficiently sparse and contrived so that it is implausible to suppose that all speakers, or perhaps any speakers, might have been trained or taught to observe P or might have constructed grammars satisfying P by induction from experience. Then it is plausible to postulate that P is a property of the language faculty, that language conforms to P as a matter of biological necessity" (Chomsky, 1976, p47).

show that language learning may be addressed from a different perspective rather than assuming an innate set of prescriptions (Elman et al., 1998; Tomasello, 2003; Tomasello and Bates, 2001). Instead of setting certain grammatical parameters in a top-down manner, children utilize general cognitive mechanisms to construct their language systems in a bottom-up fashion.

1.1.2 Cultural evolution of language

As shown above, some biological view for language evolution focuses on a nativist account for language acquisition and language universals. This view is subject to challenge on the validity of its assumption that the advantage of language and language faculty would directly determine the fitness of the human being, as exemplified in the discussion on the subjacency constraint above. In comparison, the cultural view for language evolution does not face such a problem. More importantly, opposed to the nativists' view which advocates a wired-in language faculty to be responsible for a complex human language, more and more attention has been paid to explore the possibility of language emergence from a perspective of cultural evolution. It is believed that the various complex structures we observe in modern languages are the results of a cultural evolution process. The structures emerged through cumulative interactions and cultural transmission, which Hurford (1990) calls glossogenetic language evolution. Such a view has gained prominence in the study of language origin (Knight et al., 2000; Wray, 2002, among many others).

It is the languages themselves which undergo linguistic selection, rather than the language users which undergo biological selection. Languages themselves adapt to aid their own survival (Hurford et al., 1998a). Languages with different linguistic features may have different fitness and undergo competition and selection. Communication and learning exert selection pressure for the reproduction of these features. At the time of Darwin, such a view had exsited that language evolves in a similar manner as biological organisms under the pressure of natural selection. In Descent of Man (Darwin, 1871), he notes that:

> "As Max Müller has well remarked: 'A struggle for life is constantly

> going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue'. The survival or preservation of certain favored words in the struggle for existence is natural selection." (Chapter 3)

Christiansen (1994) highlights the view that language itself evolves to be adaptive, as opposed to that humans evolve to be adaptive for using language, though his comparison of the importance of these two types of evolution may not be appropriate since they are not comparable:

> "What is often not appreciated is that the selective forces acting on language to fit humans is [sic] significantly stronger than the selective pressure on humans to be able to use language. In the case of the former, a language can *only* survive if it is learnable and processable by humans. On the other hand, adaptation towards language use is one out of many selective pressures working on humans ... Thus, language is more likely to have adapted itself to its human hosts than the other way round. Languages that are hard for human to learn simply die out, or, more likely, do not come into existence at all." (Christiansen, 1994, p126)

We can view a language as a composite with a set of linguistic items each of which undergoes selection and reproduction during language use and learning. Mufwene (2001) conceives there is a "feature pool", including units and principles of a language such as sounds, morphemes, words and idioms, as an analogue of "gene pool" in biology. Different languages, dialects, or idiolects in contact at different levels contribute to the feature pool. Learners construct their language by selecting different features from the pool and combine them together, often accompanied with modifications of these features. These individual linguistic features can be viewed as a kind of "meme", what has been coined for the replicating units in cultural evolution as a counterpart of "gene" in biological evolution (Blackmore, 1999; Dawkins, 1976). Similarly, Croft (2000) proposes a term called "lingueme" as the linguistic replicator, on analogy with "gene" and "meme" (p28)[3].

In the discussions on the adaptability of linguistic features or structures, we have to first attend to the controversy on what is the primary function of language: for representation or for communication. Bickerton (1990, p5) argues that "Language ... is not even primarily a means of communication. Rather it is a system of representation, a means for sorting and manipulating the plethora of information that deluges us throughout our waking life". The opposite belief is that "linguistic adaptation arose first in the interest of enhancing communication and secondarily in enhancing or refining thought" (Jackendoff, 1999, p272). The latter view seems more important to us. Without the impetus to express internal representations to share with others, and without being affected by communication, it is hard to imagine the development of a complex internal representation. It will also become paradoxical when taking language acquisition into account. Language acquisition has to be situated in the linguistic environment. Without a linguistic environment, children cannot learn language, as attested in several feral children cases (e.g., Curtiss, 1977). If no communication happens, there will not exist any linguistic environment for learners. Therefore, we recognize the primary, though not the only, function of language is to communicate. Language evolves under the selection forces based on the fulfillment of its communicative function.

1.1.2.1 Functional explanations

Under the umbrella of a general communicative function, linguists have identified various detailed functional constraints to explain what we observe from various synchronic language universals and historical changes. The following summarizes some of the constraints which are often cited, based on Kirby (1999a):

1. economy: the linguistic forms which are used commonly will be shortened to simplify the utterance for the sake of economy (Croft, 2000). This principle is also

[3] Croft (2000) defines lingueme as utterances which are assumed as the units of selection in language evolution, while Mufwene (2002) objects to equating lingueme with utterance, because "speakers do not learn to reproduce other speakers' utterances. Instead they learn (some of) the units and principles that enable those speakers to produce utterances" (p47).

called the "Zipf's Law": "High frequency is the cause of small magnitude" (Zipf, 1935, p29). Many abbreviated forms gradually replacing the original forms are the result of economy principle at work. For example, 'ad' for 'advertisement', and 'tv' for 'television'.

2. iconicity: the structure of the language reflects the structure of the experience of the speaker, including his perspective imposed on the world (Croft, 2000, p164). The distance between constituents in linguistic forms implies the conceptual distance between concepts signified by those constituents (Haiman, 1983). Bybee (1985) suggests that the formal closeness of an affix to its stem iconically reflects the conceptual closeness between the affix and the root. In the explanation for the universal of the order of derivational and inflectional affixes: "if both the derivation and inflection follow the root, or they precede the root, the derivation is always between the root and the inflection" (Greenberg, 1963, p93). For example, in 'computations' the derivational affix -*ation* comes before the inflectional affix -*s*[4].

3. processing: the linguistic structures evolve to make processing easier. Cutler et al. (1985) suggests that the preference for suffixes as opposed to prefixes is due to the serial nature of speech, i.e., left-to-right. Listeners will process the speech more easily if the salient information is placed early, which may result in the advantage of suffixal structures.

4. pragmatics: some linguistic structures are natural consequences of some characteristics in actual language use. For example, there exists an implicational universal that the existence of first or second person reflexive forms in a language implies the existence of third person reflexive forms. Comrie (1981, p28) suggests that third person pronouns are regularly non-coreferential in an utterance while first and second persons are, so that it is more important functionally to make the reflexive/nonreflexive distinction in third person than in first or second person referents, in order to make co-referentiality unambiguous in third person

[4]However, there are counterexamples, such as 'markedly' where inflectional affix -*ed* comes before derivational -*ly*.

referents;

5. discourse structure: the linguistic structures are shaped by discourse structures. Du Bois (1987) explains the special marking for subjects of transitive in ergative case systems as that in discourse, transitive subjects are usually given, and therefore pronominal, so that most clauses only involve one or zero nominal arguments;

There are many other lists of similar functional considerations. For example, Slobin (1977) proposes that there are four “charges” governing the use of language: (1) “Be clear”; (2) “Be humanly processible in ongoing time”; (3) “Be quick and easy”; (4) “Be expressive”. Croft (1990, p254) notes that “economy” and “iconicity” are the two manifestations of efficient adaptations of language, and that they go under the name of processing. Similarly, Kirby (1999a) suggests that the various functional explanations for language universals given above can be reduced to considerations of language processing, which may generally include parsing for the listener (i.e., the mapping of an acoustic wave onto a corresponding message and interpretation) and production for the speaker (i.e., the mapping from communicative intention to articulation) (p13)[5].

In the above discussions, the functional explanations mostly concern morpho-syntactic universals. Similar discussions have been applied to phonological universals as well. In fact functional constraints for speech articulation and perception at a low level are better grounded. A series of studies on sound systems, particularly vowel systems have demonstrated how functional explanations can make predictions consistent with the universals found in languages (Liljencrants and Lindblom, 1972; Lindblom, 1986, 1998).

Usually, functional studies for language universals first hypothesize some criteria and a functional measure from studies on certain languages, and then, they go on to test such hypothesis by applying these criteria to some other languages as an evaluation of the predictability of the hypothesis. However, most of these studies do

[5] Here we simplify the discussion by only dealing with verbal communication; similar functional considerations can be applied to written language as well.

not address the question of how the structures are formed in a language in the first place.

To address this issue, some studies take a further step to investigate what the dynamical process could be, by quantifying the hypothesis and simulating the process with computational models. Kirby (1999a) reports several computational models which employ Hawkins (1994)'s performance theory to explain language universals following this line of thought. The models first assume certain competing variants of a structure already co-exist in a population of language users in the beginning. The variants have different fitness according to a given functional measure, and those which have better fitness will have a higher chance to be passed on as a result of linguistic selection. Finally, the whole population converges to the adaptive and functionally advantageous variant. This is similar to the traditional view of natural selection in biological evolution, where genes in more adaptive organisms have higher chances to be transmitted to the next generation. However, there are significant differences between natural selection and linguistic selection, in terms of direction and mechanisms of transmission (Cavalli-Sforza and Feldman, 1981; Croft, 2000; Mufwene, 2001), which will be discussed in the next section.

Similar to the study of language universals, the selectionist view has been long applied to the study of language change. Wang (1969a) uses competition between phonological rules to explain irregularities in language change, and later Wang (1982b) advocates the principle of variation and selection from biological evolution to explain language change: language change is the result of a competition between variants, and is implemented through lexical diffusion (Chen and Wang, 1975). Lately Croft (2000) proposes an *utterance-based* selection mechanism to explain language change, in which the selection forces are not limited to the learners as proposed by other scholars such as Christiansen (1994); Deacon (1997), but also include more general constraints on adult speakers.

1.1.2.2 Controversies on functional explanations

The functional perspective may face a possible criticism, that is, the explanations seem to be *ad hoc* (Kirby, 1999a), or often constructed "after the event" (Lass, 1980). Kirby considers that some universals cannot be explained purely by functional considerations, but may derive from some innate linguistic constraints. For example, in the explanation for the universal order of the prepositional noun-modifier hierarchy, he suggests that animacy may be an innate grammatical primitive. However, it is not clear to us yet whether the animacy distinction being encoded in grammar should be an innate constraint in humans, or whether it is learnt during language acquisition.

In the study of language change, the functional approach may encounter even more challenges. In some cases, it is hard to identify linguistic functional motivations to explain the cause of a change, such as the lexical replacement of "dog" over "hound". Another selection factor, social selection, may be invoked. And very often social factors are considered as an even more important selection pressure in determining language change (Labov, 2001; Nettle, 1999c).

The above functional approach in explaining language evolution is in fact based on one important assumption, which is that some variants are more adaptive than others. Therefore they are selected as they confer higher survival and reproduction rate as in Darwinian evolution. In biological evolution, there is another important theory, i.e., the neutral theory of evolution (Kimura, 1983), which proposes that " ... at the molecular level most evolutionary change and most of the variability within species are not caused by Darwinian selection but by random drift of mutant alleles that are selectively neutral or nearly neutral" (p34). One strong argument for the empirical basis of the neutral theory is that, nucleotide changes that cause no amino acid changes, called synonymous or silent substitution, are found to occur at much higher rates in evolution than those which lead to amino acid changes (ibid, p32).

Some of the models simulating language evolution (Niyogi, 2002; Niyogi and Berwick, 1997) are considered as close to the neutral theory model (Briscoe, 2000a), as these models do not involve any selection bias between the variants which undergo differentiate reproduction only through an acquisition process which is solely dependent on the frequency of the variants present in the triggering data.

Nettle (1999c) considers the neutral model may not be suitable to account for language evolution, because the transmission of genes and linguistic norms are quite different. In biological evolution, once a mutant arises, it will be more or less automatically passed on, while in language evolution, it is not straightforward that the new mutation will be transmitted to the next generation, because language is not inherited but transmitted by children's learning. Therefore for a new mutation to come into the language, it has to overcome two problems: 1) the "averaging problem" (on average, random changes in a continuous linguistic variable sum to zero, leading to no change); 2) the "threshold problem" (when variants first arise, they are too rare to be learnt by new speakers coming into the community, who will always opt for the most common form) (p22-25). For any new linguistic variants to succeed in diffusion and lead to a language change, functional and/or social selection must be present. In other words, language changes are hardly neutral. Instead, they have to depend on certain forms of selective advantage, either social or functional.

1.1.3 Co-evolution of language and brain

The above two perspectives on language evolution in fact address different aspects of language. The biological perspective focuses on investigating how our brain is innately configured for language, while the cultural perspective focuses on how language has evolved to fit our brain. These two views can be reconciled by considering that brain and language co-evolve in the phylogenetic timescale. Deacon (1997) elaborates such a view in great depth. The main idea is that "the adaptive advantage of language communication would have provided selection for progressively internalizing certain crucial features of language structure in order to make it more efficient and more easily acquired" (p328). This is often considered as a case of the Baldwin effect (Baldwin, 1896), which emphasizes that learning and behaviors may affect evolution through genetic assimilation (Waddington, 1942).

The Baldwin effect in language evolution has been suggested in the nativist proposal as in Pinker and Bloom (1990). Christiansen (1994) challenges this view, by noting that the rate of language evolution is much higher than that of its hominid hosts, and therefore the hominid learning mechanisms involved in language learning are "chasing" a continuously moving target. Therefore he considered that the Baldwin effect cannot account for the massive innate language endowment assumed by Pinker and Bloom. Deacon (1997) gives more extensive arguments explaining the difference between his and Pinker and Bloom's view. Deacon points out that the two views differ "in the description of what has and has not been internalized in this way, and more specifically in what theoretically could have been internalized" (p340). According to Deacon, "No innate rules, no innate general principles, no innate symbolic categories can be built in by evolution... Only certain structural universal features of language could have become internalized as part of a 'language instinct', and these turn out *not* to be those that are most often cited as the core of a Universal Grammar" (ibid, p339). Because whether in Darwinian or Baldwinian evolution, for a new trait to become a regular feature of a species, "the specific adaptive demands imposed by the environment must remain unchanged over hundreds or even thousands of generations" (ibid, p328). "For a language feature to have such an impact on brain evolution that all members of the species come to share it, it must remain invariable across *even the most drastic language change possible*. . . . For genetic assimilation to take place, this persistent aspect of language must also impose *consistent invariant demands on neural processes* . . . in the same way *in all brains* under all conditions" (ibid, p329-330, emphasis original). However, it seems that no grammatical or syntactic universal meets these criteria, especially the second one which requires a constant selection pressure for evolution.

Deacon (1997) argues that ". . . the best candidates for innate language adaptations turn out to be some very general structural characteristics of the primary language medium itself, speech, and the computational demands this medium imposes when it comes to symbolic analysis" (p339). Wang (1982a) posits a set of possible candidates, such as what involve the perception of patterns in the frequency and temporal domains, the coding and storage of events and objects at different levels of memory, the manipulation of various hierarchical mental structures, and so on. A more detailed list of innate capacities will be discussed in the next section. In recent years there has been a number of computer models which demonstrate the Baldwin effect in general (Hinton and Nowlan, 1987) and specifically in language evolution (Munroe and Cangelosi, 2003; Turkel, 2002).

1.2 Self-organization: A new scientific paradigm

This thesis adopts a new perspective in studying language evolution, which is to view language evolution as a *self-organizing* process. Before a full exposition of this view, I will give a brief introduction of self-organization in general.

In the later half of the twenty century, there has been a paradigm shift in scientific investigations. In his book The Origins of Order, Stuart Kauffman (1995) points out:

> "The past three centuries of science have been predominantly reductionist, attempting to break complex systems into simple parts, and those parts, in turn, into simpler parts. The reductionist program has been spectacularly successful, and will continue to be so. But it has often left a vacuum: How do we use the information gleaned about the parts to build up a theory of the whole? The deep difficulty here lies in the fact that the complex whole may exhibit properties that are not readily explained by understanding the parts. The complex whole, in a completely non-mystical sense, can often exhibit collective properties, 'emergent' features that are lawful in their own right." (page VII)

1.2.1 A brief introduction to the theory of self-organization

Self-organization has been considered as a new perspective in the pursuit of explanations for the emergent features mentioned by Kauffman above. The philosophy of self-organization has been attested in early thinkers. The great philosopher Immanuel Kant is considered as the first to use the term "self-organization" in the discussion on the nature of living organisms in his The Critique of Judgement (Capra, 1996). Kant argues that organisms, in contrast with machines, are self-reproducing, self-organizing wholes. In a machine, the parts only exist for each other, in the sense of supporting each other within a functional whole. In an organism the parts also exist by means of each other, in the sense of producing one another. "We must think of each part as an organ . . . that produces the other parts (so that each reciprocally produces the other) . . . Because of this, [the organism] will be both an organized and self-organizing being" (Kant, 1790/1987, p253, reproduced in Capra, 1996, p21).

In the 1960s, several phenomena of self-organization in empirical physical systems have been studied independently, which led to the convergence of the theory of self-organization. A classic case of self-organization is the "Bénard instability" phenomenon in heat convection (Prigogine, 1980). In the experiment, a thin layer liquid in a flat circular dish is heated uniformly from bottom. At the beginning the liquid remains at rest. However, when the temperature difference between the top and the bottom reaches a certain critical value, patterns of hexagonal cell, called "Bénard cells", suddenly appear. Prigogine and his colleagues found that the pattern is far from equilibrium in the dissipative system[6] as the liquid is being heated continuously.

Another classic example in physics is the laser. In a laser, under normal conditions, atoms are excited by external pumping and emit an incoherent mixture of light waves of different frequencies and phases. But under some specific circumstances, the emitted light waves become coherent and form a single wave train. It is found that the emergence of such a coherent state is the result of the interactions of the many particles in the system, when the external pumping reaches a certain critical value (Haken, 1984).

Physicists have been excited by the discovery of self-organization as a mechanism to explain the emergence of various spontaneous patterns in physical systems, and have been developing rigorous definitions and analytical models of their own. Simultaneously, self-organization as a new perspective has widely been adopted by scientists

[6] A *dissipative system* (or dissipative structure) is a kind of highly ordered, stable, open system which is operating far from thermodynamic equilibrium within an environment that exchanges energy, matter and entropy. A dissipative system is characterized by the appearance of stability, but is continually changing. A simple example is a whirlpool: a similar shape is maintained, while water is continually moving through it. More complex examples include lasers, Bénard cells, and even life itself (adapted from www.wikipedia.org).

in many other disciplines dealing with complex systems observed in nature and human societies (e.g., Bonabeau et al., 1999; Jantsch, 1981). Especially, many biological systems in the natural world have been studied as examples of self-organization (Camazine et al., 2001).

The self-organizing pattern in physical systems is mostly built *of* the components themselves in the system. Similar phenomena are found in biological systems, such as schools of fish moving in a coordinated manner, groups of fireflies flashing in unison, or a multitude of raiding army ants. There is another type of self-organization in biological systems, in which the self-organized structures are built *by* organisms, for example, the comb patterns in honeybee colony, the walls built by ants, the elaborate mounds built by termites, and so on. In human societies, these two types of self-organization are both well attested. Language can be considered as a case of self-organizing system built *by* humans.

The mechanism of self-organization in biological and social systems differs from those in physical systems in two ways (Camazine et al., 2001, p12). First, there is a great complexity of the components in biological systems. The interacting components in physical systems are inanimate objects such as water molecules and sand particles, while in biological systems the components are living organisms such as ants and fish, not to mention humans, whose behavior is of greater inherent complexity as a result of sustained learning. The second difference lies in the nature of the rules governing interactions among the components. The rules in chemical and physical systems are solely physical laws related to gravity, surface tension, etc. But in biological systems, in addition to physical laws, the rules are determined by the properties of the components which are subject to natural selection. In other words, the rules of interaction are changing as shaped by selection, so that the organisms evolve to be adaptive. However, these rules do not need to be complex. Interactions among components can be surprisingly simple, even when extremely sophisticated patterns are built (ibid, p13).

The following definition taken from Camazine et al. (2001) is representative in capturing the essence of "self-organization":

> "Self-organization is a process in which pattern at the global level of a system emerges from numerous interactions among the lower-level components of the system. Moreover, the rules specifying interactions among the system's components are executed using only local information, without reference to the global pattern." (p8)

Despite the diverse appearance of self-organizing systems in various areas, there are several common properties shared by these systems (Bonabeau et al., 1999):

1. Emergence: The creation of spatiotemporal structures which arise unexpectedly from interactions among systems' components, rather than a property imposed on the system by an external ordering influence. An emergent property cannot be understood simply by examining in isolation the properties of the components, but requires a consideration of the interactions among components. The whole is greater than the sum of its parts.

2. Multistability: The possible coexistence of several stable states. Because structures emerge by amplification of random deviations, any such deviations can be amplified, and the system mconverges to one among several possible stable states, depending on the initial conditions.

3. Phase transition: The behavior of a self-organizing system may change abruptly and dramatically. There is no need to invoke a qualitative change of behavior in the components during the transition from the non-coordinated to the coordinated phase.

In a self-organizing system, especially in biological systems, there are a few basic ingredients which account for the above features of the system, as shown in the following list as a summary from Bonabeau et al. (1999) and Camazine et al. (2001). We will use the term "agents" (a term often used in computational modeling), instead of "components" used above, to refer to animate subunits in biological systems, which is closer to our discussion of the self-organization in language.

1. Positive feedback: An initial change in a system reinforces the change in the same direction as the initial deviation. Self-enhancement, amplification, facilitation

and autocatalysis are all cases of positive feedback. For example, the behavioral rule "I nest close to where you nest" leads to the aggregation or clustering of nesting birds. Positive feedback in biological systems is usually behaviorally coded in agents.

2. Negative feedback: A small perturbation applied to the system triggers an opposing response that counteracts the perturbation. It provides inhibition to offset the amplification from positive feedback and helps to stabilize the system into a particular pattern. Mechanisms such as saturation, exhaustion, or competition provide such negative feedback. In the case of nesting, a rule could be described as "I nest where others nest, *unless the area is overcrowded.* Often negative feedback arises merely as a physical constraint imposed by the environment, instead of behaviors explicitly coded within the agents' genome.

3. Amplification of fluctuations: fluctuations such as random walks, errors, random task-switching, etc. act as seeds from which patterns and structures are formed. Randomness is often crucial, since it enables the discovery of new territory. For example, foragers may get lost because they follow trails with some level of error, but this gives a chance for finding new unexploited food sources.

4. Multiple interactions: agents interact continuously with each other, either directly or indirectly. In these interactions, agents only make use of local information obtained from their own activities as well as of other's activities, and have no access to the global pattern.

In the following, we will introduce the honeybee comb in detail as an example of self-organizing system. The phenomenon under investigation is highly interesting. More importantly, the methodology that the researchers adopt and the process of investigation are very suggestive and stimulating for the study of language evolution.

1.2.2 Honeybee comb as a studied case and its implications

In a honeybee comb, three types of material are stored, that is, immature brood, honey and pollen. The arrangement of these materials forms a characteristic concentric

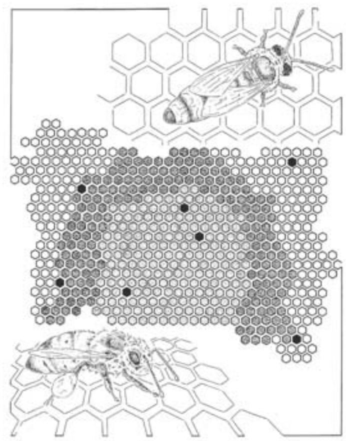


Figure 1.1: An illustration of the concentric pattern in a honeybee comb. The center of the figure presents an area of comb near the center of the colony. Symbols: lightly-stippled cell = brood, darkly-stippled cell = pollen, unstippled cell = honey, and dark cells = empty. In the upper right of the figure is the queen and in the lower left is a forager returning from the field with a pollen load. Reproduced from Camazine et al. (2001), p308 (©Bill Ristine 1998).

pattern - a central brood area, a surrounding band of pollen, and a large peripheral region of honey. Such a pattern appears to be an adaptive structure with several advantages. For example, the compact brood area may help to ensure a precisely regulated incubation temperature for the brood, and facilitate efficient egg laying by the queen and the brood care by the nurse bees. Arranging the brood in the center may also provide a better protection against predators. The location of the pollen, in a band adjacent to the brood area, may allow efficient feeding of the nearby larvae.

To explain how the adaptive structure emerges, there are several hypotheses. The first is the blueprint hypothesis. The behavior of bees may be genetically determined. They by instinct know how to organize the comb in a concentric way, i.e., the central portion of the comb is reserved for brood and pollen, and honey is to be placed peripherally. The task for proving the blueprint hypothesis will be to look for the genes in bee responsible for the concentric pattern and explain how such genes come into existence under natural selection. The second hypothesis assumes the existence of a template: there may be a temperature gradient from the center of the comb to the periphery, and the bees simply follow the temperature template to organize the materials in the comb. The template hypothesis still requires certain genes to be responsible for bees' behavior, but the relationship between the behavior and the genes is less direct than the blueprint hypothesis assumes.

However, these two hypotheses cannot account for all observations. For example, pollen and honey are often deposited throughout the comb, which is specially clear when an initially empty comb is placed in the colony. Also when some cells in the central area become empty after the developing bees mature and leave, these cells are frequently filled with pollen or honey. Such observations have probably been ignored when the blueprint hypothesis is assumed. They will be considered either as an aberrant behavior of a few bees, or abnormal situations that occur only when the rate of honey or pollen deposition is high and the storage cells are not enough in the peripheral areas.

However, these observations have intrigued scientists to hypothesize a different mechanism to account for comb-pattern formation. Entomologist Scotts Camazine of Pennsylvania State University has carried out extensive studies and obtained stimulating results. The following is a summary of his research and findings.

Rather than assuming the individual bees have an innate blueprint of the global pattern of the comb, he starts out his investigation from the self-organization perspective. The basic idea is that the pattern is an emergent property from a dynamic process involving bees' local interactions. "...it is difficult to imagine that any individual in the group possesses a detailed blueprint or plan for the structure it is building. The structures are orders of magnitude larger than a single individual and their construction may span many individual lifetimes. It seems far more likely that each worker has only a local spatial and temporal perspective of the structure to which it contributes. Nonetheless, the overall construction proceeds in an orderly manner, as if some omniscient architects were carefully overseeing and guiding the process" (Camazine et al., 2001, p309). Taking this view, he set the main direction of investigation as examining the lower level individual components of the pattern-formation process - egg laying, pollen and honey deposition, honey and pollen consumption, brood growth, etc.

Studies on tracing the above activities of bees have demonstrated the lack of a blueprint governing the behaviors of individual bees. First, for the brood deposition, the queen does not purposely confine her selection of cells to lay eggs in the center of the comb, but rather she moves unsystematically. But her movement follows one constraint, which is that she prefers to select a cell within a few cell lengths of another brood-containing cell. In other words, the queen prefers to lay eggs closely. Second, it is found that bees do not select particular regions for pollen and honey. This is revealed by an experiment, in which an empty frame is placed in a hive at the beginning of the day before foraging and removed at the end of the day. It is found that the pollen and honey are in fact deposited randomly in cells throughout the comb. Third, the consumption rate of pollen is much higher than honey. Over 90% of the collected pollen is consumed while only 60% of the honey is.

After gaining the information for the local actions of bees from experiments in the field and in literature, a Monte Carlo simulation model has been designed to simulate the formation of the comb pattern. The parameters of an early version were as follows (Camazine et al., 2001, p320)

1. The queen starts from the center of the frame and lays an egg in any empty cell that is less than 4 cells to the next nearest brood cell.
2. The maximum egg-laying rate was set at 1 egg/min, 24 h/day.
3. Only one egg is laid per cell.
4. After 21 days the brood cell is vacated.
5. Honey and pollen are deposited in randomly selected cells that are empty or partially filled with the same substance.
6. Honey and pollen are deposited in cells during daylight hours, 12 h/day.
7. The average capacity of a honey or pollen cell is 20 loads.
8. Honey and pollen are removed from randomly selected cells.
9. The ratio of honey removal to input is 0.6.
10. The average ratio of pollen removal to input is 0.95.

11. The average ratio of pollen input to honey input is 0.2. This ratio is varied over the course of a simulation to match the natural variation in pollen availability.

Unexpectedly, the simulation based on the above settings did not produce a concentric pattern. The compact brood area did not appear and the pollen was scattered throughout the comb. The failure of the simulation led to the reappraisal of the above assumptions and it was realized that the condition (8) above was not correct. Further field experiments were then performed specially to address the question of how honey and pollen are removed from the cells. Two experiments were carried out: one experiment traced the percentage of empty cells in a short term period, and the other compared the occupancy rate of cells located near the brood area versus peripheral cells. It was found that the removal of honey and pollen are not random. A pollen cell which is completely surrounded by brood is emptied ten times as quickly as a cell without brood neighbors. So the condition (8) above was modified as follows:

8'. Both honey and pollen are removed preferentially from cells near brood.

Incorporating the new constraint into the simulation, the concentric pattern readily emerges. But it is a gradual process. At the beginning, starting from an empty frame, pollen and honey are scattered in the comb, and interspersed with egg cells. This early stage shows a disorganized pattern. But later, a compact central brood area is formed, and honey spread in the peripheral region, while a narrow band of pollen is adjacent to the brood area. One of the simulation result is shown in Figure 1.2.

While the Monte Carlo simulation has successfully produced the concentric pattern, the model has incorporated much detailed information which makes the model as nearly complex as the real system itself. It is hard to analyze which parameters are essential for the pattern formation process, and which are unimportant for the system behavior. Camazine and his colleagues then developed two other abstract models, a cellular automaton model and a differential equation model, which simplify the system so as to make it possible to analyze the effects of different parameters. It is found that the important factors are the following relationships: the queen deposits her eggs near other brood cells; the input of honey is greater than that of pollen; the removal of

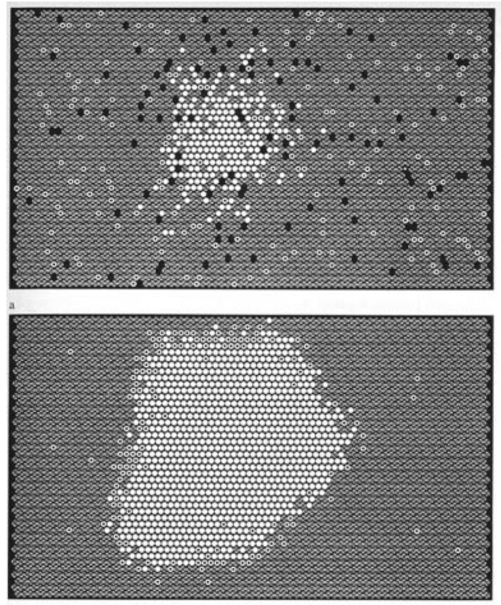


Figure 1.2: A computer simulation result showing the pattern formation on the comb. (a) day 1 (b) day 22. Symbols: open cell = brood; black cells = empty; open cells with black dot = pollen; darkly stippled cells = honey. Reproduced from Camazine et al. (2001), p325.

honey and pollen is greater near brood; the brood-development time is relatively long with respect to other processes occurring on the comb. In comparison, the choice of the particular parameters, such as the neighborhood size for determining preferential removal, the exact value of input and removal of honey or pollen, the egg laying rate, are not important. They will only affect the rate of the pattern formation, but not its general form.

The studies of the honeybee comb present an elegant example of how the self-organization framework guides the research for explaining a complex global pattern. Instead of starting with assumptions that the individual agents have an innate knowledge of the global pattern, the investigation is directed toward looking for lower level constraints. With the help of modeling, the researcher can put into the model what has been obtained from empirical analyses concerning the constraints on individuals' local actions, and test the effect of these constraints by simulation. When the simulation does not produce the expected outcome, the failure will help to identify which part of the initial assumptions does not conform to the real situation, and further field experiment can be carried out with better focus on the problem. Such interactions between modeling and empirical studies is very productive during the investigation.

In fact, not only the formation of the global pattern of the comb can be explained by resorting to lower level processes and conditions, some of those lower level conditions themselves can be further explained as emergent properties. For example, the condition (8'), which says there is a preferential removal of honey and pollen near the brood, does not need to be assumed as the result of bees' behaviors which are genetically programmed, but rather as an emergent feature from bees' local actions. Nurse bees who are responsible to feed the larvae are likely to search for food randomly with as little effort as possible. Assuming the nurse bees move on the comb in a random walk manner centered in the brood area, it will be a natural outcome that the higher emptying rate of the cells will take place near the brood than in the peripheral area.

From this example, we can see that in order to explain the emergence and formation of global patterns in a complex system with interacting agents, we should not start from a position assuming that the agents have the blueprint of the global pattern, like the innate hypothesis here which assumes the individual bees know where to put the honey and pollen, or the queen knows the center region is reserved for eggs. Instead, if we recognize the importance and the impact of interactions between agents, and try to look for the lower level constraints, we are more likely to approach the true explanation in a more effective way. It is believed that this line of thinking will be helpful to the investigation of language evolution.

1.3 Self-organization and language evolution

1.3.1 Existing proposals of self-organization theory applied in linguistics

To view language evolution as a self-organizing process is not at all a new idea in linguistics. Since the 1980s, linguists have applied the concept of self-organization to the study of language and its evolution.

Lindblom et al. (1984) is one of the earliest studies. To address the question about the emergence of phonological universals, Lindblom et al. argue that "postulating segments and features as primitive universal categories of linguistic theory should be rejected, ... While autonomous language-unique phenomena seem by no means biological plausible, demonstrating their existence or proving their nonexistence requires a single research strategy: DERIVE LANGUAGE FROM NONLANGUAGE! Only when such attempts have been exhaustive are we entitled to conclude that we are probably dealing with properties that are unique to language and that should be regarded as major discontinuities or 'mutations'. ... the more successful explanation is the one that more extensively traces the evolutionary roots of linguistic phenomena to preadaptations and extralinguistic factors" (p187). They explain the universal structure of phonemic coding comes from "a random sampling of the universal phonetic space in the presence of performance constraints" (p199), which is to achieve "sufficient perceptual differences at acceptable articulatory costs" (p193). Their study also employs a simple computer program to simulate the emergence of phonemic coding based on the above scenario.

Chen (1989; 1999a) presents a view of self-organization to explain the structure of language systems and the principle of language change. He suggests that this view incorporates Martinet's theory of structural organization and Weinreich, Labov and Herzog's orderly heterogeneity. He considers that the structural instability of a language is the internal trigger for language change to occur; meanwhile, to achieve higher stability and coherence is the aim for language change. Unorderly variation is ubiquitous in the speech community, and when such variation is associated with social factors, it becomes orderly heterogeneity, and leads to a language change. Chen provides a number of diachronic changes and synchronic variations in phonological systems in Chinese dialects as examples of self-organization in language.

Wildgen (1990) presents a brief discussion of self-organization in language from a broader view: "language genesis, language use and grammars show many features of self-organization". He conceives that "at the level of grammars we observe the effect of large population over long historical periods leading to highly stable structures which are clearly outside the conscious control of the participants in this process". This idea is close to the more recent view which emphasizes on the self-organization of language from a population and a historical perspective.

In recent years, Luc Steels is one of the strongest advocates of self-organization in studying language evolution (Steels, 1998). His main hypothesis is that "*language is an emergent phenomenon*". He argues that "language is emergent in two ways. First of all, it is a mass phenomenon actualised by the different agents interacting with each other. No single individual has a complete view of the language nor does anyone control the language. ...Second, language is emergent in the sense that it spontaneously forms itself once the appropriate physiological, psychological and social conditions are satisfied" (ibid). He and his colleagues have developed various computational models to demonstrate how the emergent phenomena come into being. They have applied this theoretical framework extensively in studying the emergence and change of shared grounded meanings (e.g., Steels and Kaplan, 2002), vocabularies (e.g., Steels, 1996a, 1997, 1999), and phonological systems (e.g., de Boer, 1997, 2000) in language.

Meanwhile, there are also a number of studies which are close in spirit to self-organization though they have not adopted the term for their theorizing. For instance, Keller (1994) has proposed to apply the *invisible hand theory* to account for language change. He considers that individual language changes can be viewed as phenomena of the third kind according to Karl Popper (1980)'s classification, which can be characterized as "things which are the results of human actions but not the goal of their intentions" (Keller, 1994, p56). In adopting a functional approach to explain language universals, Kirby (1999a) points out that "the local, individual actions of many speakers, hearers, and acquirers of language across time and space conspire to produce non-local, universal patterns of variation" (p32-33). Mufwene (2001) argues that, "language evolution is in fact a by-product of the contacts which these idiolects have with each other and how they influence each other through the mutual accommodation of speakers. The interactions of these speakers determine the overall system of a communal language" (p194).

In the above proposals of applying self-organization theory, or the like, for the study of language evolution, very often it is not clarified where and how the self-organization takes place. Sometimes it is assumed that there is an abstract "language system" which is self-organizing, such as in Lindblom et al. (1984) and Chen (1989, 1999a). In comparison, in the studies by Luc Steels and his associates, the self-organizing system is more explicit, which is the population language, and/or the individual language. In the following, we will introduce these two lines of studies.

1.3.2 Self-organization in the language system

In Lindblom et al.'s and Chen's proposals of self-organization for language structural universals and language changes, it is an abstract "language system" that self-organizes. Such a view in fact can be found implicitly in the theories of structural linguistics, where language is considered as an autonomous system which evolves to achieve structures of a better kind. A typical example can be found in the seminal paper by Martinet (1952).

Martinet proposes several factors in affecting language change, suggesting "linguistic evolution in general can be conceived of as regulated by the permanent antinomy between the expressive needs of man and his tendency to reduce his mental and physical exertions to a minimum" (ibid, p26). He proposes several accounts, including the preservation of phonemic contrast determined by the functional load of phonemes, the tendency to achieve a structure with higher integration and symmetry, and to achieve least effort or economy. Among these accounts some of them concern the speakers (e.g., the least effort), some concern the listeners (e.g., the phonemic contrast), and some of them concern both parties (e.g., the structural tendency).

We will cite the case of a change in the vowel system of a French dialect named Hauteville, given in Martinet's paper to illustrate the explanation from a system perspective. In Hauteville, a series of vowel shifts happened, which results in a reorganization of the whole vowel system. Before the shift, the vocalic phonemes of normal length were distributed at four levels of height ("aperture" in Martinet's term), as shown in Table 1.1:

After the shift, the vowel system changes to the following pattern as shown in

Table 1.1: The oral and nasal vowel patterns in Hauteville before the shift. Taken from Martinet (1952).

1	i		ü		u	|					
2		e		o		|	ẽ				
3			ɛ			|		ɛ̃		ɔ̃	
4			a			|			ã		

Table 1.2.

Table 1.2: The oral and nasal vowel patterns in Hauteville after the shift. Taken from Martinet (1952).

1	i		ü		u	|			
2		e		o		|			
3		ɛ	ɔ			|	ɛ̃		ɔ̃
4			a			|		ã	

Martinet hypothesizes a scenario for the change between the two patterns. The latter pattern is considered as showing "much more complete integration". He suggests that there were several steps for the change to progress and complete. First, "there are only two oral phonemes for the whole of the two most open orders. Since the opposition of /ɛ/ to /a/ is one not only of aperture but also of depth, it is understandable that speakers should *have tended to neglect* the difference between apertures 3 and 4, which is irrelevant in the rest of the oral pattern, and *to stress* the difference between front and back articulation" (ibid, p21, emphasis added), and the result of the first stage of change is an oral pattern with only three degrees of height, as shown in Table 1.3.

Table 1.3: The oral vowel patterns at the intermediate stage during the change of the Hauteville vowel system.

1	i		ü		u
2		e		o	
3					
4		æ			

Among the nasal vowels in the original system, "/ẽ/ was less fully integrated than /ɛ̃/ or /ɔ̃/ since it was the only unit to combine nasality with aperture 2. . . . speakers would tend to open nasal vowels[7], and therefore /ẽ/ was *exerting a pressure downward.*

[7]Martinet presents an argument why nasal vowels tend to become more and more open, which is that "nasal articulation is detrimental to the clarity of the concomitant oral articulation since it implies that part of the air escapes through the nose and is thus lost for the oral cavity proper. Yet the wider the oral aperture, the more air will flow through it, so that open nasal vowels are likely to be more distinct than close ones".

Table 1.4: The vowel patterns at a later intermediate stage.

1	i				ü				u	|			
2		e						o		|			
3	(ε̃>)		ε				[ɔ]			|	(ẽ>)ε̃		ɔ̃
4				æ		ɑ				|		ã	

In the frame of the nasal pattern, /ε̃/ could not become more open without threatening to impinge upon the domain of /ã/, which in its turn could hardly shift toward the back because of the proximity of /ɔ̃/; /ẽ/ was thus squeezed between /ε̃/ and /æ̃/. Random deviations of /ε̃/ with weak nasalization were apt to be favored since there was no longer any /ε/ occupying its former position. Eventually /ε̃/ was totally denasalized, and /ẽ/ could occupy its former position" (ibid, p21, emphasis added). The resulting system as shown in Table 1.4 is said to be attested in some dialects around Hauteville.

While in the front vowel series a fourth degree of height reappears, the height no. 3 at the back only occurs as a contextual allophone of phoneme /ɑ/. But in Hauteville the "allophones of /ɑ/ have passed to [ɔ], and /æ/ has shifted back to middle position, and thus reaching a pattern given in Table 1.2.

The way that Martinet explains the change from a system shown in Table 1.1 to that in Table 1.2 is conventional in historical linguists' analyses of language change. We note that such an attempt in obtaining descriptions for changes is necessary as the first step, so as to identify the patterns at the global level to be explained.

However, such analysis does not "explain", but rather "describe" language change. We need to go further to see how exactly the self-organization takes place. The descriptions have not addressed the actual process how the change progresses. There is no such a language that aims to achieve a fully integrated pattern, neither is any individual speaker who aims to construct a fully integrated phoneme system. The hypothesized actions such as "neglect", "stress the difference", "exert a pressure", as highlighted by italic forms in the previous paragraphs, have to be grounded with constraints concerning individual speakers and listeners, as well as the social factors during the progress of change. The change does not complete in any single idiolect's life-time. Instead, it is the differences between speakers in successive generations that makes the change progress. We propose that there are two levels of self-organization taking place in this process. In the following section, we will elaborate on this dichotomy of two-level self-organization.

1.3.3 Self-organization in the population and in the individual

The two levels of self-organization come from the consideration of the two forms of existence of language: one is the language in the mind of individual language users, and the other is the language as a collective behavior in a linguistic community.

Two levels of existence

The discussion of the dichotomy of the existence of language at two levels can be traced to Hermann Paul in the late nineteen century (Paul, 1880). He focuses on individual language users as the only object of theoretical significance and considers the "Language Custom" (or 'Sprachusus' in German) in the population as an artifact of the linguist without independent existence (see a critique of this viewpoint in the discussion of language change by Weinreich et al., 1968). Later, Ferdinand de Saussure gives one of the earliest clear expositions of this dichotomy, by pointing out that language has an individual aspect ("la parole"), and a social aspect ("la langue"). "The structure of a language is a social product of our language faculty. At the same time, it is also a body of necessary conventions adopted by society to enable members of society to use their language faculty" (de Saussure, 1910/1983, Chapter 1). He puts more emphasis on the social aspect of language, for "the language is never complete in any single individual, but exists perfectly only in the collectivity" (ibid).

Chomsky (1986) makes a related distinction, that is, between an internalized language (I-language) and an external language (E-language), but this distinction is mainly focused on the language in the individuals. An I-language is considered as tacit linguistic knowledge, so-called "competence", in an individual speaker's mind, which allows him or her to produce or comprehend a particular language; and E-language is the actual use of language as attested in speech or writing, also called "performance". It is considered that "performance" is an imperfect reflection of the competence, and language users may produce "performance errors" time and again, such as slips of the tongue[8]. Chomsky considers the proper target of study in linguistics is the I-language, and emphasizes the systematicity of the I-language and the homogeneity across speakers. These views have been heavily debated (e.g., Hockett, 1968), and we will elaborate in later sections and in Chapter 3.

Later, Hurford (1987) and Kirby (1999a) propose a modification of the I-/E-language distinction which is close to de Saussure's dichotomy: I-language "is the (internal) language as represented in the brains of the population", and E-language "is the (external) language that exists as utterances in the arena of use" (Kirby, 1999a, p318). Recently, Mufwene (2001) proposes an ecological view of language evolution, highlighting the importance of recognizing the two levels of abstraction in the study of language: the idiolect[9] and the communal language. In Mufwene's view, an idiolect is "an individual speaker's system of a language", while a communal language is an "extrapolation from I-languages whose speakers communicate successfully with each other most of the time" (Mufwene, 2001, p2). This dichotomy is close to the I-/E-language distinction formulated by Hurford and Kirby, though Mufwene's communal language as an "extrapolation" is not specified clearly whether it is a union set or intersection set of the idiolects, while Hurford and Kirby's E-Language seems more likely as referring to the union set.

The distinction for the dual existence of language is not recognized in many discussions of language evolution. For example, Christiansen (1994) advocates the view that language is an organism. He rightly advocates the view that "language adapts to its own survival"; however, the neural networks (NNs) in his computational models represents idiolects and the models only demonstrate that certain linguistic structures are more adaptive processed by these idiolects and has not shown how the communal language has evolved. Therefore those models have not provided sufficient support

[8]The distinction between "competence" and "performance" is a controversy between generativists and connectionsts (Newmeyer, 2003). We will not address the detailed argumentation here. Our position is sympathetic to that of the connectionists, considering competence as emergent properties from performance in language use.

[9]According to Weinreich et al. (1968), the term "idiolect" was first introduced in Bloch (1948, p7): "the totality of the possible utterances of one speaker at one time in using a language to interact with one other speaker is an idiolect" (cited in Weinreich et al. 1968, p124).

for the claim yet, though with some extension, the models can show the evolution of communal language when the population of interacting idiolects is implemented. His group has reported some work toward that direction recently. Compared to Christiansen's "language as an organism", Mufwene correctly recognizes the distinction: "a species, not an organism, is a more adequate analog for a language" and "the closest analog to an organism may be an idiolect" (Mufwene, 2001, p149).

Two levels of self-organization

Upon recognizing language's existence at two levels, we propose that language evolution can be viewed as a self-organizing process at these two levels, as depicted in the schema in Figure 1.3.

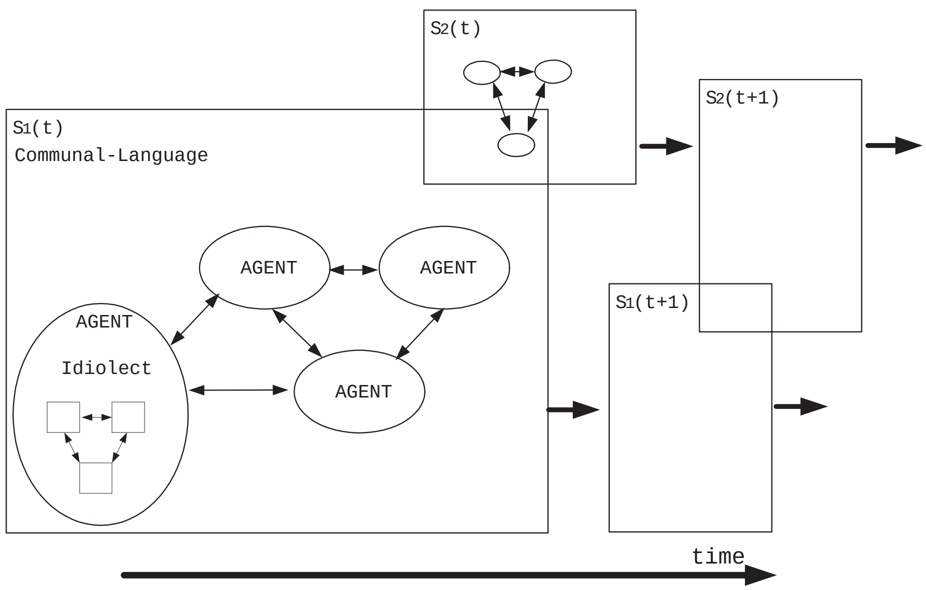


Figure 1.3: A schema of language evolution in the form of two-level self-organization. A communal language $S_i(t)$ at time instant t self-organizes through the interactions among idiolects, each of which self-organizes as well through learning from the linguistic environment.

In the figure, $S_i(t)$ represents one communal language system at some time instant t. Depending on the scale of investigation, these systems can be thought of as linguistic populations such as Chinese or English speech communities, or different dialectal populations of a language such as Mandarin or Cantonese, or different speech communities in a big metropolitan city. We note that the determination of languages or dialects is very often controversial, due to different degrees of resolution and being complexified with political or historical reasons. Moreover, there is no clear-cut boundary between two languages or two dialects in the geographical space at any time instant[10]. To simplify the situation, however, we assume there is somehow an acceptable way to designate the boundary of a communal language, as shown by the square block in the figure. Each communal language evolves along time, say from $S_i(t)$ to $S_i(t+1)$, as indicated in the figure.

A communal language is composed of a number of idiolects in individual speakers, or "agents". The agents interact with each other, through communication and learning. To simulate the language contact situation, we assume that there exist some bilingual agents who belong to two communities at the same time. Thus the two communal languages interact with each other through the bilingual agents shared by them.

Self-organization in idiolect

Within each agent, his linguistic system self-organizes. This is one of the two levels of self-organization. This process mainly happens during the language acquisition period. Adults after the critical period may still have the plasticity for change, but usually with a relatively smaller probability[11]. It has been shown that learning is a self-organizing process at different scales, from the lowest scale in the neural system (Pribram, 1981) to higher scales concerning behaviors (Pribram and King, 1996).

The language learning process is governed by both innate learning mechanisms and linguistic input from the environment. A large number of studies using artificial

[10]While dialectologists can assign individual isogloss for each linguistic feature, often no one single criterion can separate neighboring languages or dialects without violating the separation made by other criteria.

[11]The existence of a critical period is a controversial issue. Particularly, it has been challenged that the critical period is not genetically determined specifically to language (Christiansen, 1994; Deacon, 1997; Elman et al., 1998). While abundant empirical studies show ability of acquiring a second language generally declines, especially for changes in phonological system, however, it has been shown that vocabulary continues to grow and change in adulthood (Sankoff and Lessard, 1975). In specific situations in which pidgin and creole develop in colonial areas, adults may change their language more dramatically to meet the communication pressure, which leads Mufwene (2001) to conclude that adults are the main creator of creole, rather than the children as some scholars claimed (Bickerton, 1984).

neural networks to simulate language learning have shown how certain linguistic structures can be learned in a self-organizing manner (Elman et al., 1998), such as learning lexical meanings and grammatical categories (Li, 2003), learning past tense formation in English (Rumelhart and McClelland, 1986), et. It is assumed that the rich statistical information represented by low-level features is embedded in the linguistic input. Some general learning mechanisms, such as those being simulated in neural network models, can detect such distributional properties and make generalizations from them. There is no predisposed high-level prior linguistic knowledge required, but instead some low-level feature detectors suffice. Moreover, there may not exist symbolic rules, but the generalization is the emergent property from the distributed activation patterns in the neural system.

While the connectionist models are able to capture the nature of self-organization in the idiolects, similar ideas have been employed in other types of models. For example, the exemplar model (Pierrehumbert, 2001) suggests that the words in the mental lexicon are represented as clusters of exemplars, and that the relative weight of exemplars with different patterns may change over time according to the frequency and occurring context. "Representations and the grammatical structure that emerges from them (i.e., the representations) are based on experience with language. New linguistic experiences are categorized in terms of already stored representations, adding to the exemplar clusters already present and, at times, changing them gradually" (Bybee, 2002, p288). This is in fact an alternative description of how the language self-organizes internally through language use in general, in addition to connectionist models mostly focusing on language learning.

Self-organization in the communal language

The other self-organizing process occurs at the level of the population. No single idiolect can represent the communal language. Neither are any two idiolects exactly the same. Therefore the communal language does not equal to any single idiolect, nor the simple summation of the idiolects in the population, but rather an emergent phenomenon from the interactions between idiolects. Interactions here include various types of situations, such as conversations among adults, among children, between adults and children, one-to-many broadcasting either in speech or written format, etc.

The communal language is not a homogeneous system, but rather exhibits an "orderly heterogeneity" (Weinreich et al., 1968). It has been shown by numerous sociolinguistic studies that speakers of different genders and ages or in different social classes demonstrate different linguistic behaviors. Different agents may have different ways to express the same meaning. A consequence of such heterogeneity is that in the communal language, there are many co-existing forms, i.e., alternatives means of saying the same thing (Weinreich et al., 1968).

Learners faced with such heterogeneous linguistic environment will always end up with different idiolects. A further consequence of such heterogeneity in the communal language is that such heterogeneity may also exist within idiolects, as a result of language learning from the heterogeneous environment. Then one agent may use different forms in different interactional situations. These co-existing forms constitute the different "styles" or "registers" of speech as shown in sociolinguistic studies. In these studies, significant differences have been shown in different speech styles, e.g., casual speech, careful spontaneous speech, passages reading, word-list reading and minimal-pair test, sorted according to the degree of speech formality (Labov, 1972). With the presence of such internal heterogeneity, the agents have the freedom and flexibility to respond to their interlocutors by selecting appropriate means of expression[12].

As for the ultimate origin of heterogeneity, it has to do with the social cognition of humans. Language not only plays a role in maintaining convergent communication among group members and providing means for cooperation, but also acts as a way to make differentiation between groups as well, such as man differing from woman, youngsters different from their seniors, higher socio-economic class people differing from lower classes (Weinreich et al., 1968). Such differentiation results in *perpetual novelty*, which is considered as an important feature for complex systems (Holland, 1998). The communal language self-organizes in respond to the continuous influx of novelty.

[12]These selection events appear as the conscious action by the agents. At the same time, there are some unconscious or subconscious selection as well (Croft, 2000).

The emergent communal language conceptually shares some common characteristics as the honeybee comb discussed in Section 1.2.2. The local actions and interactions of the individuals in the community result in the emergence of the patterns at the global level. However, there are two significant differences between the self-organization in the honeybee comb and in the communal language.

First, in the honeybee comb case, individual honeybees do not have any idea about how a comb will look like. Their actions, such as laying eggs, filling or emptying the cells, are independent from each other and without the guidance from a comb's global structure. On the contrary, in the case of language, each individual language user has his own view of a language, i.e., the idiolect. Though the idiolects are not as systematic as linguists show from their linguistic analysis, the self-organization at the communal language level is still very different from that in the honeybee comb. Second, the interactions between individual bees do not have an influence on the principles of individuals' action. The behaviors of individual bees are governed by some innate mechanisms, and has little to do with the on-going actual formation of the comb structure. In comparison, the communal language at the global level does affect the formation of the idiolects. This is evident as attested in the process of language acquisition: children always acquire the language which is used in the community he lives, and adolescents and adults may change their language to some extent when they move to a new linguistic community.

1.3.4 Two levels of existence and three levels of analysis

From the above discussion, we believe it is important to note that language has two levels of existence: the idiolect (I) and the communal language (C), and we have to be clear which level of data is the object of our analysis. There exist three levels of analysis in the current practice of linguistics, as shown in the left side of Figure 1.4. Studies on language acquisition or developmental psychology, and psycholinguistics focus on the data of idiolects; while sociolinguistic studies are more interested in the communal language. Data from these two levels of existence are illustrated in the right side of Figure 1.4. At any instant of time, a language community consists of idiolects as many as the number of people in the community as every one has his own idiolect. The communal language can be viewed as the "common core" shared by all idiolects, i.e., the intersection set $C = \cap_i I_i$, or as the "overall pattern" of the idiolects, i.e., the union set $C = \cup_i I_i$.

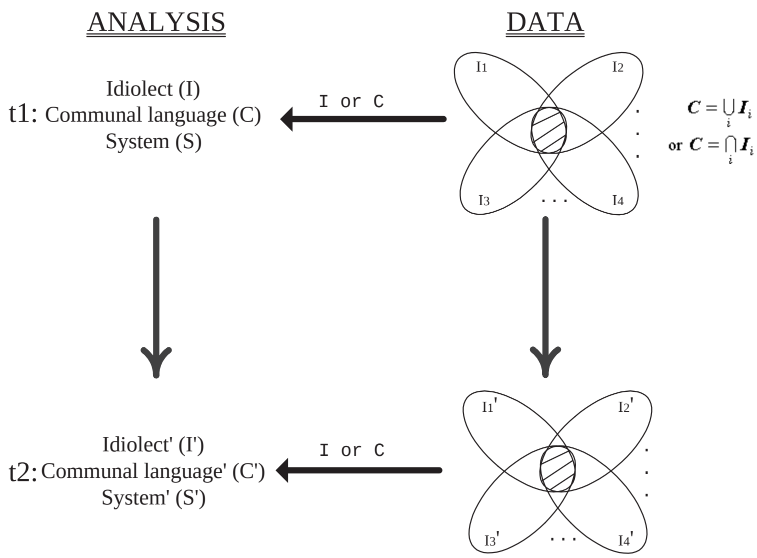


Figure 1.4: Three levels of analysis of language seen at two points of time: System, population (communal language) and individual (idiolect).

Historical linguists, in our opinion, carry out their analyses at another level, i.e., an abstract system level. At this level of analysis, it is not specified which level of data, either the idiolect or the communal language, is examined. Very often the two levels are not distinguished. As a result of the abstraction process, a language system constructed as a representative for a particular historical period is often assumed as "très délicat et très compliqué où tout se tient rigoureusement et qui n'admet pas de modifications arbitraires et capricieuses" (Meillet, 1903-4, p461) ("very fragile and complex, where everything is connected to everything in a very rigorous way, and which does not admit any arbitrary or capricious modification"[13]). As we mentioned earlier, some studies view language itself as a self-organizing system, such as Lindblom et al. (1984). These studies also show a similar position viewing language as being "tout se tient".

[13] I would like to thank Dr. Christophe Coupé for this translation.

However, this way of abstraction usually disregards the distinction between idiolects and communal language, and ignores the heterogeneous nature of language at both levels. For example, the Neogrammarian hypothesis only compares the two end points from their abstraction of the communal languages, and reaches an over-simplified regularity explanation for the change between the two end points. The intermediate stages of how a change progresses in the communal language cannot be revealed by looking at an abstract system without connecting with the idiolects.

1.3.5 A language-ready body - properties of the parts

Adopting a self-organization perspective in the study of language evolution leads us to derive the following questions: (1) what are the parts of the system, (2) what properties do these parts have, and (3) what are the interaction constraints between these parts. With these questions in mind, we will see the necessity in focusing not only on the properties in the parts of the system, but also on the factors determining the interactions between the parts.

Our scope of discussion here is in the self-organization at the level of the communal language. The "parts", or the "agents", in the self-organizing system is the individual speakers. The question (2) above concerns two aspects, one is individual speakers' initial properties or initial conditions, the other is the linguistic knowledge in individual speakers' mind after language learning.

Regarding the initial condition, it is undoubtful that there are some innate capacities which are relevant to language learning. The controversies lie in what they are, and whether they are specific to language. Many capacities required for language processing, which were previously thought as human-specific and language-specific, have been shown not to be so. Many of them are more likely to be pre-adaptations, and can be found in other animals. Wang (1982a) reviews a number of studies reported in diverse related fields such as animal communication, experiments on neural bases for speech production, neuroanatomical asymmetry of the human brain, and perceptual lateralization with spoken and written language, and so on. He reaches the conclusion that "no unambiguous evidence is yet available from any of these areas that any definable part of the human body is specialized exclusively for language or for speech" (Wang, 1991, p122). Hauser et al.(2000) and Hauser and Fitch (2003) summarize recent developments in comparative studies of language processing, specially in speech processing in animals. We incorporate these reviews and compile a partial list of the findings of these studies to illustrate the continuity between animal and human in the domain of biological bases for language:

1. Speech:

    (a) Production: Some other mammalian species also have descended larynx (Fitch, 2002; Nishimura et al., 2003[14]), suggesting that the descendent of larynx is unlikely to be an adaptive product for speech from natural selection as proposed in Lieberman (1984).

    (b) Perception: It has been shown that categorical perception of human speech sounds are also detectable in several non-human species, such as chinchillas (Kuhl and Miller, 1975), macaques (Kuhl and Padden, 1983) and Japanese quail (Kluender et al., 1987), though the categorical boundaries are not as sharp as those in humans. It seems that the perception of speech sounds does not require any language-specific processing mechanism.

    (c) Vocal imitation: Several cetaceans species (Rendell and Whitehead, 2001), especially bottlenose dolphins (Janik, 1999), and some birds such as parrots (Moore, 1996), exhibit a well-developed vocal imitative capacity employed in cultural learning.

2. Conceptual-intentional system:

    (a) Primates can acquire a wide range of abstract concepts under specific training, including tool, color, food, numbers, etc. (Heyes and Huber, 2000; Savage-Rumbaugh et al., 1980; Tomasello and Call, 1997).

[14] Nishimura et al. (2003) found that chimpanzees' larynges descend during infancy, as in human infants. But it only has the first step which is the descent of the laryngeal skeleton relative to the hyoid, while in human there is a second step, the descent of the hyoid itself. The first descent is possibly associated with developmental changes of the swallowing mechanism, and contributes physically to an increased independence between the processes of phonation and articulation for vocalization. This suggests that the descent of the larynx and the morphological foundations for speech production must have evolved in part during hominid evolution, and not in a single shift.

(b) Chimpanzees seem to possess a "theory of mind", at least at some primitive level. They can ascribe mental states to other chimpanzees and read other's intention (Premack and Premack, 2003; Premack and Woodruff, 1978).

(c) Great apes, such as chimpanzees and orangutans (but not gorillas), have been found to show self-recognition behaviors and sense of self-awareness, as revealed by experiments using the Mirror Test (Gallup, 1970; Gallup et al., 2002).

3. Other neural functions or cognitive capacities:

   (a) Non-human primates can compute complex statistical regularities of sequential patterns. For example, monkeys (cotton-top tamarins) can discriminate between sequences of syllables that differ only in the frequency or probability with which they occur in fluent and rapid streams of synthetic human speech (Hauser, Newport and Aslin, 2001).

   (b) Non-human primates can discover abstract rules. Cotton-top tamarins are able to discriminate between novel strings of consonant-vowel syllable and familiar ones, which are with two different structures, AAB and ABB, such as *wi wi di* and *le we we* (Hauser, Weiss and Marcus, 2002a). Some latest experiments also show that tamarins can detect statistical patterns built on non-adjacent regularities[15] (Newport and Aslin, 2004).

   (c) Primates have cross-modal associations. It is found that some neurons in the premotor cortex, the so-called *mirror neurons*, in the ventral premotor area in monkeys, which can get activated both when the monkey performs by himself or watch others perform complex manual tasks, such as grasping (Rizzolatti et al., 1996). Ghazanfar and Logothetis (2003) show that rhesus monkeys can recognize the correspondence between the auditory and visual components of their calls[16].

[15] The same experiments applied to adults (Newport et al., 2004) and tamarins (Newport and Aslin, 2004) reveal species similarities and differences. While tamarins can detect regularities between non-adjacent syllables and non-adjacent vowels, but not non-adjacent consonants, while humans are able to detect non-adjacent vowels and consonants regularities. These findings raise interesting questions on the nature of speech perception and statistical learning in humans and non-human primates.

The above physiological and cognitive capacities found in other animals suggest that there are a set of pre-adaptations which had not evolved specifically for language, but have existed before the emergence of language for other functions and later are recruited for language processing. Not surprisingly, there exist a continuum of these cognitive capacities between humans and non-human animals. The quantitative differences may lead to qualitative differences, a feature often observed in self-organizing systems. For example, the cross-modal association in humans may be much stronger than in non-human species. It is shown that humans are able to map sounds with the shapes or properties of objects (Köhler, 1929/1947). Ramachandran and Hubbard (2001) suggest that the extensive cross-modal associations may provide the neural basis for metaphor thinking and creativity in humans.

Furthermore, it has been shown that there are a number of sensory or cognitive capacities which may be specific to human species. For example, Hauser, Chomsky and Fitch (2002b) point out that though some apes can learn a number of symbols up to 400, their ability of learning words is not qualitatively comparable to a normal child at all. Only humans have a large number of words, which is qualitatively different from other animals in terms of scale, mode, rate of acquisition, and characteristics of displacement. Tomasello (2000) suggests that non-human primates do not use their gestures referentially, but rather to only affect the behavior of others directly. In comparison, humans use communicative signals to convey meanings or information, or to refer to things or direct the attention of others. Also it is recognized that when non-human primates learn, they only focus on the environment event involved, but do not understand the goal of the demonstrator, while in human imitative learning, infants understand other's intentions. In an experiment reported in Meltzoff (1995), 18-month-old infants watch an adult performing certain actions on objects but failing to achieve the final results of the target action (for example, an adult trying to pull two parts of an object apart but never succeeding in separating them). The infants

[16] Ghazanfar and Logothetis (2003) propose that the cross-modal identification of vocal signals shown in primates might represent an evolutionary precursor to humans' ability to match spoken words with facial articulation, as attested in the McGurk effect (McGurk and MacDonald, 1976) where visual perception overrides auditory perception: a sound [ba] is perceived as other sounds, [ga] or [θa] when a visual input of pronouncing a sound [ga] is simultaneously presented.

turn out to be able to produce the target actions of pulling the object apart. This suggests that infants are able to understand the intention of adults rather than just mimicking adults' surface behavior.

Even though there are indeed a range of cognitive abilities assumed to be unique to human, it does not mean they are specific to language. It has been argued that these abilities are domain general. Tomasello (2003) summaries two main sets of cognitive and social-cognitive skills which are particularly important for language acquisition, including *intention-reading* and *pattern-finding* (or *categorization*).

1.3.6 Competing motivations in interactions

The above summarizes the properties of individual language users, which include a number of general cognitive, sensory, motor and mnemonic capacities. According to our dichotomy in terms of self-organization, the other part of interest lies in the constraints governing the interactions between individuals in the processes of language use and language learning. As already mentioned in Section 1.1.2.1, there exist conflicting constraints regarding speaker, listener and learner. As Pinker and Bloom (1990) point out,

> "...there is a conflict of interest between speaker and hearer. Speakers want to minimize articulatory effort and hence tend towards brevity and phonological reduction. Hearers want to minimize the effort of understanding and hence desire explicitness and clarity. This conflict of interest is inherent to the communication process and operates at many levels."

In the discussion of language change, Wang (1976) has stated clearly that

> "...as the biological or social context alters, new requirements arise, and different priorities are given to old requirements. As a function of all this flux, language change may be viewed as an elaborate minimax game whose goal is an optimum response to these diverse requirements." (reprinted in Wang, 1991, p59)

We will dissect the roles of the three parties, i.e., the speaker, the listener and the learner, and analyze what are the possible constraints in these parties during the interactions. By doing so, we may approach a better understanding of the second part of analyzing a self-organizing system, i.e., the constraints of the interaction between the agents.

Speakers' constraint

There are various constraints for speakers at different levels of language processing. At the most basic level of speech production, though it is usually considered that "speech movements are strikingly effortless and highly automatized" (Lindblom, 1983, p233), we can still see the effect of the principle of "least effort" (Zipf, 1949) in shaping speech. The syllable structure in language may be considered as an energy-saving arrangement to resonate with the opening-closing movements of the mouth. The consonant-vowel co-articulation and vowel reduction widely observed in speech are the instances of the economy principle (Lindblom, 1983). In fluent speech, sounds are often under-articulated. Also speakers tend to truncate words to shorter forms, which is also called phonological erosion (Lieberman, 1963; Croft, 1990, p232). Many abbreviated forms such as 'ad' and 'lab' in English prevail over the original longer 'advertisement' and 'laboratory' once being created and quickly get conventionalized. Such minimization of articulation effort can account for some sound changes. For example, assimilation, in which phonetic sequences where the production of a segment becomes more similar to the production of an adjacent segment, is the consequence of reducing distance between two sequentially timed articulatory targets (Lindblom, 1983, p237). The widely observed consonant cluster simplification is another example of the economy constraint in articulation.

The economy constraint on production, assuming that speakers prefer least effort in articulation and shorter forms for efficient communication, can be grounded in the consideration of physiological measures such as energy, time, memory load, etc. Other constraints concern the functions of language use which seem much harder to quantify and more abstract. For instance, Jakobson (1960) proposes three functions of language use (cited in Croft, 2000):

1. Referential function: communication of information
2. Poetic function: creativity/expressivity
3. Phatic function: solidarity/conformity with social norms

Keller (1994, p94ff) gives a more detailed model of the constraints on the speaker's intentional behavior which Croft (2000) considers as an improvement of Jakobson's list[17]. Sociolinguists have studied extensively how these constraints are implemented under various circumstances with respect to social structures such as prestige, social ties, etc. How do these constraints take effect? The essence lies in the facts that language is abundant in redundancy and the grammar of each speaker is heterogeneous. One meaning can often be expressed by more than one form. For example, we can say "I hit him" and also "He is hit by me". Let alone the many synonyms and near-synonyms. The great variability for choice allows the speaker to determine his linguistic behavior according to the particular linguistic environment.

Listeners' constraint

In fact, in many of the functional explanations for language universals, the listeners' constraints are often considered as the primary selection force. For example, to explain the universal structures of vowel systems found in languages, it has been proposed that the systems tend to achieve a maximum or sufficient perceptual dispersion (Wang 1971, Liljencrants and Lindblom 1972, Crothers 1978). In language change, new phonological contrast may arise, such as the nasalized vowel as opposed to oral plain vowel, and the perceptual space becomes more and more crowded. Then listeners' failure of discriminating close sounds will lead to phonological merger and some

[17]Keller's list includes: (1) Talk in such a way that you are most likely to reach the goals that you set for yourself in your communicative enterprise . . . at the lowest possible cost; (2) Talk in such a way that you are understood. (or) Talk in a way in which you believe the other would talk if he or she were in your place; (3) Talk in such a way that you are not misunderstood; (4) Talk in such a way that you are noticed; (4) Talk in such a way that you are not recognized as a member of the group; (5) Talk in an amusing, funny, etc. way. (5) Talk in an especially polite, flattering, charming, etc. way; (6) Talk like the others talk (Talk in such a way that you are recognized as a member of the group. or Talk like the people around you); (7) Talk in such a way that you do not expend superfluous energy.

kind of lexical strategy, such as disyllabification in compensating the homophones.

In studies of syntax, Hawkins (1994) proposes a performance theory to account for word order universals, where he assumes listener, or parser in a more technical sense, has the preference for the "earliest possible temporal access of as much of the constituency information as possible" (p233), which is coined as the "early immediate constituent recognition" (EIC) principle. He develops a method to measure numerically the parsing complexity for a particular tree structure and a particular grammar. Kirby adopts the parsing complexity measured by EIC as a selection force to simulate the emergence of several language universals, such as the correlation between verb-object word order and apposition (e.g., *VO+preposition* vs *OV+postposition*), multiple branching structures (e.g., if the relative clause is comp-initial, then the noun and the adjective both precede the relative clause), the prepositional noun-modifier hierarchy ($Prep \longrightarrow (NRel > NGen > NAdj)$), and so on.

For such proposals focusing on the explanation from the point of view of parsing, there is an implicit assumption that speakers are altruistic (Kirby, 1999a). Listeners are mostly passive in the process of language use, and only the speakers have the autonomy in making active decisions in the interaction. Therefore in order to make the parsing explanation effective, we may need to assume that speakers have the intention to take into account the listener's needs and consciously make their utterance easy to be parsed by the listener. In fact this has been implicitly taken as the assumption in Keller's several maxims of the constraints on speakers' behavior. We can obtain support from empirical studies on speech errors, where speakers are found to make repair or self-corrections (Levelt, 1989, p460-463).

More importantly, one individual is either a listener or a speaker on different occasions. A listening event at one instant will certainly affect the individual's language and mind, and will in turn affect his subsequent speaking behaviors. In other words, there is a correlation between the speaking and listening behaviors, and this correlation may be a better explanation than the assumption of altruistic speaker for why the listener's constraint can play a role as a selection force for language evolution.

Learners' constraint

Children acquire their native languages with little effort and without requiring intended support from the adults. Though it has been shown empirically there are various imperfect learning during the process of acquisition (e.g., Ferguson and Farwell, 1975), the "errors" or innovations in children's grammars seldom survive into adulthood, and their grammars and/or lexicons tend more and more to conform to the conventions or norms of the speech community (Croft, 2000, p47). Therefore children are considered hardly as the main force to induce language change (Croft, 2000). However, from a wider time window, it has been often found that many language changes follow the same pattern as children's imperfect learning, and it has been suggested that these changes may be the result of imperfect learning (Andersen, 1978; Hooper, 1980; Slobin, 1977).

When discussing how language evolves under the selection force of its learnability, Deacon (1997) suggests that "languages must go through the 'filter' of children's reduced associative leaning and short-term memory constraints in order to be passed on most effectively from one generation to the next with a certain degree of fidelity. Children selectively hear some structures and ignore others, and so provide a major selection force for language structures which are 'child-friendly' " (p111). He points out that "the key to understanding language learnability ... lies in a process that seems otherwise far remote from the microcosm of toddlers and caretakers - language change" (p115).

The selection force posed by children's learning is used to account for some of the language universals or distributional features, such as center embedding being a rare structure (Christiansen, 1994; Christiansen and Chater, 1999). From the perspective of language emergence, Kirby (2002a) proposes that transmission through language learning is the impetus to drive language evolution from a holistic to a compositional system. The bottleneck imposed by learning, i.e., the new learners only have the chance to be exposed to a limited proportion of the possible meaning-form pairs, makes the idiosyncratic language hard to be maintained through transmission between generations, while a language with compositional structures will be more successful through the learning bottleneck.

For the analysis of a self-organizing system, we still need to go back to the microscopic level to see what are the constraints imposed by children's learning in determining the local actions of language use. In fact learners are subject to constraints very similar to those on listeners, which concern the ease of processing in comprehension, such as perceptual distinctiveness and memory load economy. As we recognize the heterogeneity in the communal language, one communicative function can be fulfilled by several forms. Therefore the learners play important roles in selecting which forms are more likely to be transmitted and passed on in the language community.

1.3.7 Universals, grammars and exceptions

It is well noted that all identified language universals are statistical phenomena, and very sensitive to the size and methodology of the sampling. We take the study of universals of word order correlations as an illustration below. Dryer (1992) reports his statistical results based on a large sample of 625 languages. He shows that some of the early findings or predictions based on smaller samples are not true. For example, it was believed that the word order for adjective and noun correlates in order with that for verb and object: VO languages tend to be NAdj, while OV languages tend to be AdjN. But Dryer's data show that there is no evidence of such a correlation. Table 1.5 gives the statistical result from Dryer's database. The values in the table are the numbers of genera containing languages of the given type in the given area. From the table we can see that the four types of combinations have very close frequencies, or even show an opposite tendency: the structure of OV structure goes more with NAdj, instead of AdjN as previously believed. Also, noun preceding adjective is both frequent in OV and VO languages. Moreover, the existence of exceptions in those areas where the universal tendency exists with statistical significance should not be ignored. For functionalists and nativists, these exceptions may have to remain unexplained or explained as idiosyncratic historical accidents.

From the perspective of self-organization, the universals across languages are in

	Africa	Eurasia	SEAsia	OC	Aus-NewGui	NAmer	SAmer	Total
OV&AdjN	7	24	2	4	10	8	55	110
OV&NAdj	18	4	5	15	18	14	74	148
VO&AdjN	3	6	4	5	19	3	40	80
VO&NAdj	25	3	12	2	8	5	55	110

Table 1.5: The numbers of genera of languages with different combinations of verb-object order and noun-adjective order in seven areas, reproduced from Table 17 in Dryer (1992).

fact the convergent phenomena which emerge from language use in the different language communities. The languages are not constructed by the language users with innate blueprints. To account for the universals, we need to look for the principles and constraints on the individual language users and on the local interactions between them. To account for the exceptions to the universals, we may obtain some insights from the studies on the concentric pattern in honeybee combs as shown in Section 1.2.2. The exceptions to the language universals are counterparts of the occasional irregular patterns where honey and pollen are scattered throughout the comb. These are natural outcomes in the process of self-organization in the system.

Similarly, the same kind of consideration is applicable to the study of individual languages. In the analysis of the grammar of a language, some linguists usually end up with a list of grammatical rules, and assume these rules are homogeneously present in language users, though in some tacit form. To test the validity of these rules, experiments of grammaticality judgment are often used. However, the results obtained from these tests are usually statistical, instead of categorical, i.e., the rules either exist or non-exist, but are always accompanied by irregularities and exceptions.

Very often, these data are mostly ignored by some linguists, or considered as exceptional cases in which language users have performance errors and have no theoretical interest. There seems to be an underlying assumption that every occurring utterance in any given context should have a specific "determinate grammatical structure involving an integral number of grammatical elements in specifiable structural relations with each other", on which Hockett casts his doubt (1961, p52). He points out that exceptions, such as blends, "are not rare, but extremely common", and "occur not only as 'slips of the tongue' (whatever that means), but also as planned puns, *double entendres*, plays on words, and variously in poetry and advertising" (ibid). He suggests that such phenomena could be used as evidence for "some new and very different theory of the generation of speech, that would provide at once for such 'deviant' utterances and for all 'regular' utterances" (ibid, p53).

In addition to the deviants within one linguistic system which is often abstracted from one or many speakers' linguistic behavior, the languages of individual speakers vary dramatically, as attested in either language fluency, language processing behavior, or language learning progress. Some of the earlier works can be found in the collection edited by Fillmore, Kempler and Wang (1979). For example, Ross (1979) demonstrates how individual language users differ in a grammaticality judgment experiment of 13 sentences. One of the most striking cases is that for a sentence such as "What will the grandfather clock stand between the bed and?" which would appear ungrammatical to generative linguists as it violates the "subjacency constraint". However, as Ross's data showed, among the 30 subjects, 3 judged it as "sounds perfect ... would use it without hesitation", and 2 judged as "less than perfect - something in it just doesn't feel comfortable. Maybe lots of people could say it, but you never feel quite comfortable" (ibid, p137).

Some linguists even challenge the plausibility of the grammar rules and consider that such grammaticality data only have marginal status: "they exist only as re-formed scraps of previous discourses, and, stripped of a context, they elicit intersubjectively shared judgment about grammaticality only to the extent that an obvious context can be reconstructed for them, or that they conform to the more sedimented conventions for constructing discourse, or, it must be added, that they violate or agree with explicit social canons concerning 'good grammar' " (Hopper, 1988, p119). Instead of having a static entity which is fully present at all times in the mind of the speaker, grammar is viewed as "a vaguely defined set of sedimented (i.e. grammaticized) 'recurrent partials' " (ibid, p118), which is provisional in real-time and heterogeneous across individuals. While Derwing and Baker (1977) demonstrate that language users may have some generalized morphological rules in mind, Olson (1977) argues that the presence of rules may differ in different types of language users: "For those speakers whose language has been specialized for such literate activities as philosophical argument and reading and writing prose text, the rules under discussion become part of the speakers' system for generating sentences. For those speakers whose language serves the more ordinary conversational functions of the mother tongue, those rules are not part of the speakers' linguistic system" (p111).

At present, we are far from having a clear picture of how grammars are represented and processed in our brain. However, if we view language acquisition as a self-organizing process, the heterogeneity of individual grammars will be a natural consequence of this process. The "exceptions" attested in language use is again a counterpart to the irregular patterns shown in the honeybee comb.

1.3.8 Rapid evolution as a result of self-organization

Language evolves at a rapid rate. This can be shown by extant evidence from the studies on pidgin and creole languages in which a new language emerges within several generations (Romaine, 1992). We may speculate the time scale for the phylogenetic emergence of language. If we believe that language is an innate endowment specific to human species, due to some genetic mutation which happened in the hominid line, we may have an approximate estimation of the time span for the emergence of human language.

In recent years, there have been converging estimations of the time of the rise of modern human. The research in genetics from mtDNA and Y chromosome analyses suggests that modern humans are descendants from *Homo sapiens* who emigrated from Africa around 150,000 years before present (BP) (Stringer and McKie, 1997). A recent discovery of fossilized hominid crania in Ethiopia (White et al., 2003) suggests the probable immediate ancestors of anatomically modern humans may have existed as early as around 160,000 years BP. But there had been not much sign of language till then.

One often cited estimation for the time of the emergence of a full-fledged human language is around 50,000-40,000 years BP, as suggested by a "cultural explosion" around that time (Klein, 1999). It is shown that a large number of delicate art forms, ritual burials, jewels, highly sophisticated tools and techniques, which suggests the presence of a highly symbolic capacity, have been found in archeological sites which fall into this time window. It is speculated that a full language might have been the trigger for this abrupt and dramatic cultural development[18].

Though these dates are still not conclusive, we may hypothesize that the time window for the development of a full human language from a rudimentary system without syntax, a pre-language[19], may be between 160,000-40,000 years BP, in other words, only about 120,000 years. The unusual rapid emergence of language has posed a special difficulty for the proposal of language as an exclusive product of biological mutation(s), because 120,000 years is just a moment of blink in biological evolution time scale, and it is unlikely for a biological mutation to be fixed in a population, as it usually takes millions of years for a new biological trait to emerge.

Therefore a more plausible picture would be that the biological basis for language might have been available early in *Homo sapiens*, due to some genetic mutations, but such genetic mutations did not brought forth a language organ such as the human eyes. Instead, these mutations may be responsible for the increase of brain size and encephalization (Jerison, 1977), and some new characteristics of the neural systems, for example, a significant increase of coherence of signals in the brain (Calvin, 1996). These changes have resulted in the changes of the general working of the brain, and in turn the general cognitive capacities, rather than something specific to language. These physical changes have provided the biological infrastructure for language to emerge and evolve as a cultural phenomenon. It has been noted that language evolves at a much faster speed than biological evolution. It is through cultural evolution that language could have emerged within a short period of time as 120,000 years.

As a cultural evolution process, the fast evolving speed can be explained as the

[18]We may have some flavor how a cultural explosion could have been resulted from the change of communication means. The so-called "information explosion" which is attested in recent decades may be due to the availability of a new medium of communication, i.e., the Internet, owing to the expeditious development of computer technology.

[19]Bickerton (1991) uses the term "proto-language", but as this term has been used in historical linguistics to refer to the hypothesized ancestral language by reconstruction, we prefer "pre-language" for the stage before a full human language was developed.

"ratchet effect" (Tomasello et al., 1993): "...the cultural traditions and artifacts of human beings accumulate modifications over time - so-called cumulative cultural evolution. Basically none of the most complex human artifacts or social practices - including tool industries, symbolic communication, and social institutions - were invented once and for all at a single moment by any one individual or group of individuals. Rather, what happened was that some individuals or group of individuals first invented a primitive version of the artifact or practice, and then some later user or users made a modification, an 'improvement', that others then adopted perhaps without change for many generations, at which point some other individuals or group of individuals made another modification, which was then learned and used by others, and so on over historical time" (Tomasello, 2002, p331). This scenario is very applicable to language evolution. Our simulation models to be discussed later are consistent with this view.

1.4 Organization of the thesis

To sum up, this chapter has given an overview of the theoretical background of studying language evolution. I first give a brief review of the two perspectives adopted in the field: one emphasizes the biological aspect of language and is interested in explaining what are the biological bases for language and how they have evolved, while the other focuses on the cultural aspect and strives to demonstrate how various complex structures in language can evolve through cultural evolution. The third view, i.e., the co-evolution between biological and cultural aspects of language, is also briefly discussed. Then I start to present the theoretical framework adopted in this study, i.e., self-organization, which has been a new scientific paradigm in explaining emergent phenomena in various complex systems. A study on honeybee comb is used to illustrate how effective the perspective of self-organization is in directing the pursuit of plausible explanations for complex phenomena at the global level by examining lower level characteristics of the system. A self-organization theory advocates a division and integration of the two directions of investigation: 1) the properties of the individual components of the system, and 2) the constraints and effects of interactions between the components. In such a framework, the two existing lines of study from biological and cultural perspectives can be naturally integrated.

In earlier sections of this chapter, I have introduced three levels of analysis in linguistics adopting a self-organization framework. One level of analysis is to assume there is an abstract system which is self-organizing, without attending to the reality of the process or without elaborating how the self-organization is actually implemented. The other two levels of analysis addresses self-organization at two levels, i.e., in the population and in the individual, based on the dichotomy of the existence of language at two different levels, i.e., the communal language and the idiolect. In the following chapters of this thesis, I will go further to elaborate how the two lines of self-organization are examined in empirical studies, and in computational models.

Chapter 2 will report an empirical study on homophony and related issues to exemplify the approach of analyzing the language system as self-organizing at the abstract level. First, I will report some first-hand analyses of the synchronic features of homophony in languages after a brief introduction to the nature of homophony and its rise and fall in a language. The analyses include some quantitative measure and comparison of the degree of homophony in three Germanic languages and twenty Chinese dialects. Then I suggest some possible methods to predict the degree of homophony with respect to the complexity of the phonological system of a language, and show that the degree of monosyllabicity and the degrees of homophony are highly correlated. In the latter part of Chapter 2, I discuss some observed phenomena which are interpreted as the result of self-organization within a language system, including the disyllabification in Chinese history, the differentiation of homophones in grammatical classes. In the last section of Chapter 2, I will discuss briefly some more self-organization evidence in the lexicon, reflected in the degree of monosyllabicity and the process of lexicalization.

Following the study of self-organization in the abstract language system, Chapter 3 presents a case study of language change to illustrate the self-organization in the population and in individuals. An on-going sound change in Guangzhou Cantonese, i.e., /n-/→/l-/, is reported in detail, including the methodology of data collection and analyses of three types of synchronic variation observed from the data. The individual differences and heterogeneity in the speech community are highlighted. In particular, I hypothesize the existence of two styles in lexical learning, i.e., categorical and probabilistic learning. Moreover, the observed cases of near-merger leads to some discussion on the disparity between production and perception. I will give some preliminary explanations for the observed uni-modal pattern in the diachronic change profile, and discuss further about the difference between dialectal contact and internal change based on this case study. Finally the coexistence of variants is addressed regarding its presence and persistence.

After the two empirical studies, I will go to the second part of the thesis, which is about the application of computational models in studying language evolution. In Chapter 4, I will first give an overview of the computational modeling in this area. I categorize the existing models into three time scales and four levels of resolution and discuss briefly with some models as illustration. Then I will introduce the agent-based modeling approach, which is widely used in the current computational studies. Our own studies reported in this thesis all adopt this framework. I will discuss in depth the importance of recognizing the disparity between production and perception/comprehension in modeling. The last part of Chapter 4 highlights some pros and cons of the computational modeling approach.

Chapter 5 reports two series of computational models which have been explored in this thesis. The first series of models focus on the origin of vocabulary and the existence of homonym. The second series of models deals with language change, which is simulated as an innovation diffusion mainly through successive generations' learning in a population. The effect of social networks in light of the recent development in two new types of networks, i.e., small-world and scale-free networks, is examined in detail. Also I will examine other parameters in the model, such as different functional bias of the innovation, population size, and learning styles, in particular the two types of learning styles discussed in Chapter 3.

Chapter 6 concludes the thesis with a summary of the studies reported and some future research plans. I highlight several important issues addressed in this thesis. First, there exist three levels of analyses in linguistics, that is, system, population and individual. Second, the existence of heterogeneity in the language system and in the population should be taken as the norm. Third, computational modeling is an effective methodology in complementing empirical studies. Four future directions are briefly discussed.

Two sections of appendices are given at the end of the thesis. The first section include a collection of artistic use of homophones in advertisements, and some ambiguity cases caused by (near-)homophones encountered in real life situations, and the lists of homophones of the first 5000 frequent words in three Germanic languages. The second section of the appendices consists of the experimental materials for the fieldwork study of the Cantonese sound change, including the word-list reading sheet and the questionnaire, and the data organized after transcription from the recording tapes and the questionnaires.

Chapter 2

Self-organization in the system: A case study on homophony

The existence of ambiguity is a special characteristic in human language, and homophony is one of the sources of ambiguity. In this chapter, I will examine various phenomena related to homophony in languages and interpret the characteristics of homophones in a language from a self-organization perspective. This study exemplifies analyses at a level of an abstract language system as mentioned in the first chapter.

While it has been a common belief that all languages have homophony in different degrees (Antilla, 1989, Chapter 9.5), so far there have been few attempts to examine the degrees of homophony quantitatively, not to mention cross-language comparison. There have been some lists of homophones, e.g., Higgins (1995) for English, and some case studies on the history of individual homophones or near-homophones, e.g., Bloomfield (1933, p396-398) and Malkiel (1979). However, there has been little discussion on the relation between homophony and other parts of the language system. Moreover, whether homophony affects daily communication has not been addressed systematically. In connection with language change, on the one hand, homophony-avoidance is used to explain cases of language change (Coates, 1969; Gillièron and Roques, 1912; Jespersen, 1922; Stimson, 1966); on the other hand, homophony-avoidance is often considered as a minor or even ignorable factor in directing language change. So far there has been no effective way to measure the adverse effect of homophony on daily communication; furthermore, it is difficult to predict the long term effect on the fate of homophones from the momentary and sporadic cases of confusion. However, we suspect that such long term effect does exist and the language system self-organizes to decrease the chances of confusion as much as possible. Though this should be a dynamic process, we may still obtain some hints from the synchronic distribution of the homophones in the lexicon.

In the following, I will first introduce the general definition of homophony and a brief summary of various mechanisms leading to the rise and fall of homophony in languages. Then I will compare the degree of homophony in several languages, including three Germanic languages and eighteen Chinese dialects, and propose some methods to predict the degree of homophony based on the size of the phonological resource, and the degree of monosyllabicity. To illustrate how the language system self-organizes in response to the presence of homophony, I discuss the disyllabification process in Chinese as an evidence. Also I show homophone pairs tend to differentiate in grammatical classes, as a result of self-organization in the language system. In the last section, I briefly discuss some more self-organization evidence in the lexicon, reflected in the degree of monosyllabicity and the process of lexicalization.

2.1 Some background of homophony

2.1.1 What is "homophony"?

"Homophony" refers to the case where two or more words, which are called "homophones", "having the same sound, but differing in meaning or derivation", according to Oxford English Dictionary (OED). It is easy to come up with examples of homophones, for instance, 'yuan2 yin1' as '原因' ("reason") and '元音' ("vowel") in Chinese, 'sight', 'site' and 'cite' in English, 'père', 'pair', and 'paire' in French. To extend the definition from words to more general forms, /s/ (including variants /z/ and /Iz/) in English can be also considered as homophonous, as the same form has two functions, one to form plural of nouns, and the other to form the third person present tense of verbs.

In human language, the existence of one-to-many mappings between form and meaning is considered as a feature which artificial languages do not possess. While homophony represents a case of one-to-many in speech, there is another type of one-to-many mappings in written language, which is called "homography". Homographs refer to pairs or sets of words which have the same spelling but different meanings, for example, 'bow' in "*bow* to the champion" and "tie the ribbon in a *bow*". When the words have both the same pronunciation and the same written form, that is, the overlap between homophone and homograph, they are called "homonym"[1]. "Homonyms" are two or more words spelled *and* pronounced alike but different in meanings, for example, 'bank' in "the *bank* of the river" and "a *bank* with money". The relationship between "homophone", "homograph" and "homonym" is shown in Figure 2.1.

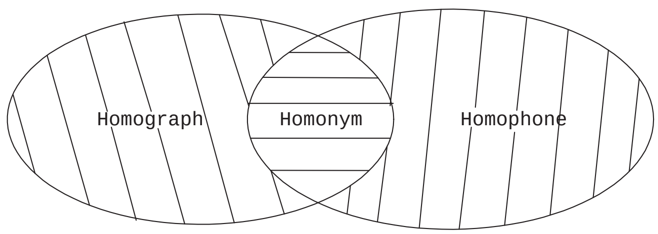


Figure 2.1: The relationship between "homophone", "homograph", and "homonym".

2.1.2 The rise of homophony

2.1.2.1 Homophones from sound change

Most of the homophones arise as the result of phonological merger, a type of sound change which is very common in languages. Words become homophonous once the phonetic distinction that kept them apart becomes lost. In English, for example, 'meat' in Middle English was pronounced similar to 'mate' of modern English, but after the Great Vowel Shift it became homophonous to 'meet' due to vowel raising, though the written forms still retain the distinction.

Chinese is a classic instance in which numerous homophones have come from sound change. In modern Chinese dialects, especially in northern dialects where many

[1] The definition of "homonym" adopted here differs from some others in the literature. Very often "homonym" and "homophone" are used interchangeably, e.g., in Bloomfield (1933) where the definition of homonym is given as that "Different linguistic forms which have the same phonetic form (and differ, therefore, only as to meaning)". Similarly, Saeed (1997) defines homonyms as "unrelated senses of the same phonological word", and "homograph" as "senses of the same written word"; "homophone" as "senses of the same spoken word". In this study, we will adopt the term "homophones" according to our definition.

mergers have occurred, we can find a lot of homophonous monosyllabic morphemes which were distinct in earlier stages. For example, in Mandarin, '急' ("anxious"), '疾' ("illness") and '即' ("immediately") are homophones with the same pronunciation [tɕɪ2] due to the loss of the consonant ending of "p,t,k" in monosyllables[2].

2.1.2.2 Homophones from borrowing

Languages are constantly in contact with other languages to various degrees, and borrowing of lexical items is the most frequent. Very often the pronunciation of the foreign words cannot fit the phonology of the target language and they are adjusted accordingly by the native speaker of the borrowing language. Occasionally the borrowed word may collide with some existing words, or even some words borrowed earlier from other languages, and homophones are thus created. For example, in English the two words 'sheik' and 'chic' are homophones with the same pronunciation [ʃi:k], but they came in different times into English: the former was borrowed from Arabic word 'shaikh' in the sixteen century, while the latter is a French loan word in middle of nineteen century, according to OED.

2.1.3 The fall of homophony

The set of homophones in a language constantly changes. While some new homophones emerge as results of sound change or lexical borrowing, some existing homophones may disappear as some words fall into disuse due to cultural changes, or some get replaced by synonymous words, or some homophones split in new sound changes.

2.1.3.1 Split from sound change

It is often believed that homophones rarely split: whenever they are subject to a sound change, they change together if there is no other interference. For cases that homophones do split, they are often explained as borrowing or analogy. Stimson (1966) gives an example of homophone split: for a sound change from Middle Chinese to Mandarin, a large number of exceptions are attested. Stimson suggests that they

[2]But these consonant endings are still kept in some other dialects, such as Cantonese, in which the three morphemes are pronounced as /gɐp7/,/tsɐt9/ and /tsɪk7/ respectively. The pronunciations are given in IPA, and the transcription of tones follows the 9-tone system (Fan and et al., 1997).

are the results of "tabu interference". When these words followed the regular sound change, they were pronounced as /bi1/ and became homophones to a highly tabu word which means "vagina" or "female genitalia". Then some other readings for these homophones, which came from other dialects or from analogy of the phonetic part of the written character, replaced the taboo pronunciation. As a consequence, some of the original homophones have split and are not homophones any more[3].

Chen and Wang (1975) report a study of homophone split in Chinese dialects using the database DOC (the database will be introduced in detail in Section 2.2.1). Compared with Middle Chinese (MC) which is assumed as a reference point to examine historical development for Chinese dialects, it is found that among the original 491 homophone sets in MC, all modern dialects show split in significantly large degrees, ranging from 17% to 58%[4]. The large degree of homophone split is used as evidence for the theory of lexical diffusion which challenges the Neogrammarian's "exceptionless of laws" in sound change.

Cases of homophone split have been observed in synchronic studies of variation. For example, the homophone pair: noun-'can' and auxiliary verb-'can', have seen a split (/keːən/ vs. /kæn/) in Philadelphia English (Labov, 1989). These examples of on-going split may give us some hints how the split actually progresses. Jurafsky et al. (2002) examine the variations of several function words, such as 'to', 'that', 'of', and find splits according to contexts. They suggest a homophone split " ...could begin with a differentiation of pronunciations that was at first purely contextual, with reduced forms occurring in more frequent, more predictable, and possibly less prosodically prominent constructions or contexts. If the contextual differences became

[3] In Chinese, taboo avoidance has a long history. In the Tang Dynasty, 韩愈 (Han2 Yu4, 768-824) wrote a famous essay, titled as 讳辩 ('Hui4 Bian4', On Taboo Avoidance), criticizing the abuse of taboo avoidance, which reflected the profusion of taboo avoidance at that time. According to the law of Tang, people whose father's name was the same as that of the official positions couldn't go for that position. At that time, a talented young man called 李贺 (Li3 He4) was forbidden to take the imperial examination because his father's name 晋肃 was homophonous to the cadre's name 进士 ('jin4 shi4'). Han argued vividly that '父名晋肃，子不得举进士；若父名仁，子不得为人乎' ? (If the father's name is called 'jin4 su4', the son shouldn't go for 'jin4 shi4'; if the father's name is called 'ren2' ("benevolence"), shouldn't the son be a 'ren2' ("man") then?)

[4] The surprisingly large degree of homophone split in Chaozhou, i.e., 58%, is due to the abundant co-existence of literal and colloquial readings of the same morpheme. This phenomenon is called "wenbai yidu" 文白异读. Morphemes' multiple readings from wenbai yidu may not be taken as good arguments as homophone split. However, the cases of homophone split in other dialects without such wenbai yidu are still convincingly abundant.

lexicalized, this would lead to some of the distinctions in reduction becoming encoded in the lexicon, and would leave an association of more reduction with the more frequent lemmas" (p21).

2.1.3.2 Replacement by synonyms

Homophones may disappear when one of the homophonous words is replaced by a synonym. Bloomfield (1933) cites such an example which was reported by Gillièron and Roques (1912). In southwestern France, the words for "rooster" and "cat" were 'gallus' [gal] and 'gattus' [gat] respectively, both descendants of Latin-origin words. Due to a sound change from [-l] to [-t], [gal] "rooster" became *[gat], homophonous to "cat". Later other words more or less synonymous to "rooster" appeared to denote rooster, such as [azan] "peasant" or [begej] "farm-helper, handyman". In other areas where the sound change did not take place, the descendant of the Latin word for rooster survived. Why was "rooster" rather than "cat" affected by the homophony? In other words, why was not "cat" replaced by other words? Dauzat (1927) suggests an explanation as that the morphemes "rooster" occurred only in this single word, while "cat" was backed by a number of derivatives, such as the equivalents of standard French 'chatte' ("she-cat"), 'chaton' ("kitten"), and 'chatière' ("cat-hole").

Another possibility for homophones to disappear is that some words are replaced by synonymous words borrowed from other languages. A language in Papua New Guinea, Haruai, provides a nice example of synonym replacement (Comrie, 2000). In Haruai society, one is not allowed to say the name of one's cross-cousin or of one's in-laws. Meanwhile, most personal names, both traditionally and in contemporary Haruai society, are ordinary content words. Thus, if a person has a taboo relative whose name is identical to that of an ordinary lexical item, that person is also forbidden to use that ordinary lexical item, and then has to find a substitute. In some cases, the taboo is extended to homophones. As a result, it is observed that Haruai has apparently borrowed corresponding words from its neighboring languages such as Kobon and Tok Pisin. For example, when the indigenous lexical item *cöc* ("tobacco") is tabooed, its homophonous word "church" is tabooed as well, and a word borrowed from Tok Pisin *haus lotu* is used for "church".

2.1.4 The exaptive usage of homophony

While homophones may cause confusion due to its one-to-many association between form and meaning, this feature is sometimes made good use for the purpose of humor or artistic exposition. Such use of homophony can be considered as an exaptation in language evolution, as homophones do not emerge or exist for the purpose of the artistic and humorous function, but language users exploit their existence to achieve special effect in communication. The examples of such use of homophony are abundant across languages. In Chinese there are many idioms or proverbs which make good use of homophones. One of the famous two-expressions is '和尚打伞 - 无法无天' ('he2 shang4 da3 san3 - wu2 fa3 wu2 tian1'). The second part of the expression makes use of a pair of homophones '发' (hair)-'法' (law). Chairman Mao Zedong of China once used this expression in a meeting with the American journalist Edgar Snow. While Mao meant was that he was not constrained either by laws or by heaven, Snow translated it by mistake as "a lone monk walking in the world with a leaky umbrella", which totally missed Mao's point[5].

In modern societies, commercial advertisements are often fond of playing with words to attract attention, and making use of homophony is one usual way to achieve the goal. A set of 20 examples collected from the advertisements in Hong Kong are given in Appendix 1.1.

2.2 Cross-language comparison of the degree of homophony

Though it is believed the existence of homophony is universal in languages, it is not known yet whether different languages have the same degree of homophony. In the following, I will report some preliminary attempts in answering this question.

[5]This example is taken from Wang (1989b, p409). Edgar's interview was originally reported in Life Magazine, April 4, 1971, p48.

2.2.1 Methodology of comparison and data description

There are several difficulties in having a quantitative measure on the degree of homophony and making cross-language comparison. First, as the lexicon in a language is basically an open set and keeps evolving, whether a given word has a homophone or not heavily depends on the size of the lexicon used for search for homophones. A lexicon with more entries and with many ancient words included certainly will be more likely to include more homophones. Therefore, to ensure comparable lexicons, from which homophones are extracted, is an important prerequisite for cross-language comparison. We have two ways to tackle this problem of comparability, owing to the availability of two sets of data. The first is a set of Chinese dialects, for which the pronunciation of the same set of Chinese characters are available. Another set of data comes from the large database of words for three Germanic languages. The details of the two sets of data will be described later.

The second difficulty in deriving a homophone list for a language is a long-standing problem for either lexicologists and semanticists, that is how to distinguish homophony from polysemy. To avoid the difficulty of differentiating polysemes from true homophones, in the study of three Germanic languages, we restrict the scope of our analyses to only homophones which have different orthographic forms, in other words, we exclude homonyms such as '(river) bank' and '(financial) bank'. It is assumed that very often when two words are spelled differently, the chance for two homophones to have the same etymology is very small. We note that this will underestimate the degree of homophony with the exclusion of homonyms.

Following the above principle of polyseme pruning, those words, such as 'work' as a noun and 'work' as a verb which have the same meaning but are used as different part-of-speech, are not considered as homophones here[6]. We further exclude those pairs which are different inflection forms of one lemma, because these words in fact refer to the same meaning even though there are distinctions in either gender, number or tense. For example, in some dialects of Dutch, the singular and plural forms of the same lemma "poor man", e.g., 'arme' and 'armen', are pronounced the same with

[6]In fact such pairs of words belonging different grammatical classes have been regarded as polysemes rather than homonyms in the field of polysemy studies (Cuyckens and Zawada, 2001, xiv).

the loss of nasal ending which is a frequent change across languages. Such pairs are excluded in our homophone list. The homophone list derived based on the above criteria provides only a partial representation of the situation of homophony, and it is unclear yet how serious this underestimation is. However, this narrow definition of homophony gives us an estimation of the lower-bound of the degree of homophony in these languages. More importantly, we can now compare different languages based on a set of explicit criteria.

2.2.2 Degrees of homophony in Chinese dialects

The data for Chinese dialects is from the Dictionary on Computer (DOC), which is an electronic database of the phonological systems of Chinese languages[7]. It was first developed in the research group led by Prof. William S-Y Wang at Berkeley in 1966 and has been upgraded and maintained through the years (Cheng, 1996; Wang, 1969b). DOC has been a fertile database which many studies of sound change in historical Chinese and Chinese dialects have used, such as the tone changes in Shuangfeng, the phonological change of Middle Chinese initials, the distribution of initial consonants in dialects, and so on (some of these studies were collected in Wang, 1977). All these studies constitute a solid empirical basis for the launch of the theory of lexical diffusion (Chen and Wang, 1975; Wang, 1969a).

DOC was mainly constructed based on the data from 汉语方音字汇 (Hanyu Fangyin Zihui, Chinese Dialect Character Pronunciation List, henceforth Zihui), which is a complied volume providing the pronunciation of over 2,700 monosyllabic morphemes (or "Chinese characters", to be more accurate) in a number of modern Chinese dialects (Zihui, 1989). The latest edition includes data from 17 dialects, including Beijing, Jinan, Xi'an, Taiyuan, Hankou, Chengdu, Yangzhou, Suzhou, Wenzhou, Changsha, Shuangfeng, Nanchang, Meixian, Guangzhou, Xiamen, Chaozhou, Fuzhou. In Zihui, the Middle Chinese phonological category for each morpheme based on the rime book 广韵 (Guang3 Yun4)[8] is also provided. In addition to these data, the latest

[7]I am grateful to Prof. C.C. Cheng who kindly provides me the latest Windows version of DOC developed by him.

[8]Guangyun is an extant edition from the Song Dynasty of the rime book Qie Yun 切韵 (601 CE), which is assumed to represent the phonological system in Middle Chinese (MC).

version of DOC supplements the pronunciations for the same set of characters in the Shanghai dialect, the reconstructed pronunciation reflected in 中原音韵 (Zhongyuan Yinyun)[9], two different pronunciations of these characters in Japanese (i.e., Kan-on and Go-on readings), and the pronunciations in Korean.

Owing to the provision of DOC, we have access to sets of morphemes with similar size in the 20 Chinese dialects. One advantage of these data is that the semantic range is approximately the same for the dialects, since all dialects use the same set of characters which is assumed to reflect a similar semantic domain. There is a recognized problem for the definition of "wordness" in Chinese. It is hard to set a clear-cut criterion to determine whether a combination of some morphemes is a word or a phrase[10]. Therefore, to use an equal set of monosyllabic morphemes seems a valid and sound basis to carry out the comparison on the degree of homophony. We note that the obtained measure is only valid in monosyllabic morphemes, and it does not reflect an overall situation of current modern dialects, as these monosyllabic morphemes may not be used as free morphemes and there are many polysyllabic words in the contemporary dialects. However, these are the data we could have convenient access so far, and the obtained measure may provide at least some preliminary comparisons, which can be extended to a better coverage when data of the lexicons of modern dialects are available.

Table 2.1 gives the numbers of entries of morphemes in DOC[11], and the syllable inventory, i.e., the number of syllables occurring in these morphemes (Syl) (tone included). The table presents a set of variables in measuring the degrees of homophony in the 20 Chinese dialects and the two historical varieties of Chinese. There are three

[9]Zhongyuan Yinyun is a rime book in the 14th century, representing Early Modern Mandarin, or so-called Old Mandarin (OM), in the Yuan Dynasty.

[10]Such difficulty does not exist in English and other alphabetic languages, as their words are separated by space. In Chinese, however, as there is no space between words, it is hard to determine whether expressions such as '牛肉 niu2 rou4' ("beef") and '彩蝶 cai3 die2' ("colorful butterfly") are words or phrases. In fact there has been no consensus for what constitutes "word" in Chinese. Various criteria have been proposed, such as the criteria of replacement, insertion and others (Chen, 1999a).

[11]Some of the characters have more than one entry in the database, because one character can have several pronunciations. One source is from tone derivation, such as a noun 数 shu4 ("number") and a verb 数 shu3 ("to count"). Each is treated as a separate entry in the database. Another source of multiple pronunciations is Wenbai Yidu, which is common in many dialects. Some characters have two readings, one literal (Wendu) and one colloquial (Baidu), such as "listen" in Cantonese has [t^hɪŋ1] and [t^hɛŋ1] respectively. Therefore the actual numbers of entries in different dialects vary, as seen in the 2nd column in Table 2.1.

Table 2.1: The degrees of homophony in 20 modern dialects and two historical Chinese (sorted according to the number of occurring syllables *Syl*.

Dialect	entries	*Syl*	*HomoSet*	*PercSet*	*HomoPair*	*PercPair*	*AverHomo* per *Syl*
Taiyuan	3933	828	580	0.70	14581	0.0019	4.75
Wuhan	3947	870	625	0.72	13412	0.0017	4.54
Chengdu	3838	938	657	0.70	11769	0.0016	4.09
Yangzhou	3766	947	642	0.68	11673	0.0016	3.98
Hefei	3693	976	661	0.68	10782	0.0016	3.78
Changsha	4174	981	653	0.67	13548	0.0016	4.26
Suzhou	3967	999	644	0.64	12077	0.0015	3.97
Shuangfeng	4020	1001	672	0.67	10802	0.0013	4.02
Wenzhou	4108	1048	682	0.65	13587	0.0016	3.92
Ji'nan	3853	1063	732	0.69	9855	0.0013	3.62
Xi'an	3875	1084	745	0.69	9397	0.0013	3.57
Nanchang	3842	1111	732	0.66	8828	0.0012	3.46
Beijing	4111	1125	757	0.67	10564	0.0013	3.66
Jian'ou	4181	1241	780	0.63	10154	0.0012	3.37
Meixian	3848	1304	785	0.60	7539	0.0010	2.95
Yangjiang	3682	1319	800	0.61	6485	0.0010	2.79
Guangzhou	3773	1367	812	0.59	6143	0.0009	2.76
Fuzhou	4398	1413	867	0.61	8639	0.0009	3.11
Chaozhou	4193	1759	919	0.52	5977	0.0007	2.38
Xiamen	5000	1855	993	0.54	8664	0.0007	2.93

measures, (1) the number of homophone sets, and the percentage in the total number of sets of morphemes; (2) the number of homophone pairs, and the percentage in the total number of pairs; (3) the average number of homophones in one syllable. The three measures of the degrees of homophony are carried out as follows.

1) *HomoSet*: the number of syllables which have homophones (the 4th column in Table 2.1). For example, *HomoSet*(Beijing)=757 and *HomoSet*(Guangzhou)=812, i.e., Guangzhou has more homophone sets than Beijing. However, we need to do a normalization as the total number of syllables should be taken into account. Dividing the number of syllables with homophones over the total number of actual syllables, i.e., *HomoSet*/*Syl*, we obtain a more indicative measure, i.e., *PercSet* (the 5th column). Now *PercSet*(Beijing)=0.67, and *PercSet*(Guangzhou)=0.59. The 20 dialects have different degrees of homophony. There exists a significant negative correlation between the degree of homophony and the number of syllables, as shown in Figure 2.2. Pearson correlation test shows a high correlation: Corr(*PercSet*,*Syl*)=-0.90 ($p<0.001$).

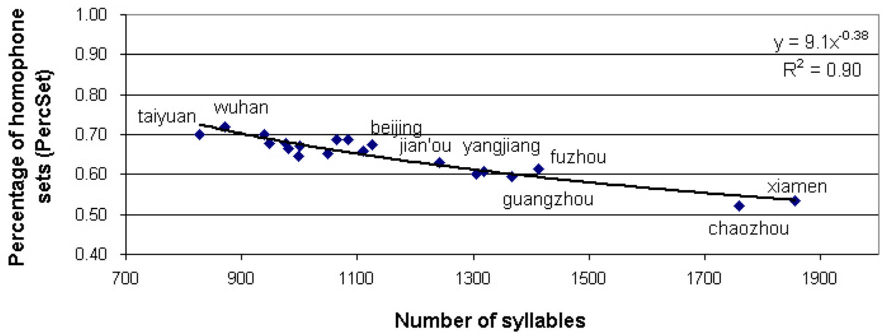


Figure 2.2: The correlation between the size of syllable inventory and the degree of homophony in terms of percentage of homophone sets. The power function and the R-squared value for the curve-fitting are given in the figure.

2) *HomoPair*: the number of pairs of homophones (the 6th column). *HomoPair*(Beijing) =10564 and *HomoPair*(Guangzhou)=6143. After normalization, by dividing *HomoPair* over the total number of pairs of morphemes, we obtain the percentage of homophone pairs as another measure of the degree of homophony, i.e., *PercPair* (the 7th column). We have *PercPair*(Beijing)=0.0013, and *PercPair*(Guangzhou)=0.0009. Similar to what reflects in *PercSet*, Beijing has a higher degree of homophony than Guangzhou. Using this measure of *PercPair*, we again see an even higher correlation between the degree of homophony and the size of syllable inventory: Corr(*PercPair*,*Syl*)=-0.96 ($p<0.001$), as shown in Figure 2.3.

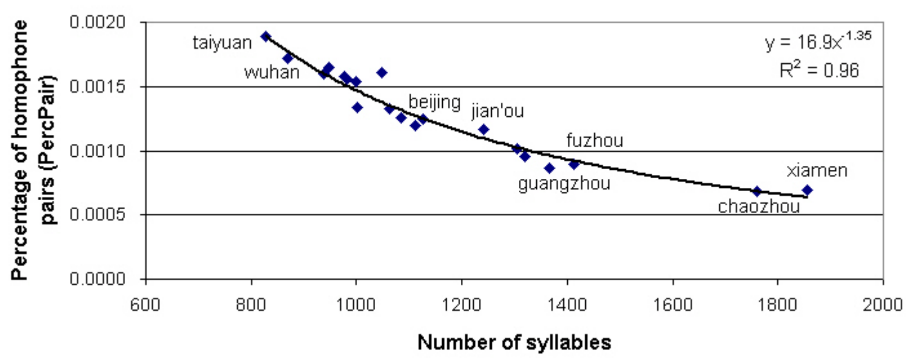


Figure 2.3: The correlation between the size of syllable inventory and the degree of homophony in terms of percentage of homophone pairs.

3) *AverHomo*: the average number of homophones per syllable (the 8th column)[12]. Again we find that the *AverHomo* has a high negative correlation with *Syl*: Corr(*AverHomo*, *Syl*)=-0.85 (p<0.001), as shown in Figure 2.4.

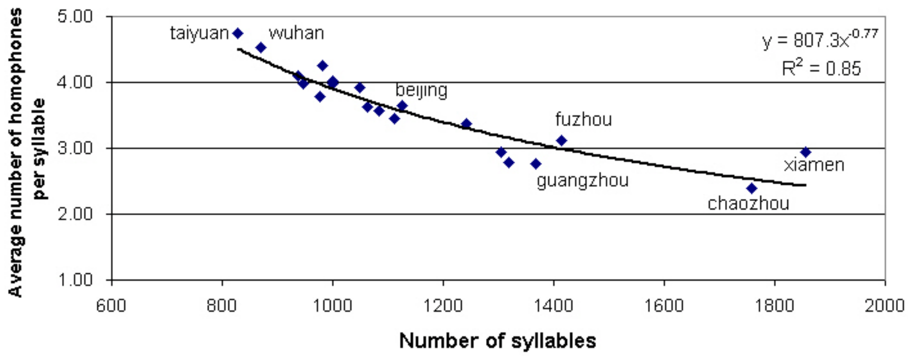


Figure 2.4: The correlation between the size of syllable inventory and the degree of homophony in terms of average number of homophones per syllable.

The three different measures discussed above all exhibit a high negative correlation between the syllable inventory and the degree of homophony. The convergent results from different measures suggest the robustness of the measure of the degree of homophony. The correlations suggest that the more syllables a language has, the small degree of homophony the language will accommodate. This correlation conforms to our intuition about the relationship between the size of the phonological resource and the degree of homophony.

2.2.2.1 Degrees of homophony in three Germanic languages

In this study, we consider another set of languages, i.e., three Germanic languages including Dutch, English and German[13]. The CELEX electronic lexical database[14] was

[12]Feng (1995) did a similar calculation and found the average homophonous morphemes per syllable in modern Standard Chinese is 5.4. His analysis is based on about 7000 characters complied by 国家语言文字工作委员会 (Guojia Yuyan Wenzi Gongzuo Weiyuanhui), which is much more than what DOC provides (only 2700 characters). Therefore, it is not surprising to see that his measure gives a much higher value than what we have here: for Beijing dialect, there are on average 3.66 morphemes per syllable.

[13]We are aware of the closeness of the three languages, as they belong to the same West Germanic language family, and have diverged later than 1000-500 BCE. The close relations and large similarities among these languages, as well as among the Chinese dialects, may make our findings lack generality. However, we believe that this work is a useful first step in examining homophony, an important feature of language, in a quantitative manner.

[14]I would like to thank Mr. Dinoj Surendran for sharing and helping in the use of the database.

Table 2.2: Information about the three lexicons from CELEX.

	lemma types		wordform types		corpus size
	pre-processing*	post-processing	pre-processing	post-processing	
Dutch	124,136	122,400	381,292	313,270	42.38m
English	52,447	41,535	160,595	77,031	17.9m
German	51,728	51,728	365,530	321,081	6.0m

*Note: It is found that there are some repetitive entries in the lexicons for the three languages. Therefore, we carried out some cleaning processing on the lexicons to remove the repeated items.

developed by the Dutch Centre for lexical Information (CELEX) of the Max Planck Institute for Psycholinguistics (Baayen et al., 1995). The latest version (edition 1995) contains lexical information for three languages, including spelling, pronunciation, morphological structure, syntactic information (part of speech and subcategorization) and corpus frequency.

Table 2.2 gives the sizes of the three lexicons before and after processing, in terms of number of word forms and lemmata, and the corpora information from which the frequency information is obtained[15]. From the table we see that the three lexicons are not comparable in either of the two types of lexicon. For lemma lexicon, Dutch has more than twice the lemma than the other two languages; for wordform lexicon, the word count in English is less than half than that of the other two. However, we only consider the first 5000 most frequent words and carry out the comparison along the frequency bands. It is assumed that the ranking for high-frequency words is reliable as the corpora used are both sufficiently large and from similar genre. We consider that this method of comparing only highly frequent words will help solve the problem posed by the incomparability in lexicon size.

For each of the three languages, we first sort the words in the order of their frequency. Then we check for each of the first 5000 words if it has (a) homophone(s) in the word list, according to our restricted criteria of homophony[16]. The homophone lists obtained in this way are given in Appendix 1.3.

[15]The corpora of Dutch is from INL by Instituut voor Nederlandse Lexicologie; English is from COBUILD by University of Birmingham; and German is from MANNHEIM by Institut für Deutsche Sprache.

[16]We confine the word list for the search of homophony within the range that all words have no less than 2 occurrences in CELEX. The sizes of the word lists for the three languages are not the same: Dutch, English and German have 120,512, 45,819, and 58,657 words, respectively. However, the criterion of no less than 2 occurrences serves as one way of guaranteeing comparability.

The first 5000 frequent words are divided into 14 frequency bands, 10 bands each including 100 words in the first 1000 list and 4 bands each with 1000 words in the remaining list. For each band, we calculate the number of words which have homophones and take the cumulative percentage as the index of the degree of homophony. Figure 2.5 shows the cumulative percentages of words which have homophones in given frequency bands in the first 5000 word lists.

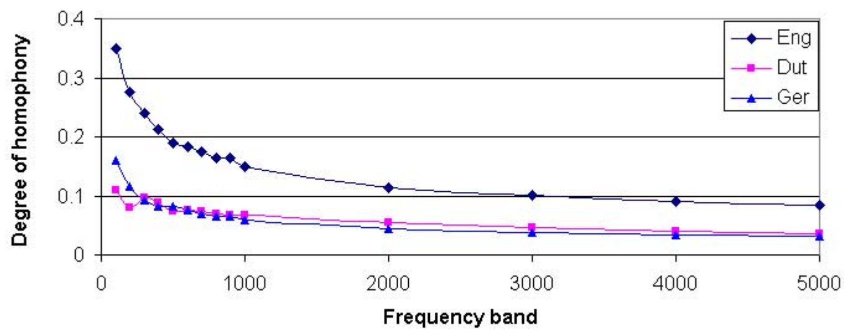


Figure 2.5: Degree of homophony in the first 5000 frequent words in three Germanic languages.

We have obtained several interesting observations. First, the degrees of homophony in the first several frequency bands are all much higher than later frequency bands in the three languages. In the first frequency band, English has 35% words with homophones, Dutch 11% and German 16%. In English, among the 35 words having homophones, 32 of them belonging to the closed class vocabulary, including function words, such as the articles 'the' and 'a', prepositions 'to' and 'in', and conjunctions 'but' and 'or', etc. In fact in the three languages, over 90% of the words in the first 100 most frequent words are such function words. It remains to be seen whether it is universally true for other languages that there exists a large degree of homophony in the most frequent words, and whether they mostly belong to the closed class vocabulary. Furthermore, we find that most of the homophones are monosyllabic words. We posit that there exists a correlation between the degree of homophony and monosyllabicity. We will examine this in more detail in a later section.

Second, while it is common that there are more homophones in high frequency bands than in low frequency bands, we also observe that the degree of homophony starts to level off at a certain value after the 12th frequency band which include 2000 words. This suggests that if we want to compare the degrees of homophony among different languages, we may only need to examine the high frequency word list to a certain extent. We see from Figure 2.5 that English has the highest degree of homophony (about 10%) while Dutch and German have similar smaller degrees (about 4%) as the level-off value. Why do the three languages have such differences? In Chinese dialects shown above we find a negative correlation between the number of syllables and the degree of homophony. Can we predict the degree of homophony based on some parameters of the language system? The following is a preliminary attempt to answer these questions.

2.3 Homophony and phonological resource

The capacity of handling a large number of words is considered as one defining characteristic for human species (Deacon, 1997). Most words in a language are represented by arbitrary associations between forms and meanings, which is characterized by the well-known Saussurean Sign (de Saussure, 1910/1983). While the number of meanings seems to be infinite, only a small number of them are lexicalized, and others are expressed by combining words according to grammar. The forms of the lexical items are built by choosing from a finite set, which we call "phonological resource". The size and characteristics of the components of this finite set of phonological resource should affect the degree of homophony. In the following, we begin with explaining how to measure the phonological resource in a language.

2.3.1 Measuring phonological resource

The "phonological resource" refers to the number of possible distinctive forms a language can make use of to construct words or represent morphemes. It depends not only on the number of sounds, i.e., the size of the phonological inventory, but also on the ways that the sounds are combined together, i.e., the phonotactic constraints. Languages differ a lot in both dimensions.

Though the number of sounds that humans can make is infinitely large as articulation exploits a continuous space in the vocal tract, the actual number of sound categories (or called “segments” or “phonemes”) which are used in distinguishing meanings by any individual language is very limited. The sounds can be mainly divided into two types, i.e., consonants and vowels, according to articulation (e.g., the obstruction of the air flow in the vocal tract and the vibration of the vocal folds). According to UCLA Phonological Segment Inventory Database (UPSID) (Maddieson and Precoda, 1990), the maximum number of segments in an extant language is 141 (Khoisan language !Xũ)[17]. The average number of segments among languages, however, is much smaller. In UPSID, the average size of segment inventory is only about 31.

Segments are organized into syllables, and word are constructed by concatenation of syllables. A syllable consists of an obligatory vowel, and optional preceding and following consonants. Different types of combination of consonants and vowels in one syllable constitute different canonical forms, such as [V], [CV], [CVC], [CVCC], [CCVC], and so on. Different languages vary a lot in the number and the complexity of legitimate canonical forms. For example, Germanic languages allow large consonant clusters, such as in English, [CCCVCCC] in ‘scripts’ and [CCVCCCC] as in ‘glimpsed’; [CCCVCCCC] as in ‘abstractst’ in Dutch and ‘strolchst’ in German; while the most complex canonical form in Chinese dialects is only [CGVN] (“G” standing for “glide”, and “N” for “nasal”), as ‘liang’ in Putonghua.

Table 2.3 lists the number of consonants and vowels and the number of canonical forms in the three Germanic languages. To ensure a valid comparison between the three languages, the criteria for determining the numbers of consonant and vowel phonemes are important. As Y. R. Chao’s classic paper discussed (Chao, 1934), the determination of phonemes in a language often has non-unique solution. Here we adopt the systems used in CELEX, as we assume that they have been made comparable in these three languages by the designers of the database.

[17]The 141 segments in !Xũ includes 24 vowels, 95 consonants and 22 diphthongs. UPSID does not include suprasegmental features, such as tones, in the phonological system. Nettle (1999b, p143) proposes another definition of segment inventory by including the combination of tones and vowels. Based on this definition, he identifies that the largest inventory is 195 in the Niger-Congo language Vute.

Table 2.3: Segment inventories, number of canonical forms, occurring syllable types, estimated and actual CV combinations, and CV exploitation rates in three Germanic languages.

	consonants	vowels	canonical forms	occurring syllables	estimated CVs	actual CVs	CV exploitation rate
Dutch	23	21	35	9,031	483	254	53%
English	24	24*	41	9,570	576	412	72%
German	25	34	33	4,225	850	217	26%

*Note: Four nasalized vowels, which occur only in foreign words, are included in the English vowel inventory.

From the inventory of consonants and vowels and the legitimate canonical forms, we see that the relations between the three variables are complex. English has fewer segments than German, but more types of canonical forms, which may be explained as the languages seem to have a trade-off between the number of segments and the ways of combining segments so as to achieve a similar size of phonological resource. However, this hypothesis does not hold for the case when Dutch and English are compared: Dutch has fewer segments than English, but also fewer canonical forms. But since we only have a small number of languages, it is hard to make more inferences.

The number of segments and the types of canonical forms may provide a measure of the potential phonological resource in a language. However, each language has a set of specific phonotactic constraints, resulting in many systematic gaps, such as no *[tl-] and *[dl-], and many accidental gaps such as no *[krɪp] and *[blɪk] in English. Therefore it is hardly possible to have an accurate estimate of the number of syllable types based on only the number of consonants and vowels, and the number of canonical forms. Jespersen (1933, p623) estimates the number of possible syllable types in English as more than 158,000, when systematic gaps are excluded. According to our calculation, however, the number of occurring syllable types in the CELEX English lexicon is only 9,570. If we assume that CELEX has included a representative number of syllable types as its lexicon size is sufficiently large (77,031 word forms), we may obtain a rough estimate of the exploitation rate of phonological resource in English, based on the above two estimates. Taking the ratio between our number (9,570) and that of Jespersen's (158,000), we estimate that the exploitation rate is only about

6%. This shows that the usable phonological resource is far from being sufficiently employed.

We may have another measure of the exploitation rate of phonological resource, by examining CV combinations only. The possible CV types can be estimated by taking the full combinations of all consonants and vowels. As shown in Table 2.3, the possible CV combinations are far from being fully utilized either. English has the highest rate (72%) and German has the lowest (26%). In fact, German has a larger segment inventory than Dutch and English, but German and Dutch have a similar number of CV combinations, while English has about twice as many as the other two. This implies that English has fewer phonotactic constraints than German and Dutch.

Furthermore, we find that the syllables are not utilized in a uniform way. Some syllables appear very frequently, such as in English [lɪ] (appearing 2850 times), [rɪ] (2016) and [ə] (1916), while a large proportion of the syllables (about 44%) only occur in one or two words. German and Dutch have similar characteristics. The three most frequent syllables in German are [gə] (4405), [tə] (3349) and [tən] (2845); and in Dutch they are [də] (17848), [tə] (12220) and [xə] (9899).

Figures 2.6, 2.7 and 2.8 show the distribution of the frequency of syllable types in the three languages. All the three curves can be interpolated as similar power-law functions ($Prob(f) = Cf^{\alpha}$, $\alpha_{eng} = -1.6$, $\alpha_{dut} = -1.3$ and $\alpha_{ger} = -1.6$), which appear as straight lines in the log-log plane. Power-law distribution is often considered as a reflection of the presence of self-organization in the system. The distribution characteristic of syllable frequencies implies that self-organization may be present in the organization of the lexicon.

2.3.2 Prediction of the degree of homophony

As shown above, the actual exploitation of the possible phonological resource is small. The number of occurring syllable types in the contemporary lexicon may serve as a representative index for the phonological resource in actual use. We propose the following hypothesis for the relation between the degree of homophony and the phonological resource in terms of the number of syllable types:

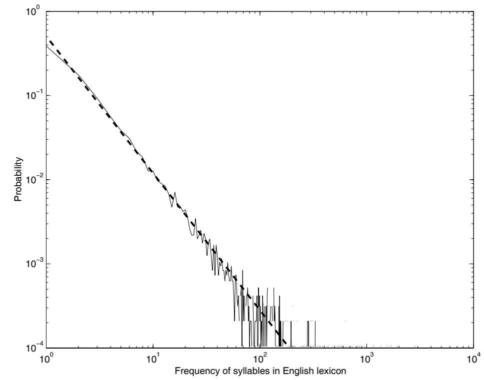


Figure 2.6: Distribution of the frequency of syllable types in the English lexicon. The solid line is the curve for the actual distribution, and the dotted line is the fitted curve with a power law.

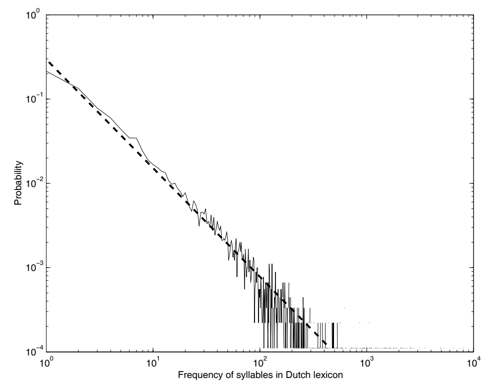


Figure 2.7: Distribution of the frequency of syllable types in the Dutch lexicon.

> Hypothesis-I: A larger number of syllable types (*Syl*) would predict a smaller degree of homophony.

The hypothesis seems straightforward. If there are more distinctive forms for constructing words, then the chance to have two words with the same forms, i.e., homophones, should be smaller. However, comparing the degree of homophony and the *Syl* in the three Germanic languages, such a correlation does not exist: English has the largest *Syl* (9570) and the largest degree of homophony (10% as in first 5000 frequency word list as shown in Figure 2.5), and German has a much smaller *Syl* (4225) than English, but also a smaller degree of homophony (4%).

Hypothesis-I seems invalid for the three Germanic languages. We recall that in our analyses of the degree of homophony in Chinese dialects, we do observe a

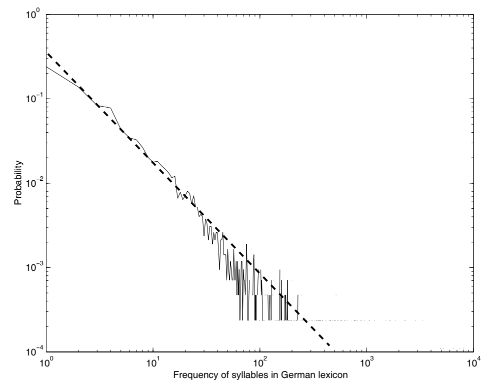


Figure 2.8: Distribution of the frequency of syllable types in the German lexicon.

significant correlation between the degree of homophony and the size of the syllable inventory, as reflected in Figures 2.2, 2.3 and 2.4. How to explain the inconsistency between the observations from these two sets of data? We note that in the case of the Chinese data, only monosyllabic morphemes are examined and many of them are not real "words" in the actual language use, while in the case of Germanic languages, the data are from real lexicons in which words have different lengths. It is easy to understand that longer words are less likely to have homophones. For two languages having the same size of syllable inventory, the one which has more long words will be expected to have fewer homophones.

It has been shown that languages differ a lot in the word mean length, and there is a high negative correlation between and the size of the segment inventory (Nettle, 1995, 1998, 1999b). Figure 2.9 shows the correlation for ten languages, taken from Nettle (1999b).

Therefore, we can confirm that the phonological resource cannot be simply determined by the size of the segment or syllable inventory; the length of words in a lexicon should also be taken into account. Due to the difficulty of obtaining a representative index of the size of phonological resource, it does not seem to be a good approach to predict the degree of homophony based on this parameter. We consider another way of predicting the degree of homophony, that is, to resort to the source of the emergence of homophony.

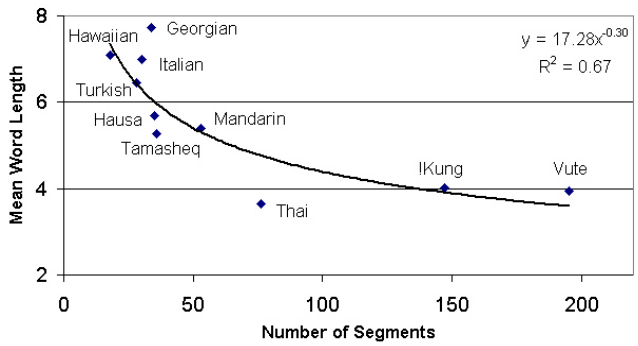


Figure 2.9: The relation between the size of segment inventory and the word mean length in ten languages. The curve-fitting power function and the R-squared value are shown in the figure. Adapted from Nettle (1999b), p146.

We know that homophones are derived mainly from sound change. When a sound merger happens, some minimal pairs affected by the sound change will collapse into a pair of homophones. Mergers in sound changes are very frequent. Therefore, the number of minimal pairs in the lexicon may be a better parameter to predict the degree of homophony. However, taking into account the full set of minimal pairs may confound the measure of potential homophony, because some pairs rarely cause confusion even though they are legitimate minimal pairs. For example, the minimal pair 'heed' and 'hard' are perceptually very distinct, as Higgins (1995) points out in his study on quantifying minimal pairs in English.

As we look into the homophone list in different languages, we find that most of the homophone pairs are monosyllabic words. This leads us to propose another hypothesis to replace Hypothesis-I in predicting the degree of homophony, as stated below:

> Hypothesis-II: A larger number of monosyllabic words ($MonoW$) would predict a higher degree of homophony.

We analyze $MonoW$ in the three languages in terms of their 5000 most frequent word lists. As shown in Figure 2.10, the percentages of monosyllabic words in the first 100 most frequent words are very high for all three languages, especially English and Dutch, both over 80%. But the percentages of $MonoW$ drop quickly and stabilize at different levels: English has a much higher percentage of monosyllabic words (32%)

than Dutch (20%) and German (14%). When we examine the correlations between the degrees of homophony and the percentages of $MonoW$ in different frequency bands, we find that the correlations are all very high in three languages, i.e., 0.99, 0.96, 0.98 respectively in English, Dutch and German. Tsou (1976) has similar observations that many examples of homophones are monosyllabic. His prediction that "in disyllabic or polysyllabic morphemes the probability for homophony is decreased geometrically" (ibid, p75) can be inferred from our data, from the observation that both monosyllabic words and the degree of homophony decrease simultaneously, especially in English.

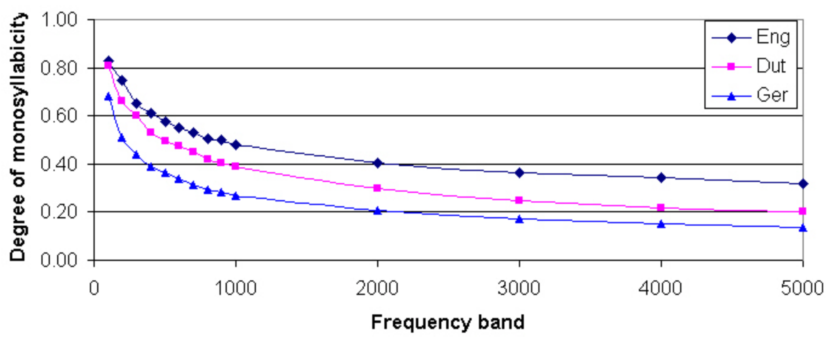


Figure 2.10: Degree of monosyllabicity in the first 5000 words in three languages.

It seems that each language has its own preference for monosyllabicity. It is not only related to Syl, which is a partial reflect of the size of phonological resources, but is also dependent on other aspects of the language system, such as the complexity of the morphological system. In the following section we will have some more discussions on monosyllabicity in Chinese, which concerns the controversy of disyllabification in the history of Chinese. We will revisit this issue in the light of the concept of self-organization, suggesting that the degree of monosyllabicity in the lexicon is an emergent property as the result of self-organization in the lexicon.

2.4 Self-organization in homophony

Do homophones cause ambiguity and confusion in daily communication? One answer to this question from common sense is that homophones do not usually affect communication, because the context, such as the neighboring words, could help to disambiguate. However, we can find many psycholinguistic experiments showing that there are differences in the processing of words with and without homophones. For instance, in Van Orden's experiment (Van Orden, 1987), subjects were asked to verify whether the referent of a word was a member of a specified semantic category. It is found that false positive rates were higher for words that were homophonous with a category member than for orthographically similar non-homophones. For example, "rows" was more likely than "robs" to be misclassified as a flower.

It has also been found that words with homophonous partners typically takes longer time than those without homophonous partners in lexical decision experiments (e.g., Ferrand and Grainger, 2003). The homophone interference effect found in these experimental situations may appear as insignificant or negligible. And if there is any confusion caused by the presence of homophony, the listener is the most affected. How will this effect on the listener affect the fate of the troubling homophones. There are two conceivable reasons. First, the confusion in the listener may affect the speaker, when the listener asks for clarification after confusion arises. Second, as we discussed in Section 1.3.6, a listening event may affect the individual's language and consequently his subsequent speaking behavior. The confusion caused by the homophony in the listener may remind the listener to avoid the use of the homophone in the same context in his own speaking. If such delay in comprehension happens often, the long term effect may turn out to be significant.

Therefore, some words which cause problems will face a de-selection pressure and consequently will be used less and less. This can be considered a self-organization process in the language system. We should be able to detect the effects caused by such homophony-avoidance process. The three phenomena described in the following sections are some evidence we will take into account.

2.4.1 Disyllabification in Chinese

The first evidence is the disyllabification phenomenon in Chinese history, which has been extensively discussed (e.g., Dai, 1990; Duanmu, 2000; Feng, 1995; Guo, 1938; Lü, 1963). It has been generally believed that at the early stage of Old Chinese, monosyllabic words are dominant in the lexicon, but the percentage of disyllabic words has increased a lot in Middle Chinese, and in various dialects of Modern Chinese such as Putonghua, disyllabic words are dominant in the lexicon. Many words which were monosyllabic in earlier times have become disyllabic. For example, 父 'fu4', which means "father", is now only used in disyllabic words, such as 父亲 'fu4qin1' in Putonghua; 睛 'jing1' ("eye" or "eyeball") has to be embedded in disyllabic compound words, such as 眼睛 'yan3jing1' or 眼珠 'yan3zhu1'. In modern Chinese dialects, monosyllabic words are only in a small proportion, and the majority of words are disyllabic. For instance, in Putonghua, monosyllabic words comprise only about 29% in the frequent word list[18]. Moreover, most new words introduced in the past century are disyllabic, and disyllabic words are increasing in modern Chinese (Duanmu, 2000; Masini, 1993).

There exist controversies on the disyllabification process, mainly in two aspects: When and Why and/or How. We will discuss about "Why" and/or "How" in the next section, where different hypothesized mechanisms to account for disyllabification are discussed. As for "When", some extreme view challenges the claim that ancient Chinese is a monosyllabic language. G. Kennedy (1951) presents such a view. He questions the dominance of monosyllabic words in ancient Chinese by calling it "the monosyllabic myth". He points out that the historical data kept in written language may not reflect the spoken language which was actually used at that time. Since the Chinese writing system is syllable based, the single characters may have concealed many original disyllabic or polysyllabic words. The characters which were originally a dependent part of a polysyllabic word usually was re-interpreted as an independent word later. For example, 凤凰 'Feng4 Huang2' ("phoenix") was originally the name of a bird but was interpreted by later dictionaries as two independent morphemes, i.e., 'Feng4' is the male bird and 'Huang2' is the female. Duanmu (2000, p145-172) also questions the claim of disyllabification, arguing that Chinese has always had many disyllabic words, and word length has been flexible throughout the history of Chinese.

[18]This measure is based on the word list 普通话三千常用词表 *Putonghua sanqian changyongci biao (3000 commonly used words in Standard Chinese)* complied by 中国文字改革委员会研究推广处 Zhongguo Wenzi Gaige Weiyuanhui Yanjiu Tuiguang Chu in 1959, as reported in Duanmu (2000).

The latter view has also been discussed in Guo (1938).

Kennedy's criticism is valid regarding the disparity between the written language and spoken language. However, such disparity was not only present in early periods, but also has persisted in the history. Even though the written language did not reflect the actual language in use, we may assume that the degree of disparity is relatively constant. The question whether ancient Chinese is a monosyllabic language is not that meaningful. What we want to know is whether there is an increase in the percentage of disyllabic words in the history as reflected by the written language.

2.4.1.1 Measure of degree of disyllabification

Scholars have carried out various statistical analyses on the classic Chinese material in different periods, such as Jiang (1991); Ma (1981); Wu (2001); Xiang (1993). However, the degrees of monosyllabicity measured by different analyses vary a lot. The reason is that the measure of the percentage of monosyllabic, or disyllabic, words is very dependent on the criteria of "wordness" the researchers adopt. The distinction between word and phrase has been one of the main problems in Chinese morphology, and there is no consensus on the criteria for the determination of a word[19]. Therefore this produces large variance among different calculations. Even for the same source book, the values could differ greatly. For example, for the book 论语 (Lun2 Yu3), Xiang (1993) estimates the percentage of polysyllabic words as about 15%, which is close to Wu (2001)'s estimation (22.2%), while Jiang (1991) only gives 2.4%.

Furthermore, it is hard to examine a large corpus by the traditional way of estimating the degree of disyllabification since it requires going through the texts character by character, to decide whether the character is a word by itself or a part of a disyllabic or polysyllabic word. Therefore, estimations have been mostly done with a small number of books or even some small sets of selected texts, and there is no common ground to make good comparison between different historical periods and different estimations by different researchers.

[19]For example, Cheng (1992) considers 父母 'fu4 mu3' ("parent") as in '为民父母，行政，不免于率兽食人，恶在其为民父母也？ in 孟子， and 富贵 'fu4 gui4' ("affluence") as in '富贵在天' in 论语 as phrases, but Wu (2001) judges them as word.

Table 2.4: A statistical estimate of the percentages of disyllabic words in several books by Jiang (1991), reproduced from Zhang (1997).

Book	Year	Words counted	Number of disyllabic words	Percentage of disyllabic words
Lun2 Yü3	372BCE-289BCE	15883	378	2.4%
Meng4 Zi3	551BCE-429BCE	35402	651	2%
Shi4 Shuo1 Xin1 Yu3	403-444	1998	190	9.5%
Lun4 Heng2	550-600	3582	270	7.5%
Bian4 Wen2 Ji2	700	2580	349	14%
Xi1 Xiang1 Ji4	1500	1473	257	17%
Hong2 Lou2 Meng4	1715-1763	2628	466	18%

Nevertheless, despite these difficulties in measure and comparison, we can still see the increase of the percentage of disyllabic words in Chinese history from some studies which compare materials in several periods with the same criteria. For example, Jiang (1991) did a statistical estimate on the disyllabic words in several books, as shown in Table 2.4. A clear increase of the percentage of disyllabic words can be seen from the table.

Sproat (2002) reports a study on using a statistical approach to measure the degree of disyllabification across three different historical periods, i.e., Pre-Han, Han and Jin/Song/Ming. His analysis is based on a large electronic corpus of historical Chinese developed by Academia Sinica in Taiwan, each period with 1.4 million characters. The percentage of disyllabic words is estimated by counting the proportion of real words among the 500 most highly associated character pairs generated from a given sample of texts. This approach decreases the degree of subjectivity of word judgement, because the researcher does not need to make a decision on every character when going through the examined text, but instead he only needs to judge the wordness of highly associated pairs of characters in the resulting list generated by the statistical algorithm. Therefore the measure is considered as statistically significant and highly comparable for different periods.

Table 2.5 shows Sproat's experimental results, from which we can see an obvious increase in disyllabification from Pre-Han to Jin-Song-Ming period. The percentage of disyllabified words seems too high compared to previous estimation, and when

Table 2.5: Estimated percentage of disyllabic words across three different historical periods, reproduced from Sproat (2002).

Period	Number of words	%Yield
Pre-Han(-206BCE)	318/500	64%
Han (206BCE-220CE)	401/500	80%
Jin-Song-Ming (265-1644)	430/500	86%

checking with the partial list provided, we see some words identified by the algorithm are not real disyllabic words[20]. However, the proposed statistical approach provides a promising technique in quantifying the measure of disyllabification for an objective comparison across different historical periods. With the algorithm and the data better refined, we may apply this method in analyzing the various periods in greater details. This may help us in zooming in on the time window when the disyllabification process started and progressed in the history of Chinese.

2.4.1.2 Disyllabification and homophony avoidance

As shown above, the increase of the degree of disyllabification since the Han Dynasty has been confirmed by various sources. The existence of a disyllabification process should be undisputable, and the controversies are more likely to be a matter of degree. The more controversial issue is the question "why did the disyllabification process arise?" or "How did the number of disyllabic words increase?"

There are several hypotheses in explaining the disyllabification process (see reviews in Duanmu, 2000, p150-158 and Wu, 2001), such as homophony avoidance (Wang, 1958), speech-tempo constraint (Guo, 1938), grammatical considerations (Li, 1990), morphologization (Dai, 1990), stress constraint (Duanmu, 2000; Lu and Duanmu, 1991), and prosodic constraint (Feng, 1995, 1998).

Controversies about homophony-avoidance hypothesis

[20] The list can be downloaded from http://www.research.att.com/rws). Some of the items in the given list , such as 子曰 'zi3 yue1' ("Confucius said") which is a phrase, should not be counted as disyllabic words. Dr. Sproat also confirmed through personal communication that the estimation may be confounded by the annotation texts which do not belong to the same period as the original texts, such as 师古 'shi1 gu3', which is actually the name of a commentator in later periods.

Homophony avoidance has been one of the most accepted hypothesized mechanisms for disyllabification (Wang (1958), Karlgren (1949), Lü (1963), and Li and Thompson (1981)). It has been argued that due to the dramatic shrinking of the phonological resources, a large number of homophones of monosyllabic words appeared, and possibly introduced many communication confusion. Disyllabic words are effective in avoiding such confusion, and therefore more and more disyllabic words appeared in the lexicon, mainly through compounding. However, this view has encountered several criticisms. In the following, we will summarize the arguments based on the review given in Duanmu (2000, p151-154), and present our views or answers to these questions, shown in parentheses following each point.

1. Most homophones can be disambiguated by context and rarely cause ambiguity in actual speech. When ambiguities do arise, a speaker can resort to a variety of ways to clarify them. It is unlikely that the entire speech community would come to agree on a single way of disambiguating each of the many homophones. (The observation that homophones do not cause ambiguity only sees part of the picture. There are many occasions in real life that homophones do cause mis-communications which are repaired subsequently, such as those shown in Appendix 1.2. Some disyllabic words may thus come into being through lexicalization.).

2. Many words which have homophones still remain monosyllabic, such as the three pronouns 'he', 'she' and 'it' are all pronounced 'ta1'. These words are high frequency words and the likelihood to cause ambiguities must be quite high. The existence of such homophones is hard to explain by the homophony-avoidance proposal. (The example of pronoun 'ta1' is not a good case for ambiguity, but rather a case of "generality"[21]. While ambiguity causes confusion, generality and vagueness do not. Instead, they may provide some advantage for language. For example, the unspecified gender may provide a convenience for the speaker to

[21]We need to distinguish between vagueness, generality, and ambiguity (Chao, 1959). Vagueness and generality refer to a semantic continuum. For example, the concept of "cold" is a type of vagueness whose specification of "being cold" within a continuum is not well-defined. On the contrary, ambiguity is always discrete. There is no overlap between the multiple interpretations of an ambiguous expression.

refer to something unspecified. In speech communication, the Chinese do not have the problem that English speakers have to use a phrase "he or she" to refer to some unspecified person without worrying about gender discrimination[22]. Moreover, the differences between English and Chinese in this morpheme reflects the different degrees of lexicalization in the same semantic domain of these two languages. English has three lexical items for the third singular pronoun, while Chinese has only one item[23] [24]. Moreover, monosyllabic homophones do not necessarily all become disyllabic; but those which often cause confusion will be more likely to get disyllabified.)

3. There is no clear evidence that classical spoken Chinese in pre-Han times mostly consisted of monosyllabic words. Many disyllabic words can be found in texts from the period of Old Chinese. Also, even though we assume a larger phonological inventory before Middle Chinese, there must have still been many homophones from the beginning in classical Chinese, as exemplified by the written forms of many morphemes which shared the same phonetic part in the Chinese characters. The reason we see less disyllabification in classical Chinese is, as pointed out by Guo (1938), that the classical written texts did not reflect the spoken language, due to the scarcity of writing materials, and written characters provide more distinctive information than speech, and therefore demand less on using disyllabic words for disambiguation. (We have acknowledged the disparity between written and spoken language in the last section, and argued for the valid observation of the increase of the percentage of disyllabic words since the Han period.)

4. Most of the disyllabic words appeared in the past 100 years or so (Lü, 1963).

[22]In written form Chinese has the option of distinguishing three third person pronouns, which is a recent development.

[23]According to Guo (1980) the third person pronouns in Chinese first emerged in the Zhou Dynasty from demonstrative pronouns (之 and 其). The pronoun 他 in Modern Chinese was first used as an indefinite pronoun in Pre-Han period and did not fully function as a third person pronoun until the Tang Dynasty.

[24]In other semantic domains, Chinese has a higher degree of lexicalization. For example, English has only "brother", while Chinese has 'ge1ge' and 'di4di'. Therefore, if Chinese 'ta1' is considered as ambiguous, then the English 'brother' should be considered as ambiguous as well, which will be easily seen as unreasonable.

However, during this period, the phonological system of Chinese has seen little change. This casts doubt on the homophony-avoidance proposal which links the disyllabification with the change of phonological system since Middle Chinese. (The appearance of large number of disyllabic words in the last century is not due to homophony avoidance, but this should not rule out the possibility at the earlier stage. We will discuss this later in more details.)

5. If there is a need to avoid homophony, why didn't it prevent the shrinking of phonological system? (Language change is beyond anybody's control. Especially, languages undergo perpetual sound changes. The shrinking of phonological system is not "preventable".)

6. There are many examples in compound constructions which show that the disyllabic forms are restricted in a way irrelevant to homophony avoidance, but more related to the prosody or stress system. For example, 煤炭商店 ('mei2tan4 shang1dian4'), 煤炭店 ('mei2tan4 dian4'), and 煤店 ('mei2 dian4') ("coal store") are all well-formed phrases, while * 煤商店 ('mei2 shang1dian4') is not. But the ill-formedness of the construction cannot be attributed to homophony avoidance (Duanmu, 2000, p153). Also some words which do not have homophones but still get disyllabified, such as 吼 'hou3'/ 吼叫 'hou3jiao4' ("roar"), and 'chong3'/'chong3ai4' ("pamper"). (It is right that the formation of disyllabic words in the contemporary language has nothing to do with homophony avoidance. The prosodic and stress constraints are more plausible explanations. The argument in item 4 above is applicable here. Moreover, this cannot deny the possibility that homophony avoidance can result in disyllabic words.)

7. There is no evidence to confirm the prediction of the correlation between the number of homophones and the number of disyllabic words. (Our study to be reported below provides some evidence for this correlation using the data from Chinese dialects.)

Our position on this issue is not to take homophony avoidance as the exclusive reason for disyllabification. We view disyllabification as the result of several mechanisms, and homophony avoidance is only one of the several reasons. There may have been several stages of disyllabification, due to different mechanisms at work. The question about the relationship between homophony avoidance and disyllabification should change from "is disyllabification due to homophony avoidance?" to a more specific one "does homophony avoidance lead to disyllabification?" What we are interested in is to show how some disyllabic words could have come about as a result of homophony avoidance. Such a process of disyllabification serve as evidence of how the language system self-organizes itself. We propose that if some homophones cause ambiguity, the ambiguity avoidance will lead to disyllabification of the homophonuous words. If homophones do not cause ambiguity, or the ambiguity can be resolved by other means, those homophones may remain monosyllabic. But these homophones may still get disyllabified, as there are other mechanisms leading to disyllabification, as mentioned earlier.

It is true that in most cases homophones can be disambiguated by the context in communication. However, there still exist situations where contexts do not provide enough information for immediate comprehension, and misunderstanding persists for a while until enough information is obtained. The postponement for further information would result in some delay of information processing in the listener. When it is realized that there is some mis-communication going on, it would require extra effort for the speaker to attempt new ways to clarify the situation. Though these may be some spurious processes, we believe that such occasional effects can accumulate and lead the language user to disfavor some expressions which have caused problems in previous communications, or remember the expression which helps to resolve the misunderstanding and re-use it again in a later similar situation.

As a consequence, many collocations become recurrent patterns and fixed constructions, and later, the independent words become lexicalized as one bound word. The disyllabic words are the crystallization of some frequently repeated collocations. For homophones, such collocations are more frequent, since the homophonous words need the context. Usually another neighboring morpheme semantically related to the target morpheme would suffice to disambiguate. Thus, many disyllabic words constituted two synonyms or near-synonyms are formed, such as such as 希望 'xi1wang4' ("wish"), 商量 'shang1liang2' ("discuss"), and 想念 'xiang3nian4' ("miss"), etc. where the two morphemes in the disyllabic word are semantically very close. One morpheme provides the semantic context for the other. In Pre-Han, the majority of the compound words are of this type (Wu, 2001).

In English, there are similar phenomena as the disyllabification in Chinese. An additional morpheme is used to disambiguate one ambiguous word, and the collocation of two words gradually becomes a fixed expression, and later may become a lexicalized word. For example, in some areas in the United States, there has been a sound change merging [ɛ] and [ɪ], which results in pairs of homophones such as 'pen' and 'pin'. It is found that these two words are expressed by adding a modifier, for example: 'ink pen', and 'stick pin' in order to eliminate the possible confusion[25]. Similarly, to differentiate the two meanings of 'funny', a pair of expressions 'funny haha' and 'funny strange' are created. Also, after 'you' was extended from the second person plural pronoun to the second person singular replacing the original 'thou', the expression 'you all' has been created to distinguish the plural from the singular. These examples show how ambiguity avoidance leads to fixed collocations of individual words. Though so far these words have not become lexical items yet, it is very likely they will become lexicalized later.

It is true that different people will have different ways to resolve the ambiguity caused by the homophones. It is impossible to imagine that the whole community will simultaneously adopt the same way of disambiguation by using the same set of disyllabic words. No individual has the global view of avoiding certain problematic homophones or has the intention or the power to lead the whole community to adopt the same way of avoiding homophony. The process of disyllabic words coming into the lexicon of the communal language does not differ in principle from the emergence of new words: at first, a novel combination of two monosyllabic words appears similar to random creation of a new word; the created expression diffuses through the

[25] I would like to thank Prof. Laurant Sagart who first suggested this example to me.

interactions in the speech community, by both horizontal and vertical transmission; when the expression is used frequently enough, it gets conventionalized as a word in the lexicon of the communal language. At the beginning, different individuals would come up with different ways of disambiguation for the same ambiguous word, and it is possible that several ways can co-exist in a communal language. This may be one of the explanations for the fact that a monosyllabic morpheme can have several disyllabic forms which are often synonyms, such as 'zao3chen2' (早晨), 'zao3shang4' (早上), 'qing1zao3' (清早) ("morning"). Also, we find that in different dialects, some synonymous disyllabic words are combinations of the same morphemes but with different orders, such as '素质' in Putonghua and '质素' in Cantonese ("quality")[26]. These variations may be the different outcomes from competition among the variants in different dialects.

We hypothesize that around the period of Qin-Han (around 200 BCE - 200 CE), the emergence of many such lexicalized disyllabic phrases started the disyllabification process, as a consequence of the shrinkage of the phonological system and the subsequent rise of the need to disambiguate the resultant homophones[27]. This is the first phase of the disyllabification process. As the proportion of disyllabic words in the lexicon increases, the prosodic or rhythmic structure is gradually formed through the generalization of subsequent generations' learning.

Feng (1995, 1998) proposes a prosodic hypothesis to account for the disyllabification process in Chinese history. He suggests that before disyllabification, monosyllables with complex canonical forms such as [CCVC] formed a valid foot. Later (starting from the Shang and Zhou Dynasty), when most consonant endings and consonant clusters were lost and subsequently tones emerged, the prosody structure requires two syllables to form a valid foot, and therefore disyllabic words become the

[26]Kosaka (1997) gives a number of such cases of the co-existence of AB and BA in the ancient texts, such as 要紧，热闹，整齐，言语，看觑，声音，点检，带携，闹吵，等待，买卖，拣选，想念，选择，名姓，喜欢，唬，战争，许诺，立站，找寻，惧怕，竞争，答对，把守，怕恐，歇宿，因为，该应，欲待. Reproduced from Li (2001).

[27]A tentative prediction based on this scenario is that in the early stage of disyllabification, those highly frequent words with homophones would have become disyllabified first. And those words were more likely to become compound words with semantically related morphemes combined together. Investigation into historical texts will help to test this prediction.

basic template for word formation and the number of disyllabic words increased accordingly. However, it is still unexplained how did the foot formation constraint come into being in the speakers at early times. At the current stage, it seems that individual speakers derive the prosodic constraint after they acquire a number of words and the phonological system. But at the beginning of the disyllabification process, how did the speakers achieve this constraint when there were still few disyllabic words and the sound changes of consonant loss were still in progress. Moreover, it is difficult for the constraint of foot structure to explain the existence of monosyllabic words amounting to at least one-third of the frequent words in modern Chinese dialects, and the large degree of variation of the degree of disyllabification (see Table 2.6 in the next section) while the syllabic structures do not differ much among the dialects.

Our hypothesis is that homophony avoidance is such a mechanism at the initial stage for the increase of disyllabic words. When the disyllabic prosodic structure is well established in the language, new lexical items are more likely to be disyllabic. This may account for the continuous increase of the number of disyllabic words in the last 100 years, especially for borrowing words from translations of many new foreign cultural items (Masini, 1993).

Correlation between the degree of homophony and the degree of disyllabification in Chinese dialects

The homophony avoidance hypothesis would predict the following correlations: a smaller phonological inventory implies a larger degree of homophony in monosyllabic morphemes, and consequently a larger degree of disyllabification. We test this hypothesis with the synchronic data, i.e., the 20 modern Chinese dialects. We have already seen from earlier analyses that there is a high negative correlation between the size of syllable inventory and the degree of homophony. As shown in Table 2.1, Guangzhou has a larger number of syllables (1367) than Beijing (1125) based on the calculation on DOC, while it has fewer monosyllabic homophones (59%) than Beijing (67%).

Lü (1963) has speculated the possible relation between the size of syllable inventory and the degree of disyllabification, comparing the northern dialects and dialects in the south: "because Cantonese has a larger syllable inventory than Putonghua, there should be fewer disyllabic words in Cantonese than in Putonghua" (p440). It is not difficult to find examples of words having been disyllabified in Putonghua which are still monosyllabic in Cantonese; for instance, 'wen2' ("mosquito") cannot be used as a free morpheme instead of "wen2zi" in Beijing, while in Cantonese the morpheme is still a monosyllabic word. However, there has been no systematic way to compare dialects quantitatively. In the following, I will report a preliminary attempt in testing the above hypothesized correlations for Chinese dialects.

We have designed a method to estimate the degree of disyllabification in different dialects. The dialect dictionary 汉语方言词汇 (Hanyu Fanyan Cihui) (1995) (henceforth Cihui) provides us a balanced set of data of the dialects for comparison. The Cihui gives a list of corresponding words for 1236 lexemes in the 20 Chinese dialects[28]. We first counted the number of monosyllabic words in the whole list[29], and calculated the percentage of monosyllabic words, denoted as $PercMono1$. The degree of disyllabification ($PerdDisy1$) is estimated as $1 - PercMono1$. Table 2.6 gives the degree of disyllabification of the 20 Chinese dialects, as well as the syllable inventory and the degree of homophony which are shown in Table 2.1.

We find there is a significantly high negative correlation between the size of syllable inventory and the degree of disyllabification: Corr(Syl, $PercDisy1$)=-0.74 ($p<0.001$). Also, there is a high positive correlation between the degree of homophony and the degree of disyllabification: Corr($PercSet$, $PercDisy1$)=0.76. Figure 2.11 and Figure 2.12 show the relation between the two pairs of variables and the curve-fitting functions. These two correlations provide strong support to the above hypothesis.

The above method to calculate the degree of disyllabification has some problems of overestimating the degrees of disyllabification, because in the list there are some words which were never monosyllabic in the first place, and are polysyllabic in all

[28] The selected meanings are mostly common concepts of daily life, such as "sun", "face", etc. We assume that the percentage of monosyllabic words for this set of meanings is representative for the degree of disyllabification of dialect.

[29] I would like to thank Wang Feng who helped prepare the data for the analysis.

Table 2.6: Comparison of degrees of homophony and degrees of disyllabification in 20 Chinese modern dialects

Dialect	Syl	PercSet	PercDisy1	PercDisy2
Taiyuan	828	0.70	0.60	0.40
Wuhan	870	0.72	0.62	0.40
Chengdu	938	0.70	0.62	0.42
Yangzhou	947	0.68	0.61	0.40
Hefei	976	0.68	0.61	0.40
Changsha	981	0.67	0.62	0.41
Suzhou	999	0.64	0.61	0.40
Shuangfeng	1001	0.67	0.63	0.43
Wenzhou	1048	0.65	0.53	0.31
Ji'nan	1063	0.69	0.59	0.36
Xi'an	1084	0.69	0.61	0.41
Nanchang	1111	0.66	0.60	0.38
Beijing	1125	0.67	0.62	0.41
Jian'ou	1241	0.63	0.55	0.31
Meixian	1304	0.60	0.60	0.39
Yangjiang	1319	0.61	0.51	0.24
Guangzhou	1367	0.59	0.50	0.24
Fuzhou	1413	0.61	0.51	0.25
Chaozhou	1759	0.52	0.50	0.23
Xiamen	1855	0.54	0.54	0.29

modern dialects, such as 玻璃 'bo1li' ("the glass"), or 蝴蝶 'hu2die2' ("the butterfly"). These words did not go through a "disyllabification" process from monosyllabic to disyllabic, and they should not be included in the estimation. We tried another way of calculation. We only take into account those meanings which are expressed by a monosyllabic word in at least one dialect, assuming that the original form for the meaning is very likely to be monosyllabic at earlier stages and has been retained in at least one dialect. This is based on the assumption that it is rare that a disyllabic word would become a monosyllabic word again. Based on this method, we obtain a better measure of the degree of disyllabification, $PercDisy2$, as shown in Table 2.6. We again calculated the correlations of interest. It is found that the correlations are even higher compared to those using $PercDisy1$: Corr(Syl, $PercDisy2$)=-0.76 ($p<0.001$); and Corr($PercSet$, $PercDisy2$)=0.78.

These high correlations provide a strong argument for the homophony avoidance hypothesis, because the existence of such a correlation is hard to explain by other proposals, such as prosodic or stress, for disyllabification. There has been no arguments

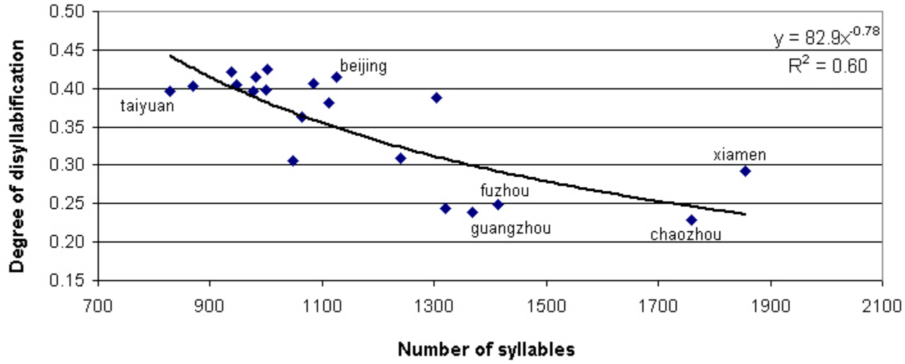


Figure 2.11: The sizes of syllable inventory versus the degrees of disyllabification in 20 Chinese dialects. The curve-fitting function is shown in the figure.

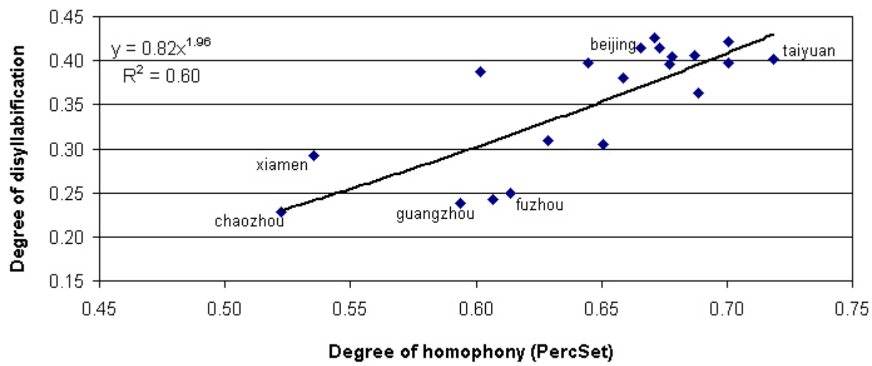


Figure 2.12: The degrees of homophony versus the degrees of disyllabification in 20 Chinese dialects. The curve-fitting function is shown in the figure.

to show that the prosodic or stress constraint is related to the size of the syllable inventory.

2.4.2 Grammatical differentiation between homophones

A pair of homophones sharing the same grammatical class are more likely to cause confusion than words belonging to different grammatical classes. Kelly and Ragade (2000) did some statistical analysis on the English homophones to test this hypothesis. They found that the existence of homophone pairs having the same grammatical class is not statistically biased as discussed below. However, many words belong to more than one grammatical class. When taking into account the frequency effect, it is found that the frequently used grammatical class of a word is statistically biased not to be the same as that of the word's homophone. For example, for a pair of homophone

Table 2.7: A summary of the experiments in Kelly and Ragade (2000). Homophone pairs are differentiated in usage in terms of grammatical classes.

conditions	pairs of homophones/ total pairs of words	mean from 50 cycles of random pairing	standard deviation	t value
existence constraint test	139/502	140.1	9.47	0.84 ($p>0.30$)
frequency constraint test	73/253	84	5.94	13.09 ($p<0.0001$)

"weight" and "wait", both can be a noun and a verb, but considering only the most frequent usage, one is a noun, and the other is a verb.

Kelly and Ragade carried out two statistical tests. The first is called "existence constraint test". A list of 502 homophone pairs is prepared[30]. It is found that there are 139 pairs of words which are from the same grammatical class. In order to test if the distribution of homophone pairs is some random effect, Monte Carlo experiments were run for fifty cycles of random word pairings. In these cycles, each word in the analysis was randomly paired with another word. It turns out that the mean number of pairs with the same grammatical class from these 50 cycles is 140.1. There is no statistical difference between the estimated value (140.1) and the observed value (139) as shown by t-test. The results are summarized in Table 2.7.

The second test is called "frequency constraint test". Another set of data with 253 homophones pairs was prepared, each word only marked by its most frequently used grammatical class. It is found that there are 73 pairs with the same grammatical class. The Monte Carlo experiments are performed again to this set of data, and the estimated mean value for random pairs of words with the same grammatical class is 84. The t-test shows that this value from random pairs is significantly different from the observed value, which suggests that homophone pairs tend to differentiate in different grammatical classes. As Kelly and Ragade explain, "this restriction seems reasonable if one assumes that the usage frequency of a word will be depressed if it has a greater chance of impairing comprehension". This can be taken as another evidence

[30] In the list, some words may appear in more than one pair, if they have different usages in different grammatical classes.

Table 2.8: Statistical results of the differentiation of homophone pairs in usage in terms of grammatical class in English.

English	pairs of homophones/ total pairs of words	mean from 50 cycles of random pairing	standard deviation	t value
existence constraint test	410/1125	426	13.5	8.6 (p<0.001)
frequency constraint test	207/448	203	10.6	-2.7 (p<0.005)

Table 2.9: Statistical results of the differentiation of homophone pairs in usage in terms of grammatical class in Dutch.

Dutch	pairs of homophones/ total pairs of words	mean from 50 cycles of random pairing	standard deviation	t value
existence constraint test	180/499	185	11.4	3.6 (p<0.001)
frequency constraint test	55/175	70.6	6.7	16.4 (p<0.0001)

of the self-organization in the lexicon.

We apply the same tests to the three Germanic homophone lists. The statistical results of the Monte Carlo experiments are shown in Tables 2.8, 2.9 for English, Dutch respectively[31]. Unexpectedly, the result of English is different from what Kelly and Ragade found. The existence constraint test shows that whole set of homophone pairs differentiate in grammatical class in a way significantly different from random data; but the frequent constraint test disproves the differentiation, as random pairs of words have more differentiation in grammatical class than homophone pairs. The homophone data we compiled are different from what Kelly and Ragade use, and it is not clear yet what accounts for this discrepancy. In Dutch, both tests show grammatical differentiation with statistical significance, as shown in Table 2.9.

[31] The German data has some problem and therefore the experiment is not applied to German. In the German database, different grammatical usages of one wordform are assigned an equal frequency.

2.5 Self-organization in the lexicon: Monosyllabicity and lexicalization

Now we come back to the controversy of the degree of monosyllabicity in Pre-Han times. As we mentioned, there are two extreme views on the degree of monosyllabicity in Pre-Han or later ancient Chinese. We may obtain another perspective on this question, by linking the degree of monosyllabicity with the phonological resources of the language and making cross-language comparisons. There have been various hypotheses for the reconstruction of OC (e.g., Baxter (1992); Sagart (1999). Based on the reconstruction of Sagart, we have a rough estimate that OC had about 3000 syllables. The size of the phonological resource in terms of monosyllables for the lexicon in OC is smaller than one third of modern English. As phonological resource of OC is much smaller than that of English, we would expect that the percentage of monosyllabic words in the lexicon would not be much more than that of English, if we assume that the degrees of lexicalization due to communication constraints do not differ too much among languages. As the monosyllabic words only take up 30% of the lexicon in English, it hard to believe that OC could have several times more monosyllabic words in the lexicon than English. Thus from the cross-language comparison we may have a rough idea how to evaluate the different estimations of monosyllabicity.

The discussion may continue fruitfully on the validity and plausibility of various criteria for "wordness" and the distinction between lexical words and phrases. However, to address the "wordness" problem from a dynamic perspective may be more revealing and would provide us some new insights for the self-organization in the language system. The lexicon is always changing. While new words incessantly appear, old words fade away. During this dynamic process, there is a tug of war among several conflicting factors to constrain the degree of monosyllabicity of languages.

On the one hand, monosyllabic words may serve the purpose of efficient communication - the shorter the better, at least from the speaker's point of view. Therefore many sound changes lead to phonological attrition; for example, a vowel in an unstressed syllable is often lost and consequently a syllable disappears. Such case can be easily attested in English, such as abbreviations, e.g., “coz” from “because”.

On the other hand, there is a natural tendency for language to lexicalize frequent collocations by concatenating several monosyllabic words, resulting in new words which are usually disyllabic or polysyllabic. As discussed earlier, some disyllabic words in Chinese may come from lexicalization, i.e., two frequently co-occurring monosyllabic words at early times may become a disyllabic word at a later stage. The lexicalization can be reflected by the writing system, such as in English. At the beginning, the words are separated by a space, but later when the phrase becomes more and more frequent, the space between the two independent words is replaced with a ’-’ and may disappear altogether later. For example, we have ‘handout’ now, which was from a phrase ‘hand out’. We can see many such cases, as some are still in the variation stage, such as ‘hand-picked’ and ‘handpicked’ found in CELEX. There is often some phonological reduction involved in this process of lexicalization, such as the widely attested forms in America, i.e., “wanna”, “gonna”, “gotta” from previous “want to”, “going to”, and “got to”. Here we have a nice example of the word ‘goodbye’ to illustrate such a lexicalization process in English. The word “goodbye” actually has gone through several stages as attested in the literature. Below are some excerpts from the OED.

1588 “I thanke your worship, God be wy you.” (SHAKESPEARE. L.L.L. III. i. 151)

ca1659 “But mum for that, his strength will scarce supply His Back to the Balcona, so God b’ wy” (CLEVELAND Lond. Lady 54)

1719 “Good B’ w’ ’y! with all my Heart” (D’URFEY Pills III. 135)

1818 “And so your humble servant, and good-b’ye!” (BYRON Juan I. ccxxi,)

1860 “We then bade Ulrich good-bye, and went forward.” (TYNDALL Glac. I. xviii. 122)

There is a chain of lexicalization in terms of changes in word length in the lexicon. Each change may start as random and sporadic innovation. However, language use poses a selection force in regulating the words, in terms of length and frequency of the words, etc. We would like to emphasize that the selection is implemented through the interaction of language users. The individual language user does not have the global view of how the system would turn out, nor the intention of constructing a system with a target structure.

We believe that the percentage of monosyllabic words in a language must be an emergent property, and different languages may share similarities in certain aspects. We have seen from the above analyses on the three Germanic languages that these languages exhibit similar distributions of monosyllabicity in the high frequency bands as shown in Figure 2.10 earlier.

To make the situation of monosyllabicity clearer, we have done some detailed analyses on the distribution of words with different word lengths in the three Germanic languages and modern Chinese. Since we have noticed that the lexicon structure may be dependent on frequency, we measure the distributions in the lexicon along the frequency bands. Figure 2.13 shows the distributions of words with different lengths from the first 100 words to the first 20,000 words in the lexicon. In the first 100 words, English and Dutch have 83% and 81% monosyllabic words respectively, while German has only 68%. The percentages of monosyllabic words decrease in a similar pattern: they drop quickly for the first several thousand words and then level off to a relatively stable value. Different languages level off at different degrees: English at about 21%, Dutch at 10% and German at 7%. Furthermore, the words in English are mostly of short length; monosyllabic, disyllabic and trisyllabic words make up about 85% of the words in the 20,000 most frequent word list, while in Dutch and German these three types of words take up only 76% and 73% of the lexicon. The degree of monosyllabicity seems to be language specific.

We analyze a Chinese word list in a similar way. We have access to the ranked word list in frequency order, which is compiled from the CKIP Corpus by Academia Sinica in Taiwan. Here we only show in Figure 2.14 the data up to the first 5000 words, assuming that the proportions remain relatively stable thereafter. We can see that the percentage of disyllabic words in Chinese is dominant, about 70% in the first 5000 words. But we can still observe a high proportion of monosyllabic words in the first several hundred words, which is very similar to the Germanic languages.

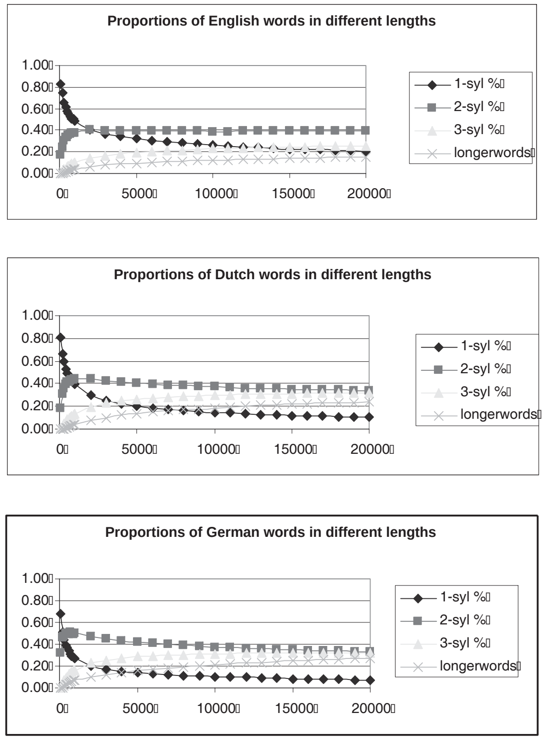


Figure 2.13: Proportions of words in English, Dutch, and German in different lengths among the first 20,000 frequent words.

From the above discussions of monosyllabicity and lexicalization, we can see the self-organization characteristics of the evolution of the lexicon. The distribution of monosyllabic words in different frequency bands reflects the frequency effect on the word length. The increase of disyllabic words in Chinese history reflects the effect of phonological attrition on the construction of the lexicon, and the effect of homophony. As the phonological resources decrease, sound changes produce mergers which result in a large number of homophones. However, we assume that the neighboring words provide the information for disambiguation and thus assures communication efficiency. In Chapter 5 I will report a simulation model to illustrate this idea.

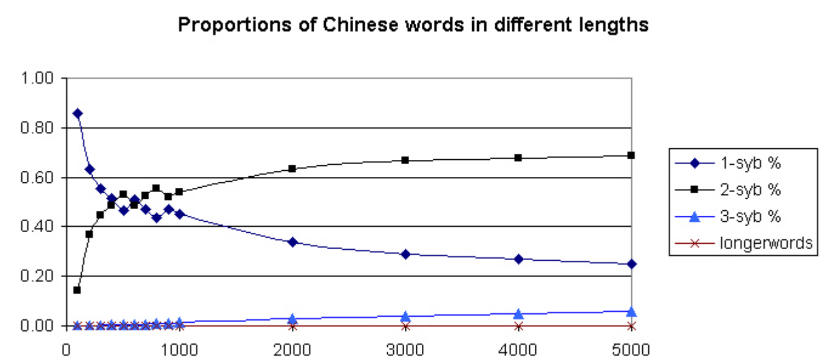


Figure 2.14: Proportions of Chinese words in different lengths among the first 5000 frequent words.

Chapter 3

Self-organization in the population and individuals: A case study on sound change

3.1 Why individual language users?

In the last chapter, we views language itself as an autonomous system which self-organizes in the process of evolution. This approach is similar to the structural or functional account for language change, in which the explanation for change resides in *the language system*. Constraints on speakers and/or listeners are proposed. However, these constraints are taken into account separately, and how these constraints work as selection forces is seldom elaborated. This can be exemplified by Martinet (1952)'s account for vowel changes described in Chapter 1, where it is supposed that a vowel "exerting a pressure downward" due to the tendency to achieve higher degree of integration of the vowel system. However, this type of analysis applying to language change does not explain how the change is implemented. It is not elaborated how the process of change would have been carried out in the speakers. Does every speaker go through the same process, by recognizing this asymmetry in his phonological system and purposely change it to approach a more integrated vowel system?

While it is obvious that no change occurs in a way that all speakers in a speech community simultaneously change all at once, we have to consider how language change progresses through diffusion through speakers across generations. Only when we embody our investigation in the individual language users situated in a speech community and interacting iteratively, can we have the chance to approach a full picture.

Saussure has a brief account of the importance of the individual language users: "the community of language users . . . is the *one and only reality*" (de Saussure, 1910/1983, Section 3.5, emphasis added), "everything which is diachronic in languages is only so through speech. . . . Speech contains the seeds of every change, each one being pioneered in the first instance by a certain number of individuals before entering into general usage. . . . (the innovative) form, constantly repeated and accepted by the community, became part of the language. But not all innovations in speech meet with the same success." (ibid, Section 3.9). How the seeds of innovation sprout from the individuals, and how the innovations get accepted as part of the language, these are the questions pertaining to the "implementation" problem of language change, which was clearly raised in Weinreich et al. (1968) and Chen and Wang (1975).

In most of the previous accounts for language change, the implementation problem is often overlooked. As McMahon (1994) remarks, while the structural linguists assume "phoneme change", the subsequent generativists speak of "rule change" (p35). The structuralists examine the evolution of a structured "language system", but do not specify whether the system is the individual or the community system. The generativists are clear about their object of analysis, which is the individual language users, but they continue to overlook the effect of social factors on language use and the actual heterogeneous situation in language use. Such a standpoint is reflected in the following often-quoted passage from Chomsky:

> Linguistic theory is concerned with an ideal speaker-listener, in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual

performance (Chomsky, 1965, p3-4).

However, as Mathesius (1911) points out very early on, the homogeneity of language is not an "actual quality of the examined phenomena," but "a consequence of the employed method" (p2). Since the 1960s, a surge of development in sociolinguistics (Labov, 1966, 1972, 1994, 2001) provides a new perspective in systematically studying the heterogeneity in the language community. Speakers differ greatly in their language behavior, stratified by various social factors, such as age, gender, education level, social class, and so on. Weireich et al. note that "it is not enough to point out the existence or importance of variability: it is necessary to deal with the facts of variability with enough precision to allow us to incorporate them into our analyses of linguistic structure" (1968, p169).

However, sociolinguists may have gone to an extreme, as the importance of studying individual speakers is underrated. Labov (1969) states that:

> " . . . the construction of complete grammars for 'idiolects', even one's own is a fruitless and unrewarding task; we now know enough about language in its social context to realize that the grammar of the speech community is more regular and systematic than the behavior of any one individual. Unless the individual speech pattern is studied within the over-all system of the community, it will appear as a mosaic of unaccountable and sporadic variation" (p759).

It is not straightforward to us that the individuals' speech patterns will be of no interest by themselves. As we have discussed in Chapter 1, each idiolect is the result of self-organization within the individual speaker. The self-organizing process and the characteristics of resultant idiolect, including the variation patterns which sociolinguists are interested in, should provide us an informative window to understand how the variations occur and develop.

Also sociolinguistic studies may be confusing the distinction between individual language and community language. When the statistical characteristics are found from the analyses of the language data in the community, some linguists tend to impose such characteristics on the individual language users. However, we caution not to extrapolate the regularities observed at the community level to the individual speakers, following the same spirit in our discussions of the honeybee comb. We cannot assume individual honey bees each have in mind a blueprint for constructing a hexagonal cell from observing its presence in the bee comb. While the linguists are able to identify regular patterns in an on-going sound change at the community level, it is not necessarily true that the regularity also resides in the individual speakers. Instead, the change is implemented through diffusion among speakers and across generations. Individual speakers are the loci of change. We need to know what the characteristics of individual speakers are to allow the regularity of change to occur.

Furthermore, as Shen (1997) points out, while sociolinguistic studies have mainly focused on demonstrating the linguistic variability and the social factors correlated with the variations, there is not enough effort devoted to understanding how individual differences are related to linguistic change as a whole (p11). Sociolinguistic studies mostly focus on the individual differences observed from linguistic behaviors; rarely have studies gone further to ask where such individual differences come from. Are they purely due to the idiosyncratic linguistic environment, or can they be attributed to the differences in language learning styles or even some differences in general cognitive development? Would the existence of such individual differences affect language change?

In order to answer the above questions, we advocate the self-organization perspective in studying language change: it is important to take into account *both* the properties of individual speakers, or "agents", *and* the interactions between them. These two aspects constitute the two central components of a self-organization system. Specifically, we highlight the dramatic heterogeneity in the population during the self-organization process in on-going changes, and explain the heterogeneity by delving into the individuals' production and perception. While early discussions on the properties of the agents mainly focus on the properties they share, we emphasize in addition their differences, not only in their specific linguistic behaviors, but more importantly in more general learning styles.

In the following, we will introduce an interesting sound change, which is an on-going change between /n-/ and /l-/ in a Chinese dialect, Cantonese. We have carried out some fieldwork and collected some synchronic data. We will report the detailed analyses on three types of variation exhibited in and across speakers, and the differences between production and perception. Two different learning styles of lexical learning are proposed based on the observed different linguistic behaviors. In the end of the chapter, we will give some preliminary explanations for the historical profile of the change.

3.2 The empirical case study of Cantonese /n-/→/l-/

3.2.1 Historical background of the change

In Cantonese, a group of words which are believed to have had /n-/ initials at earlier times[1] are changing to be pronounced with /l-/ initials in recent decades. There has been no systematic study to trace the origin of this change. But various pieces of evidence have suggested that this change may have started before early 20th century. One piece of interesting material related to this change is a place name in Hong Kong where the local vernacular is a Cantonese variety. There is a small island whose English name is 'Lamma Island'. Its Chinese name is '南丫' which is pronounced as 'nam nga' according to standard Cantonese. The first Chinese character, '南', with an /n-/ initial, corresponds to the first syllable 'lam' of the English name. The mismatch between English's /l-/ and the Cantonese /n-/ suggests the presence of /n-/→/l-/[2].

[1] The conventional term in historical Chinese for this group of words are called ni-initial words ('泥' 母字), as exemplified in the rime book 切韵 (Qie1 Yun4). It is assumed that MC is ancestral to Cantonese.

[2] The English name can be found in coastal maps far back in 1786 (Empson, 1992, Map 1-8), and it has been consistently used in later maps. However, the Chinese names had variants. In a map drawn in 1810 (ibid, Map1-13) the Chinese name for the island is shown as '蓝麻' which is pronounced as 'lam ma'. The first time for the two names, '南丫' and 'Lamma', being together was in a private venture map made around 1925-1930s (ibid, Map 2-11). It is puzzling, however, that the spellings of the Chinese characters were also provided in that map, which is 'namma'. It is a mystery why '南' was used if it was really pronounced as 'nam' at the time of the map. Why did the map-maker not notice the discrepancy between the Chinese and English names? One possible explanation is that the pronunciation with an /l-/ initial for '南' was already available, and probably widespread then. So even though the map-maker noted down a standard spelling for this character with an /n-/, he did not realize the mismatch of pronunciation between the characters and the English name. If this is true, the change of /n-/ → /l-/ should have started earlier than 1920s.

There are some other examples from the names of streets in Hong Kong (which are supposed to have existed for a long time since the British colonization) and some translations for foreign words, as shown in the upper part of Table 3.1[3]. In these names, the /l-/ foreign syllables are represented by Chinese characters which are supposed to have /n-/ originally. The mismatches suggest that local native speakers, at least the people who gave the translations, have already pronounced these characters with /l-/. The change of /n-/→/l-/ not only has affected the pronunciation of Chinese characters, but has also affected the perception of the /n/ sound in foreign words. There are some examples of mismatches in the opposite direction, i.e., /l-/ characters are used to represent the original /n-/ syllables of foreign words, as shown in the lower part of Table 3.1.

Table 3.1: Some evidence of the change /n-/→/l-/ in loan words or translations in Hong Kong

English	Written Cantonese	Cantonese pronunciation of target character
Cleveland St.	加宁街	ning
Lemon	柠檬	ning
Albany Rd.	雅宾利道	lei
Bisney Rd.	碧荔道	lei
Knight St.	励德街	lai
Macdonnell Rd.	麦当劳道	lou
Nassau Rd.	兰秀道	lan
Nation Rd.	礼信街	lai
Norfork Rd.	罗福道	lo
Northcote Rd.	罗富国径	lo
Bossini	堡狮龙	lung
Guinness	健力士	lik
Tiny computer	泰来电脑	loi
Sydney	雪梨	lei
notes	碌士	lok
number	林巴	lam
Heineken	喜力	lik

As early as 1940s, Y. R. Chao had observed in Guangzhou Cantonese that "about one out of four persons in Canton[4] has no initial *n*, and pronounces an *l* in words

[3] Most of the examples are taken from Lee and Murashima (2002). The case of "number" also appeared in Cheung (1972, p221).

[4] Canton is the old, English name of Guangzhou city in the past.

beginning with n for other speakers; for instance, *lee* for both 'you' and 'plum', whereas the pronunciation of the majority is *nee* for 'you' and *lee* for 'plum' " (Chao, 1947, p18). In a dialectal survey reported in Yuan (1960), the same change has been mentioned, but only with rough description. It is said that *most of* the Guangzhou speakers have a clear distinction between /n-/ and /l-/ and *only a few* people pronounce original /n-/ words with /l-/ initial. Also it is mentioned that the /l-/ in some speakers is pronounced "softly" and hard to differentiate with /n-/. The change seems to have progressed faster in Hong Kong, as discussed briefly in Cheung (1972, p1-2) and in Bauer (1983). Bauer remarks that "for many speakers, particularly young people, /n-/ has merged with /l-/, that is, speakers use /l-/ instead of /n-/ but not the reverse. ...While many people have no distinction between /n-/ and /l-/, their speech may show variation between the two with a tendency to pronounce /n-/ words with /l-/ even though they hear no difference between them and cannot make one either" (ibid, p29).

There are also pieces of evidence for the /n-/→/l-/ change attested in some dictionaries which incorporate contemporary colloquial pronunciations. For example, in a compiled homophone list published in 1953 in Hong Kong (Wei, 1953), at least four words, which are still transcribed with /n-/ in current dictionaries, are shown to have alternative pronunciations with /l-/[5].

The merger of /n-/ and /l-/ in fact is quite common in a wide range of Chinese dialects. According to the dialectal reports (Karlgren, 1937/1994; Yuan, 1960/2001; Hou, 2002; Zhan et al., 1990), similar changes of /n-/→ /l-/ are attested in dialects such as Nanjing (Yuan, 1960/2001, p30) and Nanchang (ibid, p128), etc. In some dialects this change is phonetically conditioned, depending on the following vowel. The change in the opposite direction, i.e., /l-/→/n-/, is also found, e.g., in Chongqing (ibid, p30). Moreover, in some dialects, it is reported that the /n-/ and /l-/ are in free variation, e.g., Lanzhou (ibid, p30).

The above reports are mainly based on analyses of a small number of speakers interviewed. It is not clear yet how accurate and comprehensive the above descriptions

[5]The four morphemes showing pronunciations with /l-/ in Wei (1953) include '弄' ("arrange") (ibid:32), '匿' ("hide") (ibid:33), '鸟' ("bird"), and '尿' ("urine") (ibid:34).

are in capturing the situation of the changes between /n-/ and /l-/ in the past of Guangzhou Cantonese and in other various dialects. The descriptions of the different changes, either /n-/→/l-/, /l-/→ /n-/, free variation, or the phonetic conditions may be just a partial picture reflected by a limited number of speakers. To have a full and accurate picture of the situation will require a larger sampling of the speech community. We arrived at this conclusion because we observed dramatic individual differences from the empirical study to be reported below.

While the above studies about the state of change are mostly based on a small number of speakers' speech data, several studies have been carried out to investigate in depth the actual situation, using systematic sociolinguistic methodology (e.g., Yeung, 1980; Pan, 1981; Bourgerie, 1990). These studies all examine the situation of Hong Kong Cantonese. They study the correlation between various social factors and the synchronic variation. Their data show consistently that the proportion of /n-/ is smaller in younger groups. The speakers in the youngest group in Yeung (1980) even show a complete loss of /n-/.

Following the trend of the change given by these earlier studies on Hong Kong Cantonese, we would expect that the change /n-/→/l-/ should have now progressed further and there should be less /n-/ instances found in the current Hong Kong community. A brief remark of the situation may be found in Zee (1999) " ... [n-] has merged with the syllable-initial [l-] in the speech of the young adult speakers" (p155)[6].

Compared with Hong Kong, the situation of Guangzhou is more complex. It is known that Putonghua (PTH) has been promoted as the standard language in mainland China as a language policy since 1955 (Chen, 1999b, p23). In Guangzhou, PTH is the official language used in schools since kindergarden and in public media including most of the radio and TV programs. In addition to the cultural diffusion, the PTH influence is strengthened by another force, that is, demographic diffusion.

[6]There is an interesting case showing the recent status of two morphemes in the process of /n-/→/l-/, found in a recent local newspaper in Hong Kong. In the front page of South Morning Post of Nov. 6, 2003, in celebrating its 100th anniversary of the first publication, the headline says 'Bak Lin Sau Hing' ("100 year anniversary"). The morpheme '年' ("year"), which is supposed to be pronounced as 'nin' according to the dictionaries, is spelled with an /l-/, suggesting the new pronunciation of /l-/ has already become a convention of the contemporary Hong Kong speech community. Interestingly on the same page, another morpheme '南' ('nam', "south"), which is also subject to the sound change, is still spelled with an /n-/ initial.

There have been several large waves of immigrations from other areas since 1950. According to an official demography statistics Guangzhou Renkou Zhi (1995), there have been three largest waves of immigrants in 1950, 1955 and 1956, each amounting to more than 400,000 people. At the same time the emigration was almost at the same degree as the immigration[7], as shown in Figure 3.1. Even though the immigrants might not be all PTH speakers, nor the emigrants all originally Cantonese speakers, considering the population size of the Guangzhou city was only around 1.5 million then, we can easily imagine how the population composition has changed after such a huge replacement of the local residents. In recent decades, due to its leading position in economic development, immigrants from the north continue to arrive steadily, though it is harder to estimate the amount of influx from the north due to the greater flexibility of population movement. The arrival of the large number of native PTH or northern dialect speakers and the increase of PTH's social status in daily communication must have a strong impact on contemporary Cantonese. Most of the local Cantonese people are bilinguals, though with different degrees of fluency in PTH. The contact with PTH through the immigrants and bilinguals should have have an impact on the on-going change in Cantonese of interest. As the words with /n-/ initials are still pronounced as /n-/ in PTH, it is expected that the learning and use of PTH will bring back the /n-/ initials in Cantonese speech.

In comparison, Hong Kong has been less subject to the influence of PTH. PTH has not been recognized as an official language until Hong Kong's return to China in 1997, and the proportions of PTH speakers or other northern dialect speakers so far are still very small with respect to the local population size[8]. Therefore the sound change of /n-/→/l-/ shared previously by Guangzhou and Hong Kong may have diverged in the last several decades. However, as this study will not address the comparison between Hong Kong and Guangzhou, we will leave this topic for exploration in later studies.

[7]While these data may not be completely accurate and inclusive, we assume that they provide a minimum estimation.

[8]According to the census data provided by the Hong Kong government, the proportion of Mandarin speakers in the last three censuses in 1991, 1996 and 2001 was only 1.1%, 1.1% and 0.9% respectively (Data from *http : //sc.info.gov.hk/gb/www.info.gov.hk/censtatd/chinese/hkstat/fas/*01*c/cd*0062001_*index.html*). But due to the recent flourishing development of the various connections between mainland and Hong Kong after 1997, especially the promotion of tourism to HK recently, the influence of PTH may increase dramatically in the coming years.

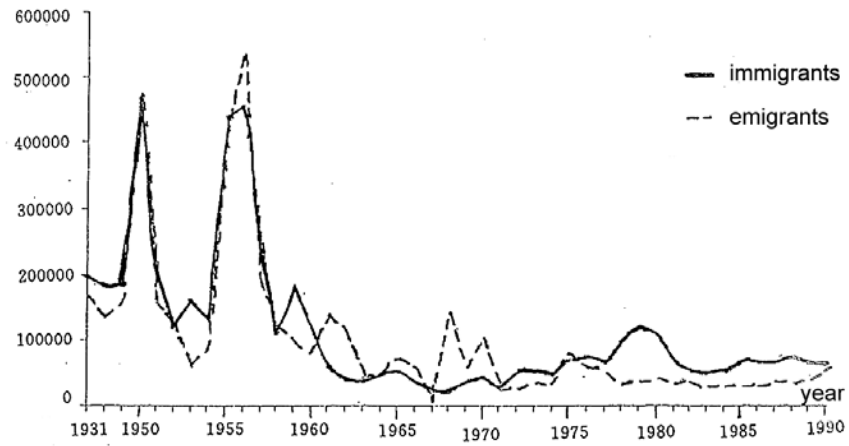


Figure 3.1: The number of immigrants to and emigrants from Guangzhou during the year of 1931-1990 (adapted from Guangzhou Guangzhou Renkou Zhi, 1995).

We will instead focus on the situation in Guangzhou, where two possible directions of change are in progress, i.e., the original /n-/→/l-/ and a reversal change /l-/→/n-/. At the moment it is not clear yet which change is leading the direction of the global change. But for the sake of simplicity, in the following discussion, "change" and "rate of change" refer to the change /n-/→/l-/ unless specified otherwise.

3.2.2 Methodology for data collection and processing

We use two methods to collect data of the synchronic variation between /n-/ and /l-/ in Guangzhou Cantonese. The first is to ask the speaker to read a word list. Word-list reading collects data of a formal speech style where speakers are normally very careful about their speech. Speakers may have stylistic variations and different speakers have different degrees of variation across styles. We expect this method will elicit variation at its lower bound. This would provide us a baseline for further comparison with casual speech in which more variation is expected. Meanwhile, word-list reading provides an effective way to collect a large number of words subject to the sound change, while recording spontaneous speech usually only collects data of a limited set of frequent words, and infrequent words may not have enough occurrences. Moreover, as we are particularly interested in the variation of specific words in single speakers, word-list reading can easily elicit an equal number of multiple tokens for a target word for comparison, while it is hard to obtain such data in spontaneous speech.

In Guangzhou Cantonese, there is almost no authentic "correct" pronunciation available to ordinary individual speakers. Due to the logographic nature of the Chinese writing system, word reading obtains fewer hints from the written characters, compared to that in English in which the spelling of words provides more phonetic information. Therefore we assume that the linguistic behavior in the n-/l- variation may not be too biased in word-list reading from less formal speech styles. However, we note that there is one possibility to introduce bias in the word-list reading, that is, some speakers may be aware of the correspondence between PTH and Cantonese, and they may consciously transfer the /n-/ in PTH to their pronunciation of words in Cantonese. The transference should be more likely in the situation of word-list reading than casual speech, as the written characters may facilitate influence from PTH. If so, the word-list reading experiment may attest more /n-/ than speakers would have in their casual speech. If the speaker is very conscious about the transference and consistently applies the rule, the word-list reading may give a very biased estimation of the degree of variation in the speaker. Bearing in mind the possible differences in casual speech, we also examined some speakers' casual speech when data require, by asking a few questions to elicit tokens of a few target morphemes which are frequent in daily communication.

We selected 33 target morphemes, among which 29 morphemes are assumed originally to have /n-/ as they belong to the ni-initial group in Qie Yun, and the other 4 morphemes (including 粒 "grain", 挪 "move", 鸟 "bird", and 朽 "decay") belong to other initial groups in MC[9]. Each morpheme is embedded with a disyllabic word or phrase, such as '农民' ("peasant") and '你好' ("how are you"), the first characters '农' and '你' being the target. In addition, the list includes 17 disyllabic words with the target morphemes in the second position, with the aim to test the effect of the position of the morpheme in a word. The 50 (32+17) target words are repeated three times and the total 150 words are randomized and divided into three sections. A copy of the

[9]粒 and 挪 belong to lai (来)-initial, 鸟 belongs to duan (端)-initial, and 朽 belongs to xiao (晓)-initial. 挪 and 鸟 are pronounced with /n-/ in PTH, while 粒 remains /l-/ and 朽 remains /x-/. These four morphemes are included because they are all given /n-/ in a standard Cantonese pronunciation dictionary (Zhan, 2002), which suggests that they were pronounced as /n-/ some time ago, at least in some speakers.

sheet with the list of words for the production test is given in Appendix II.

The collected production data are transcribed using a double-blind method to ensure the reliability and accuracy of the transcription. Then the three occurrences of each of the 50 words are compared to see whether there are different pronunciations for the same target morpheme. For each speaker, each target morpheme is assigned a state as one out of the four possibilities, either "n", "l", V or O, according to the consistency of the three tokens. If all three tokens are the same with /n-/ initials, set as "n"; if all /l-/, set as "l". If different initials appear, the state is set as V. O for other cases, such as when speakers fail in recognizing the morphemes or pronounce it with initials other than /n-/ or /l-/[10]. In order to test the reliability of occurrence of V from 3 tokens in one production test, we repeated the experiments with some of the speakers at a later time and compare the two sets of longitudinal data.

Due to the constraints of fieldwork conditions, it was hard to perform real perception tests for a large number of speakers. Therefore, we used a questionnaire of homophone judgment as a substitute of the perception test. The questionnaire design follows the methodology developed in Shen (1997), in which the subject is asked to judge whether three morphemes are considered as having the same pronunciation.

The target morpheme with an /n-/ initial is aligned with two other morphemes which share the same rime and tone but have an /l-/ initial. As some morphemes may have more than one pronunciation in different words, the morpheme in question is embedded in a disyllabic word and is underscored, in order to guide the subject to the target reading to make the judgment. For example,

<u>你</u>好	<u>李</u>平	<u>理</u>由
[lei] "you"	[lei] a surname	[lei] "a reason"

If the speaker has the sound change /n-/→/l-/, he would judge the three morphemes as homophones. Otherwise, he is expected to choose the target as different from the other two; the first morpheme '你' will be selected in the above example. Owing to the convenience of implementation, it is easy to collect a large amount of data

[10] For example, some young students or adults with less education cannot determine their Cantonese pronunciations for infrequent morphemes such as '腻', '诺', '聂', '朽', or pronounce them with initials other than /n-/ or /l-/.

efficiently which is hard to achieve with real perception tests due to the constraint of fieldwork condition. Shen (1997) used this method to collect more than 360 subjects' data for a phonological change in Shanghai, which are sufficiently large for him to examine the lexical diffusion pattern of the on-going change with a fine resolution in the apparent time. But as it is difficult to find many target morphemes for which we can come up a triplet as the above example, we only test 20 target morphemes. The questionnaire also includes two control items. A copy of the questionnaire is enclosed in Appendix II.

However, we note that what the questionnaire reflects is a speaker's perception of his own production. As the speaker's production is not traced during the experiment, we do not know what pronunciation the subject uses to do the homophone judgment. When the speaker chooses the target morphemes as different from the other two, we would like to assume that the speaker pronounces the target morpheme with an /n-/ initial, but this may not be true. It could be a sporadic special pronunciation which differs from an expected standard pronunciation. Therefore the questionnaire may include some untrue /n-/ instances. Such cases can be inferred when comparing the questionnaire data with the production data. Conversely, when the speaker judges a target morpheme as homophonous as the other two morphemes in the line, we cannot be sure if it is because the speaker has changed the target morpheme to /l-/ initial in his production, or he produces an /n-/ initial but does not distinguish it from /l-/ initial in his own perception. We in fact attest the latter case in our data, which will be discussed in a later section about near-merger.

3.2.3 Summary of collected data

The data collection was mainly carried out in Guangzhou in the first week of September and some more in late October, 2003. Speakers of the first age groups were mostly chosen from a year-one class of a middle school, and all around 12 years old; The majority of speakers of the second age group were from two year-one classes of two high schools, all around 15 years old. The first experiment was carried out in the first week of a new semester when these classes were all newly grouped and most of the students did not know each other before, so that their mutual influence can be ignored. In doing so, we can approach a large number of speakers effectively while ensuring a relatively random sampling. The speakers of the older age groups were mainly chosen through personal networks of the author. The diversity within the groups is maximized as much as possible, such as different social sectors, different educational background. But due to the time limit of the fieldwork, the numbers of speakers in the older groups are not as sufficient as the younger group.

We collected questionnaires from 155 speakers, and obtained production data from 42 of them[11]. The speakers were divided into six age groups, the distribution of which is shown in Table 3.2.

Table 3.2: Distribution of speakers in age groups and genders

		Questionnaire			Production		
Age group	Age range	M	F	Total	M	F	Total
1	10-14	22	22	44	2	5	7
2	15-24	44	37	81	6	5	11
3	25-34	3	6	9	2	5	7
4	35-44	3	2	5	3	1	4
5	45-54	4	6	10	4	5	9
6	> 54	1	5	6	0	4	4
	Total	77	78	155	17	25	42

We processed the data collected from the questionnaires and the product tests, according to the method described in the last section. Appendices 2.3 and 2.4 list these data in detail, including the pronunciation of the target morphemes in the 50 words by 42 speakers in the production tests (some speakers have two tests), and the 20 morphemes in the questionnaires filled in by 122 subjects who did the questionnaire twice.

3.3 Synchronic variation in the on-going change

3.3.1 Three types of variation

In the past investigations on sound change, two types of synchronic variation have been highlighted: one is the variation among speakers in the linguistic community,

[11] I would like to thank Ms. Yao Yao who helped in transcribing the collected data from fieldwork.

termed here as Type VT-I; and the other is the variation among words in the lexicon, termed here as Type VT-II.

Considering an on-going sound change from an unchanged form (U) to a changed (innovative) form (C) which involves a number of words, it is easy to find cases of words in the variation state (V), that is, both forms of the word co-existing in the linguistic community. At a given time, some speakers use the C form while others use the U form for a particular word. Different speakers may exhibit variation in response to the change. Some speakers adopt more changed forms than others. Most of the sociolinguistic studies focus on the analyses of this type of variation (VT-I) (Labov, 1972). By comparing the different degrees of U and C in individual speakers, they show that the linguistic community exhibits "orderly heterogeneity" (Weinreich et al., 1968), as the variations show correlation with various factors such as gender, age, socioeconomic status, education background, speech style, etc.

The second type of variation, VT-II, lies in the lexicon. Most sound changes affect not only a single word, but a number of words in the lexicon. It has been theoretically argued and empirically demonstrated that sound changes progress through lexical diffusion (e.g., Chen and Wang (1975); Wang (1969a, 1977), among many others). The words do not change all at the same time or at the same rate. Some words may lead the change and others join the change at later times. Words under the same phonetic conditions, even homophones, may have different rates of change. Such variation in the lexicon has been widely observed in on-going changes (Bauer, 1983; Shen, 1997), from historical linguistic analyses (Wang and Cheng, 1977), and even in children's language acquisition (Ferguson and Farwell, 1975; Hsieh, 1972).

The above two types of variation are usually discussed at the group level, either groups of words, or groups of speakers. For variation VT-I, as the analyses are focused on the differences between speakers, all the words subject to the change are taken into account as a whole or as groups, and the differences between individual words are not scrutinized. For variation VT-II, words are examined individually; however, the analyses of the status of different words are often based on a single speaker of a group of speakers' data, and the differences between individual speakers are often not explored.

In this study, we would like to draw the attention to a finer level of analysis, which is on the variation of individual words within individual speakers, termed here as VT-III. For a particular word undergoing a sound change, the two competing forms may co-exist in one speaker. Therefore, for each word involved in the sound change, three possible states, i.e., the unchanged (U), changed (C), and variation (V), not only exist in the population, but also in individual speakers. The existence of state V in a speaker may not be surprising at all. Even from a cursory observation or reflection would reveal many such instances. For example, many English speakers would pronounce both [rut] and [rʊt] for 'root', both [rɪ'sətʃ] and ['rɪsətʃ] for 'research'. Wang (1969a) remarks when proposing the hypothesis of lexical diffusion: "at any given time in any living language, we should expect to find several sets of morphemes with dual pronunciations. . . . In actual fact, of course, many of the dual pronunciations are used by the same speaker" (p14). Quantitative data concerning such variation can be found in the literature. For instance, a study of the on-going change of syllabic [ŋ̩]→[m̩] in Hong Kong Cantonese shows that 36% of the 75 speakers pronounce the word 'five' as both [ŋ̩] and [m̩] in their spontaneous speech, while other speakers consistently use one or the other pronunciation (based on data in Bauer (1983, p324-325).

However, though the existence of such VT-III within speakers may appear to sociolinguists as an obvious fact, there have been no specific attention to this type of variation. Our interest in VT-III is due to the following considerations. First, we are interested in the internal representation of such variation, and a detailed examination of the situation may give us some hints. Very often in describing a change, the occurrence of the variation of a word in a speaker is assumed to be governed by a phonological rule with or without conditions. For example, for the change of interest, a rule may be proposed as follows,

(1) [n-]→ [l-] / <conditions . . . >.

However, we doubt how real such a rule is in individual speakers. The observed variation of some words may just be the result of the alternative occurrence of two variant forms, which are learnt as two individual lexical items, in the way two synonyms are learnt. This conception of the variation shares some similarities with the exemplar model proposed by Joan Bybee. In a recent paper (Bybee, 2002), she adopts this model to explain the /t,d/ deletion in American English, which has been addressed extensively with different methods (e.g., Guy, 1980). Instead of explaining with a rule, she suggests that "all phonetic variants of a word are stored in memory and organized into a cluster", and that "these exemplar clusters ... change as experience with language changes. Repeated exemplars within the cluster grow stronger, and less frequently used ones may fade over time, as other memories do" (Bybee, 2002, p271). The exemplar model is able to account for the frequency effect on the words' variation, which is hard for the rule models to address unless they assume every word has a rule for its own. However, it is not clear yet if the exemplar model can effectively address the evident rule-like behavior in some speakers. Before committing to any particular model to account for the variation in words and in speakers, it is necessary to carry out a careful examination on a fine level of variation, i.e., VT-III in individual words in single speakers.

Furthermore, the interest in variation within speakers concerns the question "what would the child learn being exposed to variant forms during an on-going sound change". Shen (1997) has raised this question and remarks that "issues related to phonological acquisition during a sound change have not been seriously addressed" (p154, emphasis original). When a sound change happens, the variation in the community and in the lexicon creates a complex situation for the learners. Some studies have shown that children are able to acquire the sophisticated variable patterns in some on-going sound changes (Roberts, 2002). However, these studies do not pay attention to the differences between children. Adults' data show that some speakers learn to use variant forms, while some speakers do not. Such differences should be traced back to acquisition. They may be ascribed to the difference of the linguistic environment that the individuals have been exposed to. But how about the differences between people who are of the same age, who are siblings or even identical twins? If there still exist significant individual differences for these cases, then we need to resort to differences in individual learning styles or more general cognitive capacities. The individual differences shown in an on-going change may reveal some patterns which are not easy to observe in an idealized acquisition situation in which a homogeneous linguistic environment is assumed.

In addition to the variation between speakers, we are interested in knowing how individual words vary. Most of the time, the studies on the synchronic variation between speakers do not pay attention to the differences between individual words. To one extent, the words are not differentiated at all, but assumed to be all in a state "characterized by the presence of more or less free variants, so that the speakers have the choice between alternative expressions. ...most often the choice will appear as being due to pure chance" (Vogt, 1954, p367) (quoted in Wang (1969a), p14). Such description of "free variation" between changing sounds can be attested in many reports of phonological systems. However, as noted by Fischer (1958), " 'Free variation' is of course a label, not an explanation. It does not tell us where the variants came from nor why the speakers use them in different proportions, but is rather a way of excluding such questions from the scope of immediate inquiry" (p47-48).

Since the mid-1960s, the variable rule analysis method has been proposed to study the synchronic variation systematically, by sorting out the effects of various linguistic and non-linguistic constraints. It is assumed that a single variable rule is at work uniformly across speakers, with different probabilities applied under different conditions (Labov, 1969; Sankoff and Labov, 1979). In these analyses, words are often grouped together according to various language internal criteria, such as the phonetic environment, morphological and/or syntactic conditions, etc. There have been few studies, however, which examine the relevant lexicon word by word. And very often since data are collected from spontaneous speech, it is hard to control the number of tokens of relevant words and the variation is often measured based on the whole set of tokens of a group of words. In such analyses, characteristics of individual words may have been overlooked.

On the other hand, it has been repeatedly shown that "every word has a history of its own" in a sound change (Malkiel, 1967). During the process of a sound change through lexical diffusion, words with the same phonetic condition may fall into different stages of the change (U, V or C) at some point in time. For example, 'hoot', 'root' and 'foot', all subject to the o-shortening, belong to three different states at present (Wang, 1979). Such variation, however, is often observed based on the linguist's own data or on a corpus which is a collection of data from a number of speakers. In this study, we suggest it is worthwhile to examine the status of individual words in each speaker. It would not be surprising to see that different words have different histories in one speaker as well.

In the following, I will present the observed three types of variation in order, i.e., variation among speakers (VT-I), variation among words (VT-II), and variation in words within single speakers (VT-III). The analyses are mainly based on the data of one production test.

3.3.2 Variation in the population: VT-I

We observe a large degree of variation across speakers. Speakers within one age group exhibit dramatic differences, even more among speakers across different age groups. Figure 3.2 shows the heterogeneity in the three young age groups, illustrated by the percentages of four states among the whole set of target morphemes in each speaker. For example, among the 11 speakers in age group 2, while one male speaker (no. 3) has 100% /l-/, another male speaker (no. 6) has 8% /l-/ and 76% /n-/. In the presence of such remarkable individual differences, we should be cautious in describing a language or a dialect only based on one or two speakers. It seems necessary to obtain data from a sufficiently large number of speakers in the speech community, in order to have a more complete picture of the synchronic situation.

We take the percentage of /n-/ as the variable to measure the degree of heterogeneity within the group. Figure 3.3 shows the percentages of /n-/ of individual speakers in the six age groups, the median values, the average values, and the standard deviation in the six age groups. For the three young groups, the differences between maximum and minimum values are quite large. For example, in group 3 the values

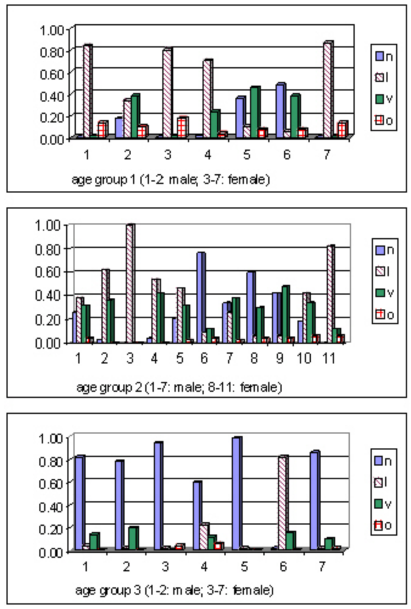


Figure 3.2: The distribution of states of target morphemes in speakers of three young age groups.

spread from 2% to 98%, and in group 2 from 0% to 76%. We need to address why there exist such large degrees of variation in the population, and at the same time, the extreme cases are also important and deserve special attention: Why does a speaker in group 3 (no. 6) pronounce only 2% of the target morphemes with /n-/ initials while the average of the group has about 70% /n-/? Why does a speaker in group 2 (no. 6) have 76% /n-/ initial while the average value of the group is only about 26%? Should these extreme cases be ignored as outliers with no theoretical relevance? We suggest the opposite: these extreme cases may reveal individual differences which would be missed if we only pay attention to the average cases, or only examine the overall statistical pattern of groups.

3.3.3 Variation in the lexicon: VT-II

While speakers in the population exhibit great heterogeneity, the relevant words in the lexicon also show a large degree of variation. This is an inevitable intermediate stage of an on-going sound change as a result of lexical diffusion. Some morphemes have changed faster than others, i.e., having a larger frequency of /l-/, assuming the change as /n-/→/l-/. Figure 3.4 shows the distribution of the three states (n, l, or V) for

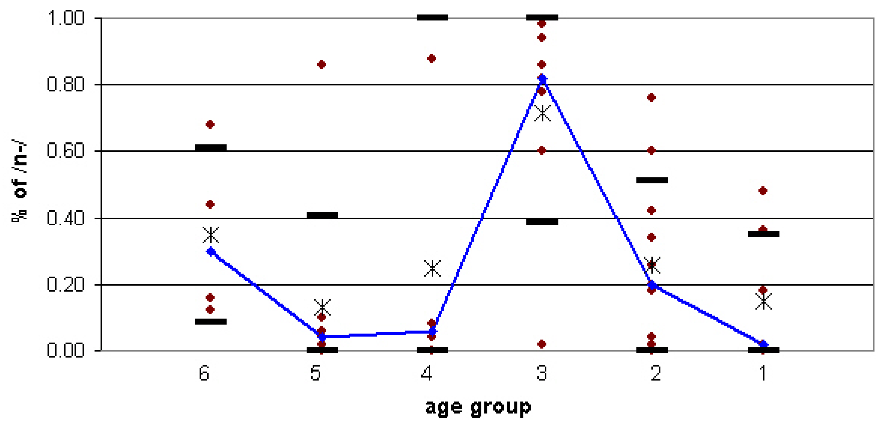


Figure 3.3: An illustration of heterogeneity in the six age groups (from left to right: older to younger groups). The percentages of /n-/ in speakers are represented by dots. One dot may represent more than one speaker. The line connects the median values, the star denotes the average value and the horizontal bars denote the standard deviations in each group.

each of the 33 target morphemes. The graph in the upper panel shows the situation when the target morphemes are placed in the first position of disyllabic words, and the lower panel is for morphemes in the second position. The different rates of change imply that morphemes are not affected by the sound change all at the same time, but change at their own paces. This demonstrates what lexical diffusion theory predicts: sound change takes place "as a kind of diffusion from morpheme to morpheme" (Wang, 1969/1991, p7), instead of changing all relevant words or morphemes in block.

Not only do different morphemes change at different rates, but the same morpheme in different position in disyllabic words exhibits differential rates of change. Comparing the two panels for the same set of 17 morphemes, we can see that morphemes behave differently when they are in different positions of words. One of the largest differences is the morpheme '年' ("year"). In '去年' ("last year") the percentage of /l-/ is 41% while in '年代' ("age") only 14%. Examining the consistency of morphemes between the two positions in the 42 speakers, we find that speakers vary greatly: four speakers show consistency over 90%, while one speaker only shows 12%; the average consistency of the two positions for all speakers is about 66%.

As the same morpheme in different words exhibits different rates of change, would this suggest that the diffusion is based on individual words instead of morphemes?

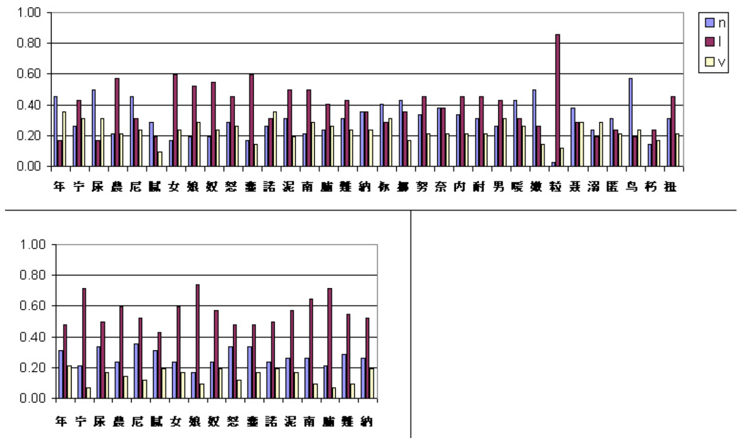


Figure 3.4: The percentages of different states of 33 target morphemes in the 50 words. Target morphemes in the upper panel are in the first position of a disyllabic word, while the 17 morphemes in the lower panel are in the second position of words.

We cannot draw any clear conclusion so far, because we lack a type of data in which the target morphemes are in the same position of different words, for example, '年代' and '年月'. Interestingly, however, we observe an effect of position: morphemes in the second position consistently have larger percentages of /l-/ than those in the first position, except for one morpheme '囊'. The difference is statistically significant ($t = -3.19$, $p<0.005$, one-tailed t-test). The situation of /n-/ and /l-/ is not symmetrical: while the percentages of /l-/ in the second position are significantly larger than those in the first position, the percentages of /n-/ in the first position are not significantly larger than those in the second position.

The consistently larger percentages of /l-/ in the second position are puzzling. It could be due to the speakers' greater awareness of their pronunciations. It is possible that due to the word-list reading setting, speakers are more aware of the pronunciation of the first morpheme in a disyllabic word than that in the second position. If so, the larger percentages of /l-/ in the second position may suggest that when speakers are less aware of their production, they tend to pronounce more /l-/ than /n-/. This is consistent with what earlier studies on Hong Kong Cantonese have found (e.g., Bourgerie, 1990; Pan, 1981): speakers produce more /l-/ in casual speech than in formal situations. This explanation has one implication, which is that the speakers have the knowledge that the /n-/ pronunciation is the formal or "correct" one. It is possible for some speakers who have generalized such a rule which transfers PTH's /n-/ pronunciation to Cantonese and applied it consciously in the word-list reading task.

However, if the above explanation is true, then one question arises: if speakers are really consciously applying the rule, why don't most of the speakers transfer all relevant morphemes, but only a small number of morphemes? The variable rule theory may suggest that it is because there are different probabilities in the rule application on different words: some words have a higher probability than others for the rule to apply. Early studies have shown that various linguistic factors systematically determine the rates of change, including phonetic condition, frequency, etc.

We indeed observe the effect of phonetic condition from our data. In particular, there is a pair of homophonous morphemes in the list, i.e., '南' ("south") and '男' ("man") /nam4/. Their percentages of /n-/ and /l-/ are quite close: the first has 24.3% /n-/ and 45.9% /l-/ while the latter has 29.7% /n-/ and 40.5% /l-/, and they have the same percentage of V state (29.7%). Moreover, examining the individual speakers, we observe that 68% of them show the same state for these two morphemes; and that no speaker has a consistent /n-/ for one morpheme and an /l-/ for the other. This pair of words may remain homophonous and has no tendency to split.

Among the first 10 morphemes which have the largest percentages of /n-/[12], six of them have a high front vowel, either unrounded /i/ or rounded /y/, following the initial. Furthermore, among the first 10 morphemes with the highest percentages of /l-/[13], it seems that a dissimilation effect may play a role in the change, as 6 out of the 10 morphemes have nasal endings, either /-ŋ/ or /-m/. This dissimilation effect has also been reported in Shantou, a variety of Min dialect (Karlgren, 1937/1994:356-357)[14]. Systematic analyses on the effect of phonetic environment using the variable

[12] The first 10 morphemes having the largest percentages of /n-/ include 鸟类 (/niu5/), 尿桶 (/niu6/), 嫩绿 (/nyn6/), 年代 (/nin4/), 尼姑 (/nɐi4/), 挪动 (/nɔ4/), 暖气 (/nyn5/), 你好 (/nei5/), 奈何 (/nɔi6/), and 聂耳 (/nip9/).

[13] The first 10 morphemes having the highest percentages of /l-/ include 粒子 (/lɐp7/), 姑娘 (/lœŋ4/), 西宁 (/lɪŋ4/), 牛腩 (/lam5/), 西南 (/lam4/), 花农 (/loŋ4/), 少女 (/lɸy2/), 女兵 (/lɸy2/), 囊括 (/lɔŋ4/), 花泥 (/nɐi4/).

[14] Contrary to the dissimilation effect in Shantou dialect, in another Chinese dialect, Kejia, the nasal endings /-m/, /-n/, and /-ŋ/ have the assimilation effect to retain /n-/ initials (Karlgren, 1936/1994, p356-357).

rule method may be able to test the significance of these conditions, though the effect of the phonetic conditions may be blurred by other factors, such as the frequency of words, in such a complex bi-directional change.

It has been shown repeatedly that the frequency of morphemes or words is one important factor determining their rates of change (Bybee, 2002; Hooper, 1976; Phillips, 1984). Due to the lack of reliable frequency data of Guangzhou Cantonese, it is hard to test this hypothesis in the current study. However, from a cursory check, the frequency effect does not seem applicable at least in some words. In the above 10 morphemes with highest percentage of /n-/, there are both high frequency constructions, such as '你好' ("how are you"), as well as literary constructions which should have low frequency in speech, e.g., '挪' ("to move").

3.3.4 The puzzle of 'li4' ("grain")

The morpheme '粒' (/lɐp7, "grain") deserves some detailed discussion. It has the largest percentage of /l-/: 36 out of the 42 speakers give consistent /l-/; only 2 speakers consistently pronounce it with /n-/ initial and 6 speakers give varying pronunciations. This morpheme can be assumed to have had an /l-/ initial in Middle Chinese, since it was grouped with lai-initial ('来' 母) morphemes in the rime book Qie Yun (切韵). Then it is not surprising that the morpheme has a specially high /l-/ percentage, if the /l-/ pronunciation is supposed to be the original one from Middle Chinese and there has been no change from /l-/ to /n-/ in Cantonese before PTH's influence.

The specially high percentage of /l-/ of '粒' may be reinforced by the influence of writing system as well, because the written form of the morpheme '粒' has a phonetic component '粒' which is unambiguously pronounced with an /l-/ initial. Therefore it is possible that the phonetic component in the written form available to the speaker in the word-list reading task would increase the occurrence of /l-/ initial. If that is one reason, a naming task without the visual cue of the written form would elicit more /n-/ than what we have observed.

However, where does the /n-/ pronunciation come from? In fact the pronunciation with an /n-/ initial has been attested in historical Cantonese repeatedly, such as in a Cantonese-English dictionary compiled by Eitel in 1877. A dictionary of Cantonese pronunciation published recently (Zhan, 2002) also gives /n-/ for this morpheme. It is unclear whether these transcriptions really reflect the majority of the community of the time of the compilation of the dictionary, or just a small number of speakers who were accidentally selected as the informants. In Hashimoto (1972) it is noted that a small number of speakers pronounce '粒' with /n-/ initial while /l-/ is used by the majority. We suspect that Hashimoto's observation is closer to the real situation of her time, since this situation is reflected in our data about 30 years later. However, as the direction of change in Cantonese seems to be from /n-/ to /l-/, we still need to explain how the opposite direction of change could have happened, i.e., how this particular morpheme originally with an /l-/ initial has acquired the pronunciation with an /n-/ initial, and remains in the population even though only in a small proportion.

The pronunciation of /n-/ initial for '粒' may have originated from some speakers who have mixed /n-/ with /l-/ and thus pronounce /l-/ as /n-/ in some or all morphemes, or from some speakers who have hypercorrection (Ohala, 1993) and deliberately change /l-/ to /n-/. The former has been attested in one speaker in this study: a male speaker in the age group 3 sporadically pronounces some /l-/ initial morphemes with /n-/. Unfortunately, there is no formal analysis in this study for this change /l-/→/n-/ for morphemes other than the 27 targets. Nevertheless, the question still remains: how could such sporadic pronunciation in individual speakers have spread sufficiently to be noted down by lexicographers? Regarding its idiosyncrasy, '粒' with /n-/ could have been borrowed from other languages or dialects, for example, minority languages such as Zhuang, where '粒' is pronounced as /nat/ (Zhuangdong Yuzu Yuyan Cihui, 1985, p164), or other Chinese dialects, such as Yangjiang Cantonese variety, where '粒' is pronounced with an /n-/ initial (Hanyu Fangyin Zihui, 1989)[15]. The history of this particular morpheme will be an interesting question for further study.

[15]In Yangjiang dialect '粒' is also the only morpheme in the set of morphemes in 'lai' initial ('来母') of Middle Chinese, which is now pronounced with /n-/(Hanyu Fangyin Zihui, 1989). Another Chinese dialect, Chaozhou, has more interesting splits of the 'lai' initial morphemes: while most of the morphemes remain as /l-/ initial, a number of morphemes have both /n-/ and /l-/ pronunciations for colloquial and literal reading respectively, and some other morphemes (e.g., 冷，领，岭) are pronounced with /n-/ initials.

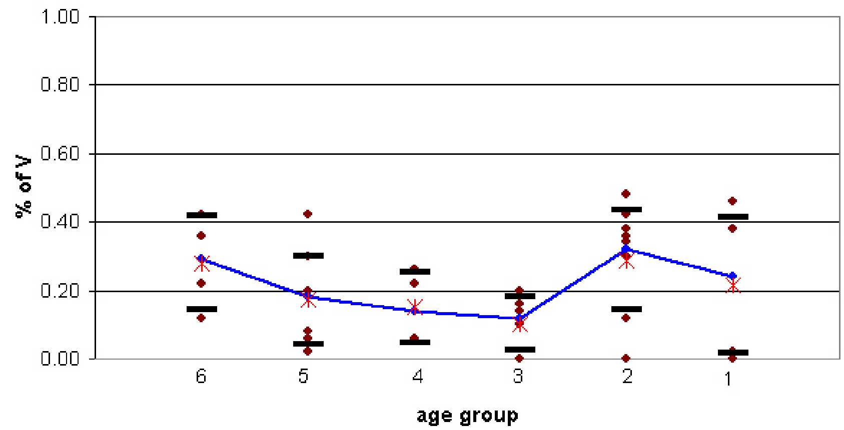


Figure 3.5: Degrees of variation type VT-III in 42 speakers, each represented with a dot. The line connects the median values, the star denotes the average value and the horizontal bars denote the standard deviation in each group.

3.3.5 Variation in the individual: VT-III

After we have seen the variation between speakers and variation between words, we will examine the variation within single speakers. This is the variation VT-III this study would like to highlight. The degree of VT-III in a speaker is measured by how many morphemes among the target set of morphemes are in the V state. Figure 3.5 shows the VT-III in each of the 42 speakers, arranged in 6 age groups.

Most of the 42 speakers (92%) show VT-III in the first production test. The degrees of VT-III in the speakers are all less than 50%, and the largest degree is 48% (24 morphemes out of 50) in a female speaker in group 2. The overall average for all speakers is only 21%. The average and median values in the six age groups remain relatively constant. A test using ANOVA shows that the differences between the six age groups are not statistically significant ($p = 0.11$).

Four speakers show no existence of VT-III at all; they pronounce all three tokens for each of the 50 morphemes consistently. Among them, two speakers have no /n-/ initials, while the other two have no /l-/ initials (except for the morpheme '粒' which is the special case explained above). It seems that when a speaker has no V, he must consistently adopt only one of the two competing sounds for all relevant morphemes. According to this observation and the preceding ones, it is impossible for a speaker to allow a subset of morphemes consistently with /n-/ initials, while the rest of the morphemes show /l-/ initials. This leads to the following hypothesis, which will be discussed further in connection with Table 3.3:

> VT-III Hypothesis: In an on-going sound merger A→B which affects a group of words, if a speaker has a set of morphemes consistently with A, and some other morphemes consistently with B, then he must have some morphemes in the V state, i.e., these morphemes have both A and B pronunciations.

On the one hand, the existence of V state seems inevitable in most speakers. On the other hand, the degrees of VT-III seem limited in speakers, which suggests that the variation is not "free variation", as discussed earlier. Were it free variation, which means the occurrence of /n-/ or /l-/ is completely random, i.e., each with a probability of 50%, then the probability of having a morpheme in V state from three tokens would be 75%[16]. If the 50 morphemes are independent from each other in their pronunciations (though this may not be true as morphemes with similar phonetic shapes may have a higher probability to change together), the average probability of all morphemes in V state is just 75%, much higher than the observed maximum value 48%.

The above analyses are based on the data of one production test, in which each word has only three tokens. We are aware that the small number of tokens from a word-list reading task may not be sufficient to elicit variation. Given more chances, speakers may exhibit more VT-III. To test this hypothesis, 10 speakers from the 42 speakers in the first experiment were asked to take a second production test some time after the first test[17]. Comparing the data of the two sessions, we do observe a certain degree of inconsistency.

We measure the consistency between the two sessions as the percentage of morphemes which share the same state in the two sessions. Figure 6 shows the consistency

[16]When the two pronunciations of a morpheme are in completely free variation, i.e., the probabilities for /n-/ and /l-/ to occur are equally 50%, the probability of eliciting the variation state (V) with 3 tokens is 1- 2*(0.5*0.5*0.5) = 0.75. If the variation is not completely free, say, the probability ratio for /l-/ and /n-/ to occur is 1:3, then the probability for V is 1- 0.75*0.75*0.75- 0.25*0.25*0.25 = 0.56. However, if the variation is too biased, the V state is not meaningful anymore.

[17]The distance between the two tests varied for different speakers, some two hours, some two weeks.

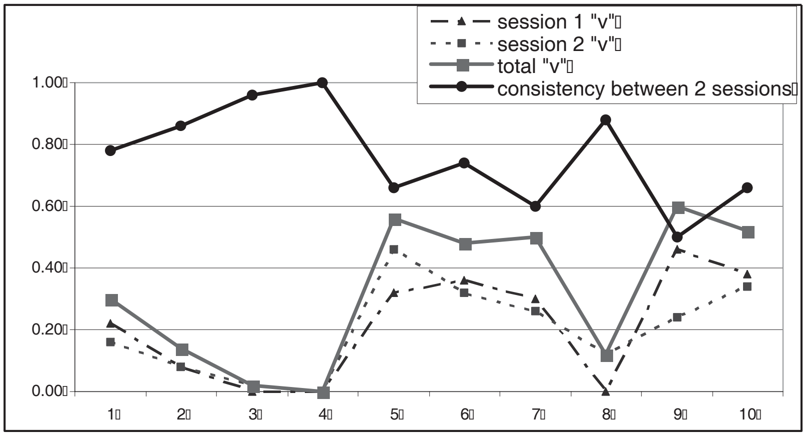


Figure 3.6: The percentages of V in two sessions of word-list reading and the consistency between the two sessions..

of the 10 speakers, in terms of the percentages of V state in the two sessions and the total percentage of V. Nine out of the 10 speakers show inconsistency. The average consistency of the ten speakers is about 76%. The worst case is a 11-year old primary school student (the 9th speaker in Figure 3.6), who only has 50% morphemes consistent between the two tests. The degrees of VT-III exhibited in her production of the two sessions differ greatly (46% vs 24%) and the total degree of variation combining the two tests is 60%. The inconsistency in speakers reminds us again the importance of sufficient sampling in order to obtain a representative picture of the dialect or language of interest, not only in terms of sufficient number of speakers, but also sufficient number of tokens for each word in single speakers.

Though there exists a certain degree of discrepancy between the two tests, the degree of VT-III based on one production test is still informative. There exists a high correlation between the VT-III from one test and the combined VT-III from two tests (Pearson correlation coefficient r = 0.96). Moreover, when more tokens are elicited, the degree of variation does not increase as the worst case would predict. In other words, the combined degree of VT-III is always smaller than the sum of the VT-III variations of the two tests.

In order to check how much the degree of VT-III would increase when more tokens are available, we asked the speaker who showed the largest degree of VT-III in the first two tests to perform two more tests. It turns out that her VT-III increases from

60% to 70%, and to 74%. Therefore, it seems true that given more tokens the degree of VT-III of this speaker will continue to increase. However, we speculate that it is unlikely that her degree of VT-III will keep increasing until 100% to show a complete “free variation” between /n-/ and /l-/, because after 4 production tests this speaker still has 13 morphemes with invariant pronunciation. Among them, there are some infrequent morphemes pronounced with /n-/, which usually occur in literary words such as ‘挪动’ (‘to move’) and ‘鸟类’ (‘the species of bird’). It is highly possible that the pronunciations of these morphemes are determined by the corresponding PTH pronunciations. Meanwhile there are frequent constructions such as ‘你好’ (“how are you”) which are found to be consistently pronounced with /n-/ even in her casual speech. These pronunciations seem unlikely to vary.

For other speakers who have smaller degrees of VT-III, it is even less likely that they will exhibit “free variation” for the whole set of morphemes. Therefore, though the linguistic community as a whole incorporates all instances of variants of the morphemes, a speaker only learns to use the variants of a small number of morphemes. The “free variation” phonologists have discussed may be true at the population level, but it does not seem to be true at the individual level. Though we do not exclude the possibility that some speakers would show free variation between the two sounds in all the relevant morphemes when casual speech is examined, we suspect that for most of the speakers, the VT-III is only applicable to a limited number of morphemes.

The limitation of V could be due to the availability of the variants in the linguistic input to the individual speakers in their learning period. In other words, the individual learners only hear a small number of variant tokens. Another possible reason for limited VT-III may be due to some cognitive constraint in lexical learning, which would limit the capacity of speakers to learn variant forms for one meaning, similar to learning synonyms. It would be interesting for future work to know what else psychological or cognitive constraint would pose such a limit.

While recognizing the existence of VT-III in most speakers, we need to explain why some speakers do not show the co-existence of variants as other speakers? It could be due to the limited number of tokens collected in the production tests, which is not enough to elicit the variation. Providing more tests, the variation may appear in these speakers. The cases of speaker no. 3 and 8 illustrate this point, as they have no VT-III in the first test but exhibits VT-III in the second test.

Another reason for the absence of VT-III may be due to the word-list reading task, in which the speaker becomes more aware of their speech and therefore deliberately willing to be consistent in their pronunciation. It turns out that speaker no. 4 belongs to this type. She consistently has only /n-/ in the two production tests, but observations from her casual speech reveal some /l-/ instances in morphemes such as '男' ("man"), '女' ("woman"), and '你' ("you"), though she consistently maintains /n-/ for morphemes like '难' ("difficult") and '南' ("south"). It seems implausible that the speaker can keep track of her pronunciation during the process of reading a long list of words (150 words in one production test) and thus show zero degree of VT-III. Instead, we hypothesize she consciously transfers the PTH initial of the target morphemes into her Cantonese pronunciations. In other words, she seems to have a rule /l-/→/n-/ operating when reading the word-list.

However, even though 3 out of the 4 speakers who show no VT-III in the first test appear to have variation with more tests, or in another speech style, we cannot exclude the possibility of the existence of speakers having no V. As we mentioned earlier, one possible situation is that a speaker is embedded by chance in a linguistic environment without any variation. Though this is very unlikely in current Guangzhou Cantonese as our data has revealed a great deal of heterogeneity in the community, we expect that in the early and last stage of a change, such environment could be found. Another reason could be that some speakers may have a specific learning style in lexical learning and phonological development, which does not accommodate variant pronunciations for the words. This will be discussed in more detail in the section on individual difference in learning styles.

3.4 Disparity between production and perception

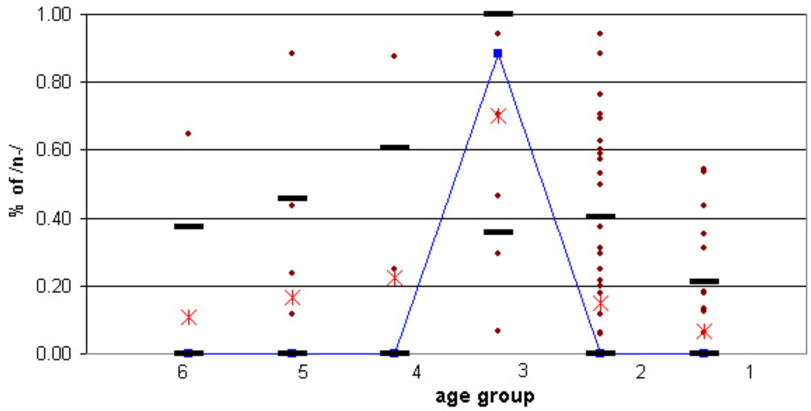


Figure 3.7: The percentages of /n-/ in 155 subjects in the six age groups shown in the questionnaire. The line connects the median values, the star denotes the average value and the horizontal bars denote the standard deviations in each group.

3.4.1 Variations VT-I and VT-II in questionnaire data

The above analyses show the heterogeneity exhibited in speakers' production. The data we obtained from the 155 questionnaires also demonstrate a similar degree of heterogeneity. Figure 3.7 gives the percentages of /n-/ in the six age groups, together with the median values, average values and standard deviations within each group. Speakers in the same age group exhibit a large degree of differences in the percentages of /n-/. Different morphemes show different rates of change as well.

Comparing Figure 3.3 for the production data and Figure 3.7 for the questionnaire data, we observe a similar uni-modal pattern, as represented by the median values of /n-/ in the six age groups, though production data consistently show more /n-/s in the older and younger groups. The similarity between the two sets of data is unexpected. On the one hand, the production test has a much smaller sample size than questionnaire; on the other hand, the questionnaire mixes a speaker's production and perception. It seems that the data collected by the two methods are quite comparable, and we may have the flexibility in choosing one from the two, dependent on the purpose of the study. If the study is to reconstruct the diachronic profile of the change, the questionnaire seems to a more effective method as it allows collecting a large amount of data efficiently. But if the study is to investigate the individual linguistic behavior in the on-going sound change, the production data are more informative, as it can

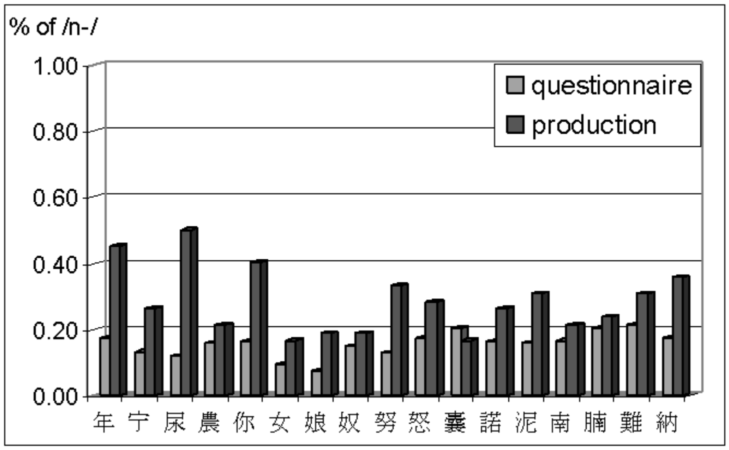


Figure 3.8: Percentages of /n-/ in 17 morphemes from production and questionnaire data.

collect more tokens and elicit variation in words which is difficult for questionnaire.

Figure 3.8 shows the percentages of /n-/ in the 17 morphemes reflected in the questionnaires, in comparison with the results from the production test. For most of the morphemes, the percentages of /n-/ from the questionnaires are much smaller than those in the production test, except for one morpheme ('囊') whose high percentage of /l-/ in the production we do not have any explanation for yet. The differences between production and questionnaire will be explained in the next section.

3.4.2 Cases of near-merger

All age groups, except group 5[18], show smaller percentages of /n-/ in the questionnaire than in the word-list reading. The discrepancy is because that there are some speakers whose data from the questionnaire show the merger of /n-/ and /l-/ in some target morphemes, but in the production test they consistently produced /n-/ for these morphemes. These are cases of near-merger, which have been reported in several early studies (see a review in (Labov, 1994, Chapter 12). Among the 42 speakers, 23 of them (about 54%) show the near-merger of /n-/ and /l-/.

One of the earliest examples of near-merger was presented in Labov (1972), a study on New York City of a vowel merger between /oh/ and /ohr/. For a minimal

[18]In age group 5, two speakers' questionnaires are problematic, because they give unexpected /n-/ for some morphemes but they consistently pronounce /l-/ in word-list reading. This results in an unreasonably high percentage of /n-/ in questionnaires in that age group.

pair "sauce" and "source", six speakers produced the two words with a marked phonetic difference, but judged the words to be the same. This inconsistency between production and perception is considered as paradoxical: "how does a person learn to articulate each member of one category in one way, and each member of the other category in another way, if he or she cannot recognize the difference between the categories?" (Labov, 1994, p368). Early studies of near-merger have been mainly vowel mergers, in which the phonetic distinctiveness of merging vowels is not clear cut since the vowels vary in a continuous dimension[19]. The near-merger in consonants reported in this study may provide more striking evidence for the paradoxical situation. The two merging consonants, /n-/ and /l-/, are phonetically so different in a discrete manner[20]. How could such a disparity between production and perception be possible?

To answer why the paradoxical situation of near-merger is possible, Labov and his associates have proposed an explanation as "suspension of phonemic contrast" (Labov, 1994, p391ff). They suggest that the semantic contrast utilized by the two sounds are reduced, and therefore even those who have clear distinction of the two sounds in their own production do not attend to the distinction in stimuli presented to them. However, when the semantic distinction is important in a linguistic task, the subjects show higher accuracy in distinguishing the two sounds. This can be reflected in the comparison of the identification accuracy between an experiment of semantic disambiguation in the coach test and a categorization task in the commutation test, as shown in Figure 8 in Labov et al. (1991, p68). We expect that if a more refined experimental design is used to make sure that the speakers do focus on the pure

[19]In most of the studies reported by Labov et. al, the vowels are shown in a 2D space constituting of the first two formants, i.e., F1-F2. It is not clear whether it is true that two vowels are really distinctive when they appear distinct in the F1-F2 space. The differences in these two dimensions may be compensated by higher formants, so that two vowels differing in F1 and F2 do not necessarily mean they are distinctive vowels. On the other hand, re-examining the data given in Labov (1994), the example of near-merger in Philadelphia suggests that the paradox may not really exist in some cases. As shown in Figure 9.4-9.8 therein, when the phonetic differences are distinct enough, the subject has high accuracy in the commutation test; the errors mostly occur in those stimuli which fall in the overlapping area of the two merging sounds.

[20]Though the two sounds /n-/ and /l-/ seem different in a discrete manner in the articulatory dimension, some early studies have mentioned that some Cantonese speakers' pronunciations of the /n-/ and /l-/ are perceptual indistinguishable (Yuan, 1960/2001, p181). It is suspected there is an intermediate sound between /n-/ and /l-/, such as Karlgren (1936/1994) where he calls the sound "naso-oral" (ibid, p176).

acoustic difference between the sounds, speakers showing the near-mergers would be able to tell apart the two distinctive sounds. In other words, the paradox in fact will not exist in some properly designed experimental condition. If the speaker consistently makes the distinction between the two sounds in his production, he should be able to distinguish the phonetic differences between them. But in the actual communication situation where there are abundant linguistic and non-linguistic information available for semantic decision, the fine phonetic distinction is often ignored.

We find that the near-merger in a speaker only happens to a limited number of morphemes among the group of relevant morphemes. In our data, the largest degree of near-merger happens in a single speaker is 12 out of 17 target morphemes; 4 other morphemes in this speaker are consistent between production and questionnaire, and the remaining one morpheme has variation in production. This reflects another aspect of lexical diffusion at level of individual speakers: in an on-going change, the individual morphemes/words subject to the change exhibit different characteristics; the change does not affect the whole set of relevant morphemes simultaneously.

3.5 Individual differences and two hypothetical lexical learning styles

Speakers are tremendously diverse and varied in their language abilities and behaviors. Linguists and psychologists have shown individual differences in language acquisition in children (Shore, 1995), and in language use in adults (Fillmore et al., 1979). What we have shown in the previous sections has illustrated the considerable differences in speakers' linguistic behavior in the process of the on-going change between /n-/ and /l-/. In the following, we will summarize these individual differences and discuss some hypothetical lexical learning styles.

3.5.1 Different types of speakers

According to the possible combination of the presence of three states, we categorize the speakers into 7 types (T1-T7), as shown in Table 3.3. Types T2-T4 clearly exist in the current data. speakers Type T2-T4 may appear to have both /n-/ and /l-/ in their phonological system, but they differ a lot in the actual membership of the two phonemes. The existence of T1 and T5 in the current situation needs further confirmation. However, they should certainly exist at stages when a change begins or approaches completion. The last two types, T6 and T7, have not been attested. Type T6 reflects a situation that the variation applies to all the relevant words, exhibiting a kind of "free variation" in the speaker. Though our data from the formal speech style suggest that such speakers are unlikely, however, we do not exclude its possibility when data from other less formal speech styles are examined. Type T7 means that in a speaker's lexicon the relevant words split into U and C, but no word appear to be in the V state; this type is considered as impossible according to our *VT-III Hypothesis* stated earlier.

Table 3.3: Seven types of speakers categorized by different distributions of morphemes in the two phonological categories.

	U (/n-/)	V (/n-/ & /l-/)	C (/l-/)
T1	x		
T2	x	x	
T3	x	x	x
T4		x	x
T5			x
?T6		x	
*T7	x		x

Note: The symbol "x" indicates that there are morphemes manifesting that phoneme, "?" means uncertain and "" means logically impossible.

3.5.2 Explanations for individual differences

The individual differences may be attributed to one's specific linguistic environment. Each individual's linguistic environment and experiences are idiosyncratic, and therefore it is not surprising to find such heterogeneity across speakers. However, the linguistic environment may not be able to fully account for the individual differences. Language learning is a self-organizing process, dependent both on the linguistic input and the individual learning styles. Different learners taking in the same input may end up with different systems. To illustrate this point, we compare the situation of a pair of identical twins. They are two Hong Kong Cantonese male speakers (HK1 and HK2). They grew up together and went to the same schools until college. Therefore it is expected that their linguistic environment should be very similar during language acquisition period, and consequently that their linguistic behavior should be very similar as well. However, when we compare their pronunciations of the set of target morphemes based on data from the two production tests, we are surprised to observe a large amount of differences. Table 3.4 shows the distribution of morphemes in the 3 states in the two speakers and the number of shared morphemes with the same pronunciation. Among the 5 morphemes that HK1 consistently pronounces with /n-/, his brother only shares one and pronounces all the other four consistently with /l-/ instead.

Table 3.4: Comparison of the pronunciations of 50 morphemes of two identical twin Cantonese speakers (HK1 and HK2).

	HK1	HK2	number of morphemes shared
/n-/	5	10	1
/l-/	33	18	11
V	12	22	5

The data from the twins show how individuals can differ from each other in their linguistic development, even when they are exposed to very similar linguistic environment. Similar findings have been reported in longitudinal studies of twins' phonological development. For instance, Leonard et al. (1980) compare the development of word-initial consonants in the first 50 words of two female identical twins who were assumed to have been in the same linguistic environment. It is found that they differed as much as any two children who grow up separately.

Therefore, besides differences in linguistic input, there should be other factors accounting for the individual differences. Individuals' learning styles may be one of the determining factors. To identify possible stylistic differences, comparison between outliers and the majority may provide valuable information. In this study, the speakers without VT-III, i.e., speakers of types T1 and T5 shown in Table 3.3, are considered as extreme cases, opposed to the majority having VT-III, i.e., speakers of types T2 -T4 and T6. Accordingly we categorize the six types of speakers (T1-T6) further into two types: with and without VT-III. These two types will be discussed in detail shortly, under the terms of "probabilistic" and "categorical" learners respectively.

The above categorization is based on the observed differences in speakers' lexicon composition, which is at the behavioral level. There should be some differences in the underlying cognitive mechanisms related to language learning and processing between these two types of speakers. In the studies of individual differences in language acquisition, the various contrastive learning styles are mostly proposed at the behavioral level, such as referential versus expressive styles in vocabulary development (Nelson, 1973), cautious system-builder versus imitative learner in phonological acquisition (Ferguson, 1979), etc. Some linguists tend to emphasize that the stylistic differences in children's early language should be mostly ascribed to differences in the linguistic input and interactions, and suggest caution in linking the differences with learning capacities (Pine, 1994). However, when we encounter differences which cannot be explained purely by the environmental factor, such as the case of identical twins, we may have to seek explanations in general learning mechanisms which may be revealed by other learning processes such as music or chess. Some general dichotomies such as analytical versus holistic, risk-taking versus conservative, and social vs. object oriented (Shore, 1995) may be plausible candidates from this perspective.

3.5.3 Two hypothetical learning styles

In the discussions of learning styles, little attention has been paid to the situation when a language change is in progress, in which variants of the same linguistic item co-exist in the linguistic environment. It is true that even without language change, the language community is already heterogeneous. Previously attention has been paid to examining how children can extract the invariance among the diverse input to form consistent categories. However, on-going change providing such a complex situation for children to learn, the different learning styles may be revealed more readily. Considering the n-/l- case, for the same word, there are two distinctive forms existing in the linguistic environment. What do the children learn and use finally?

In addressing this issue, Shen (1997) has suggested two polar models of learners, i.e., lexical learner and phonemic learner. When a sound change is in progress, say from U to C, and the change involves a number of morphemes/words, the linguistic environment that a learner is exposed to is usually heterogeneous, as exemplified in the Teachers' pool shown in Table 3.5a. A lexical learner (Table 3.5b) would learn the words separately and maintain the two phonological categories in his lexicon. A phonemic learner (Table 3.5c) may be the most reasonable for some linguists who believe in language change as "phonemes change" (Bloomfield, 1933). For a phonemic learner, despite the existence of variation in his linguistic environment, he will only acquire one phoneme, either U or C. For a sound change to succeed, it is expected that more learners of the next generation learn C than U.

Table 3.5: Hypothetical types of lexical learning in the same linguistic environment: a) the linguistic environment; b) lexical learners; c) phonemic learners; d) probabilistic learners.

a. Teachers	W1	W2	W3	W4
Speaker-1	U	U	U	C
Speaker-2	U	C	C	U
Speaker-3	C	U	U	U
...				
b. Lexical Learners	W1	W2	W3	W4
Learner-1	U	U	C	U
Learner-2	U	U	U	U
Learner-3	C	C	C	C
...				
c. Phonemic Learners (categorical learners)	W1	W2	W3	W4
Learner-1	U	U	U	U
Learner-2	C	C	C	C
Learner-3	C	C	C	C
...				
d. Probabilistic Learners	W1	W2	W3	W4
Learner-1	U/C	C	U/C	U
Learner-2	U	C	U/C	U
Leaner-3	C	U/C	C	U/C
...				

Shen (1997) rejects the hypothesis that all learners are phonemic learners, because his data show that in two language communities, there are significant percentages of speakers (60% in Shanghai and 58% in Wenzhou) whose lexicons have the two merging phonological categories, instead of one. He further uses a probabilistic model to demonstrate that it is not possible that all learners are lexical learners either, as the distribution of the rates of change is significantly different from that in hypothetical model which assumes 100% of lexical learners in a community. He thus draws a conclusion that a realistic population may be a mixture of the two types of learners.

The two types of learners proposed by Shen only allow one form for each morpheme. However, as our study has shown, actually most speakers learn to use two forms for some words. Therefore we hypothesize another type of learner, called a "probabilistic learner" (Table 3.5d), who acquires two variant forms for some morphemes and only one form for other morphemes. In other words, these learners will turn out to have some morphemes in V and some others in C or U, corresponding to speakers of type T2-T4 in Table 3.3. The propensities of allowing the co-existence of variant forms across probabilistic learners must be gradient within a continuum. Some learners may acquire one form at a very small number of exposures and allow the co-existence of both forms more often than other learners.

While the proposed "probabilistic learner" accounts for speakers of types T2-T4 in Table 3.3, Shen's phonemic learner may account for speakers of types T1 and T5. We term here Shen's phonemic learner as "categorical learner" for a more general usage, since the categories may refer to units other than phonemes. It seems that Shen's lexical learner does not exist in our data, as no speaker has consistent /n-/ and /l-/ for separate sets of morphemes, unless they have morphemes in V.

It is hard to explain how a categorical learner is possible when the linguistic environment is filled with distinctive variant forms for each of the morphemes/words subject to the change. We suggest three possible explanations. One is that the speaker is specially situated in a homogeneous environment, which is only likely when the change just starts or nearly completes. The second is that the learner is especially insensitive to the presence of variants, and once he learns a certain form he will stick to it and disregard other forms; such a learner may have a rigid phonological system, which may be the consequence of the "system-builder" style found in some children's phonological development (Ferguson, 1979). The third source of categorical learner is suggested by some observation in our data, that is, the learner is able to generalize a rule for sound change from the heterogeneous input. In the current case of n-/l-, there could be some extreme speakers who are specially aware of the correspondence between Cantonese and PTH, and consciously transfer the /n-/ to their pronunciation of morphemes in Cantonese consistently in all speech styles. But this type of speakers is still highly susceptible, because for them to be able to generalize a rule, they must have already learned the two variants. They may appear as consistent and categorical in a certain speech style, for example in a highly formal style such as word-list reading, but the variation may appear in other speech styles. Speaker no. 4 in Figure 6 discussed earlier reflects such a case. This type of speaker in fact is not a "categorical learner" according to our definition[21].

3.6 Interpretation of the diachronic change: Dialectal influence or internal change

After we have analyzed the synchronic variation exhibited in the data, we are interested in seeing if we can interpret the diachronic profile of the change from the observed synchronic variation.

3.6.1 A uni-modal pattern

Some empirical studies have shown that language changes often proceed as an S-curve diffusion pattern, the current case study of n-/l- in Guangzhou Cantonese does not seem to be so. Take the percentage of /n-/ in the set of target morphemes in each speaker as the index for the linguistic variable of interest. We observe an unusual uni-modal pattern, as shown in the data from word-list reading (Figure 3.3) and from questionnaires (Figure 3.7): there is a peak of percentage of /n-/ in age group 3, while the older and younger groups have much less /n-/; furthermore, they are less by similar degrees. (The slightly higher percentage of /n-/ in group 6 than group 5

[21]We note that speakers may change their learning styles during the process of learning, or even after the acquisition period. We are thankful to Prof. Ron Chen who raised the discussion in this aspect. While our current fieldwork data are only from the synchronic dimension, it will be an interesting topic for future work to attest the existence of speakers undergoing changes of learning styles.

in Figure 3.3 is mainly due to the existence of a large number of near-merger cases in two speakers in group 6.)

The uni-modal pattern is intriguing, but not surprising, as it is a complex situation where two competing changes are in progress: the original /n-/→/l-/ and the recent reversal /l-/→/n-/. The second one is mainly due to the influence from Putonghua (PTH) on Cantonese, introduced by immigrants from PTH or northern dialect speaking regions, and by the official language policy of promoting PTH as the standard language. We speculate that the change would exhibit an S-curve if there were no reversal change. To test this hypothesis, we may compare the situation in Guangzhou Cantonese with other Cantonese varieties which have less or no influence of PTH. However, it is hardly possible to find such a candidate dialect, as the influence of PTH is overwhelming across the country. Hong Kong may be a possible one, as we discussed in Section 3.2.1.

3.6.2 Influence from PTH on Cantonese

We have discussed briefly the influence of PTH on Guangzhou Cantonese in Section 3.2.1. It is mainly through two channels, cultural diffusion (by education policy) and demographic diffusion (through immigrants). Here we will elaborate what have happened in these two channels.

As mentioned earlier, there have been large numbers of immigrants to Guangzhou since early 1950s. In the past, the prestige of Cantonese in the city of Guangzhou was high, and speaking PTH was considered a stigma for northern immigrants and they would probably be treated differently from local people in various public places. Therefore many immigrants have tried to learn and communicate with local people with a self-taught Cantonese. One direct consequence is that many morphemes, which are pronounced with /l-/ by native Cantonese speakers, have been likely to be pronounced with /n-/ initials in these PTH speakers due to self-teaching, because the corresponding of pronunciations of these morphemes are with /n-/ in PTH. Subsequently, Cantonese learners who have been exposed to the influence of these immigrants, especially the children of these immigrants, would have a high probability to learn some morphemes with /n-/ initials from them. We compare the percentages of /n-/ in speakers from native and non-native Cantonese families, and have found that there is a significant difference between the two groups of speakers ($t=-2.9$, $p<0.005$): there are more /n-/ in speakers from non-local families.

Table 3.6: Comparison of the percentages of /n-/ of speakers from local and non-local families

	from local family	from non-local family
number of speakers	33	9
mean	11%	29%
variance	211	292

The usage of PTH in local communities also affects Cantonese speakers. We know that the acquisition of /n-/ is universally earlier than /l-/ from studies on children's phonological development, such as the recent studies in PTH (Zhu and Dodd, 2000) and Cantonese acquisition (Wong and Stokes, 2001). The acquisition order of the two sounds suggests that if the linguistic environment provides /n-/, the child should be able to learn the sound /n-/. In other words, if we do not observe /n-/ words from the speaker, we can assume that during the time of phonological development, there is no /n-/ available in his linguistic environment. This may account for the fact that those speakers whose parents are not native Cantonese speakers have much less /n-/s than those from local families, because the former had less chance to be exposed to /n-/ stimuli.

However, we cannot make the opposite inference. If the speaker is able to produce words with /n-/ initials now, it is not necessarily true that he has learned /n-/ from his Cantonese linguistic environment during his childhood, because /n-/ can be learned after childhood. If a child learns to speak PTH at school, he will be able to maintain the perception of the distinction between /n-/ and /l-/. This will allow him to learn pronunciations with /n-/ for Cantonese words in adulthood, if the speakers recognize the existence of such pronunciations.

Moreover, some speakers may be aware of the correspondence between PTH and Cantonese, and thus consciously transfer PTH's /n-/ to Cantonese pronunciation.

However, individuals vary a lot in their degrees of transference. Some speakers, especially those with less education and whose PTH is less proficient, show no sign of such transference. Some speakers apply the transference only to some lexical items, instead of applying to all relevant morphemes or words. Each speaker has his own set of affected morphemes. We can see this from the fact that the percentages of /n-/ vary in different speakers, and most of them are much smaller than 100%. Were it implemented as a phonemic change, then we would have expected a complete change to 100% of /n-/ in all the speakers who applied the transference. Again the actual data demonstrate that there is no phonemic change, but rather the change is implemented through lexical diffusion. In fact there may be no dynamic diffusion process in the individual speakers. Most of the speakers may stabilize their pronunciation for most of the words around puberty. Only in some special cases, the speakers apply transference consistently on all relevant morphemes, as exemplified by the speaker no. 4 in Figure 3.6 who has /n-/ for all words in both production tests. But even for this speaker, the transference is only consistent in formal speech, such as the word-list reading, and her casual speech reveals the existence of /l-/ for some morphemes.

Words have different outcomes in such a complex situation. For those words and morphemes which are used in daily communication, it is likely that the speakers will learn and use what they were exposed to in their learning environment, be it /n-/ or /l-/; for those infrequent morphemes and words, if they are not sure of the pronunciations due to lack of exposure, they will determine the pronunciation by analogy with the phonetic component of the written character, or by transferring the pronunciation of PTH to Cantonese. For example, about 14% and 7% of the speakers, who have mostly /l-/ initials, pronounce some infrequent morphemes like '溺 ' and '匿' with /j-/ respectively, and 14% of the speakers pronounce '朽' with /x-/, as a result of analogy from the phonetic component in the character (Hashimoto, 1972). Meanwhile, some speakers, who have mostly /l-/ initials, pronounce '腻', '鸟' and '嫩' with /n-/, which is very likely to be the result of the transference from PTH. Due to these different individual learning strategies, several splits of the original set of morphemes will occur in individual speakers and in the speech community: while some morphemes remain unchanged with /n-/, some others will acquire new pronunciations with /l-/, /j-/ or /x-/. The phenomena we observe from this study may give us some hints on how historical changes of split would have started and progressed.

3.6.3 Explanation for the uni-modal pattern

Now we will provide some tentative explanation for the observed uni-modal pattern. First, the average percentage of /n-/ seems relatively constant across the three older groups (around 24% when only counting morphemes with consistent /n-/ in one production test, or around 44% when taking into account both /n-/ and V morphemes); the difference between the three older groups is not statistically significant (ANOVA, between groups, $p=0.50$). This suggests that the /n-/→/l-/ change had not progressed much during the period of 1940-70, or the reversal change has already played a role. However, we do not know how many /n-/s attested in the speakers of the older groups are the residues from the original change of /n-/→/l-/, and how many have re-appeared due to PTH's influence. A comparison with Hong Kong Cantonese of the same age groups may provide some possible answers.

Second, there is a sharp increase in the percentage of /n-/ in the 3rd age group. The difference between group 4 and 3 is statistically significant ($t=-1.9$, $p=0.058$); so is the difference between groups 3 and 2 ($t=3.12$, $p=0.004$). The significantly larger percentage of /n-/ should be partly due to the increasing impact of PTH's use. As we mentioned earlier, a major promotional campaign of PTH was launched in 1955, but the enthusiasm ebbed considerably after 1959 (Chen, 1999b, p26), and there was a lack of continuous effort in promoting PTH during the tumultuous period of the Cultural Revolution (1966-76). Later in the late 70s and early 80s there was a second promotion campaign of PTH, though with less enthusiasm and energy than in the mid-50s. A resurgence of PTH appeared. Speakers in the age group 3 may have been the group that was the most influenced by this second period of promotion from their schooling at that time. A cursory observation to support this argument is that the competence of PTH of the 3rd group as a whole is much better than the three older groups. Due to the PTH's influence, there were more possibilities for individual speakers to transfer PTH's /n-/ to Cantonese.

But we have difficulty in explaining why the degrees of /n-/ drop significantly again in the two younger groups. Conceptually it could be explained as an inherent drift (Sapir, 1921) in Cantonese bringing /n-/ to /l-/ and that this drift tendency is still at work. The peak of reversal change in the early 1980s was just a perturbation of this inherent tendency, and after the perturbation the original drift continued. However, this is an abstract interpretation and we still need to figure out how the drift is implemented if it is true. There are several speculations. One is that the parents of the two younger age groups are mostly in the 4th age group. If we assume that parents are more influential in children's language learning, then it seems reasonable to predict that the speakers in the two younger age groups have small percentages of /n-/ similar to those in the 4th group, as we observe in the data. However, this argument seems to be inconsistent with what early sociolinguistic studies have shown: language learning is more affected by siblings and friends rather than parents (Weinreich et al., 1968). However, it is still worthwhile to compare data of pairs of parents and children to see if the proposed argument is reasonable in this case, because we have shown earlier that speakers from non-local family are significantly different from those from local families. Another explanation is to consider the decrease of /n-/ in the two young age groups as a kind of "age grading". That is, most of the youngsters do not have the same degree of awareness of PTH's relationship with Cantonese as the elders and thus they have less transference from PTH to Cantonese. We group the 50 words into two categories, i.e., frequent and infrequent, according to their degree of literacy which is roughly decided by our own judgment. It appears that age group 3 and group 2 are significantly different in the pronunciation of literary words (t=2.69, p=0.006) but not significantly different regarding colloquial words (t=0.48, p=0.32). It suggests that the fewer /n-/ in speakers in the 2nd group may be partly due to their less transference in the literary words than the speakers in the 2nd group. But the difference between group 3 and group 1 is not significant, which may require other explanations.

An alternative explanation of the uni-modal pattern is not related to the diachronic change, but rather related to the sampling problem of the data. The uni-modal pattern may reflect a social stratification characteristic in terms of educational level. We notice that the speakers in age group 3 on average have the highest educational level than other groups. In this study, the speakers are categorized into 6 educational levels, graded as graduates from (1) primary schools, (2) high schools, (3) technical schools, (4) local colleges, (5) universities, and (6) graduate schools, respectively. The average levels of the six age groups are respectively 3.75, 3.11, 4.5, 5.0, 2.1, and 1.7. If we further group educational levels 1 to 4 as low education (LE) and 5-6 as high education (HE), there is a significant difference between LE and HE groups (t=3.1, p=0.005).

Table 3.7: Comparison of the percentages of /n-/ of speakers with high and low level of education.

	High education (HE)	Low education (LE)
mean	30%	10%
variance	380	156
number of speakers	10	32

This suggests that the high educational level in the speakers of age group 3 may partially account for the high percentage of /n-/ compared to the other groups. One piece of background information concerning the education situation may further explain the distinction between LE and HE in Guangzhou. There are many non-local students in universities and the usage of PTH is much more common than other schools. In primary schools, high schools and technical colleges, local students are the majority, and their daily communication is still mostly in Cantonese. Therefore the educational level at university may create a division on the degree of exposure to PTH.

As we have recognized that the specially high percentage of /n-/ in group 3 is at least partly due to the special composition of the group, we will be prudent in asserting the credibility of the observed uni-modal pattern. It is possible that the percentage of /n-/ would drop to a similar level as other groups, when more speakers with lower education background are included. However, as the current peak of /n-/

in age group 3 is so remarkably higher than other groups, it is hard to conceive that the uni-modal pattern would totally disappear. It would be worthwhile to test and validate the above various explanations in further studies.

3.6.4 Differences in actuation: Dialectal contact or internal change

While we hypothesize that the reversal change is due to the PTH influence on Cantonese, we do not categorize the change exclusively as a case of borrowing or dialectal contact. In fact it is very difficult to separate clearly language contact from internal change. The distinction between the two types of change mostly lies in the actuation. Once the change due to contact has started in some bilingual speakers, the following stages will share the same diffusion process. As Wang (1979/1991, p79) remarks, "Although it is obviously important to know whether a change is actuated internally or externally", the implementation by "a process of lexical diffusion should be the same". The process includes not only diffusion across words in the lexicon, but also across speakers in the language community, both through the learning of the new generations.

Considering the reversal change under discussion: not all speakers are subject to the influence of PTH directly. Only a few speakers, who are fluent bilinguals and conscious of the correspondence between PTH and Cantonese, are affected directly by the contact between the two dialects within their idiolects. The /n-/s in some morphemes produced by these speakers serve as input for the learners in later generations. The learning process in these learners would be the same as learning other linguistic items.

Moreover, we have to be cautious when we speculate that a speaker is subject to the PTH's influence. In the discussion of the problem of attributing sound change to language contact, Cheng and Wang (1972) point out that "borrowing is not phonetically selective, i.e., it is highly implausible that the borrowing dialect will accept from the lending dialect only those morphemes which share a given phonetic or phonological characteristic. Since borrowing is motivated by lexical need or cultural imitation, it is unthinkable that the borrowed items are all and only those which satisfy a given phonetic condition" (reprinted in Wang, 1977, p91). This discussion is about languages/dialects at the population level. But we can generalize a similar argument on individual speakers: if we assume that a certain speaker has an influence from PTH on his Cantonese, we should expect to see other evidence of such influence, be it in the phonological and/or grammatical system, as the influence should not be restricted to the n-/l- case only. Further experiment to examine more linguistic variables in the few speakers with mostly /n-/ pronunciations may allow us to test this hypothesis.

3.6.5 More on *coexistence of variants*

The current study of an on-going change highlights the phenomena of co-existence of variants both in the speech community and in the individual speakers. Such observation is not surprising at all, as discussed in the beginning of this chapter: the whole enterprise of sociolinguistics since early 1960s is to examine and explain the existence of such variations. However, most of the time, such existence of variation is in the middle of a diachronic change, reflecting the state of an intermediate stage. Will such variants coexist for a long time, or in other words, will coexistence of variants be stable in a language? The answer seems to be no. Kroch (1994, p185) remarks that "due to their sociolinguistic origins, the two forms often appear in different registers, styles, or social dialects; but they can only coexist stably in the speech community if they differentiate in meaning, thereby ceasing to be doublets".

The process of how doublets arise and disappear can be illustrated by a study on the past tense development in English. Taylor (1990) has found that the appearance of doublets peaked in the thirteenth to fifteenth centuries which is supposed to be due to the large immigration of Scandinavian speakers from the southern areas into the north and northeast of England. Adult Scandinavians as second language learner of English simplified inflections. For example, they generalize "-ed" past tense forms, most often starting from less frequent verbs. For a period of time, children in subsequent generations may learn from an environment with both forms, and develop a diglossic capacity. In general, doublets do not last long. According to Kroch (1994), the average life span of doublets, measured by citations in the *Oxford English*

Dictionary, is about 300 years, though Lightfoot (1999, p99) considers this figure "misleadingly high"[22].

There were competing past tense forms. In some cases the innovative weak form has come down to present-day English; in other cases it is the old strong form which has survived. In no case has a Middle English doublet survived to the present day, except in one context where the competing forms take on different functions. For example, "fitted" and "fit" coexist in some dialects, as the former is causative and the latter is not, e.g., "She fitted him with a new suit" versus "That collar fit him last year" (Lightfoot, 1999, p100). We may find some recent examples such as "mice" for plural form of the animal "mouse", while "mouses" for the computer accessory.

Why can't such doublets co-exist for a long time? Are there any inherent cognitive constraints in the language user to prohibit the coexistence of variants? The "Blocking Effect" proposed by Aronoff (1976) from the viewpoint of language acquisition may be one of the candidates. He argues that there is an economy restriction on lexical items such that morphological doublets do not generally exist. If a child knows "went", he won't use "goed". Blocking effect seems to preclude the coexistence of variants in individual language users.

The economy explanation by Aronoff does not sound convincing. Two other hypotheses in acquisition studies may provide a better account for similar observed phenomena. They are the "principle of contrast" proposed by Clark (1987), and "mutual exclusivity assumption" by Markman (1989). These hypotheses suggest that children may be born with such constraint of avoiding different names associated with the same object, though it is unknown yet whether this constraint is specific to language.

Croft suggests the impossibility of coexistence of variants from another perspective, that is the nature of language as "convention". In his theory of language change, he proposes a *First law of propagation*: "there appears to be a natural human tendency

[22]The duration of the co-existence of morphological doublets may be much shorter than co-existence of synonyms. From Wang (1966)'s study on lexical replacements in the Donghan-Sui period of ancient Chinese, we find that many basic words, such as "eye", "tree", "sweet", have co-existing synonyms for a long period of time, usually over 700 years, though the shift of dominance may complete less than 200 years, such as the word '住' for "live" took over the place of '居' within the period between late Han (around 200 CE) and Dong Jin (around 400 CE).

for a community to select one alternative as the conventional signal for a recurrent coordination problem", or " there appears to be a natural human tendency to increase the conventionality of one variant of a lingueme in a community at the expense of another, albeit over a long period of time in many cases" (Croft, 2000, p176).

However, it is not clear yet to what extent these constraints apply in different individuals, and to different language components. If the effect is working categorically, then we should expect no coexistence of variants in individual speakers at all, even in the middle of on-going changes. However, the co-existence of variants is very abundant, as shown in many synchronic variation in sociolinguistic studies. In our current study we observe a significant degree of variation, at least within some individual speakers and in some relevant part of their lexicon. It seems that the blocking effect does not apply uniformly to all speakers, if it is true at all for some speakers. As we have discussed above, there may exist a large degree of individual differences.

Chapter 4

A review of computational modeling in language evolution

It is known that "a basic task of science is to build models – simplified and abstracted descriptions - of natural systems" (Belew et al., 1996, p432). In recent decades, with the growing availability of increasingly powerful computers, scientists in various disciplines have adopted computational models to study natural phenomena. In the study of language evolution, computational modeling is a burgeoning area, and a rapidly increasing literature has accumulated. As a reflection of the development of the field, a large proportion of work collected in the series of recent anthologies on the study of language evolution relies on computational models (such as Briscoe, 2002; Cangelosi and Parisi, 2001; Christiansen and Kirby, 2003; Hurford et al., 1998b; Knight et al., 2000; Wray, 2002). Several comprehensive reviews of the state of the art of this area can be found in Parisi (1997), Kirby (2002b) and Wagner et al. (2003).

In this chapter, instead of giving a comprehensive review of this large and diverse literature, I will focus on the several divisions which are useful for an overview of the field, in particular the three time scales of evolution and four levels of resolutions that models adopt. Then I will introduce the agent-based modeling approach which suits the theoretical framework of self-organization in studying language evolution. The relation between production and perception/comprehension is discussed as a special consideration for the modeling study. At the end of the chapter, I will summarize the pros and cons of computational modeling, highlighting the promising future of this approach.

4.1 Modeling at different time scales

Wang (1978) notes that in the study of language change it is necessary to distinguish different time scales, "for it seems that the questions, the data and the methods would not be the same for all these time scales". He suggests three different time scales, that is, microhistory, mesohistory and macrohistory. Microhistory of language is reckoned across a very thin slice of time, in years or decades, and mainly deals with "changes in progress" (Labov, 1972). Synchronic variation dependent on social factors and language learning in early years are the main subjects of studies in this time scale. Mesohistory is concerned with changes that occur across centuries or millennia, which constitute the great bulk of the literature of language change. Macrohistory of language addresses the largest time perspective which deals with the question how language emerged during human evolution. This last area poses the greatest challenges, as the relevant primary data are mostly not available.

The distinctions between these three time scales can be rephrased as another set of ternary distinctions from a different perspective, that is, language acquisition, change and origin. In addition to the differences in the time scale, studies in these three areas also differ in the questions and objects they focus on. Language acquisition, sometimes called "ontogenetic emergence", focuses on the process of the formation of idiolects in children, while language change and emergence both concern the communal language but at different stages. Language change deals with the processes of evolution of modern languages, from some form to another, both of them being fully functional. In comparison, the study of the origin of language, sometime referred as "phylogenetic emergence", investigates how a full human language, featured by compositionality and hierarchical and recursive structures, could have evolved from a rudimentary communicative system, i.e., a pre-language. Studies on language change deal with the stable states in the fitness landscape of evolution whose dynamics are mostly smooth and gradual[1], while language origin concerns the emergent states (Wang, 1978) which may go through some sharp phase transitions.

Having recognized the distinctions, however, it is important as well to note that the study of these three areas are intimately related to each other. Language change and origin are indispensable from language acquisition, because the former two are impossible to understand without studying language acquisition. Language change progresses mainly through language acquisition; without transmission by new generations, a language will die out, and there will be no more language change concerned. The study of language origin need the knowledge contributed by language acquisition, because the initial condition found in children to learn language will provide a target list of innate endowments for the study of language origin to explain how human beings have evolved to have language.

Meanwhile, the study of change in modern languages should provide hints to hypothesize what the phylogenetic process of language origin could have been in the past. The various factors identified in the study of language change, such as social factors, the heterogeneity in the communal language, the transmission across generations and within generations, should play similar roles in the remote past. The uniformitarianism principle is assumed in this thesis. Language origin and change should both proceed in a self-organizing way through interaction and acquisition in the language community, though the two processes have different dynamics due to their different initial conditions.

Empirical studies in these three areas have developed independently largely, each with its own focus and methodologies. Language acquisition and language change typically have grown into two large distinct areas in linguistics. Occasional links between language change and language acquisition are explained mainly by hypothetical inference from a functional perspective (e.g., Slobin, 1977)[2]. The process of change implemented by acquisition was seldom worked out. One of the few examples in empirical studies is Lightfoot (1999) which explains the change of word order in English

[1]This is not to deny that there are abrupt changes in the evolution of modern human languages. Language split and development of creoles should be evidence of such abrupt changes.

[2]The plausibility of some of these hypotheses has been challenged. For example, Kiparsky (1968) proposed that the source of double negation in English was due to children producing double negatives. But this view has been discredited (Croft, 2000).

as the result of language acquisition in a language contact situation.

Computer modeling has been widely employed as a powerful tool to examine various concepts and theories proposed in the field. Different models applying different representations and architecture have been developed to address various questions. One noticeable feature from modeling is that the connections between the three distinctive areas are made necessary. To model language change and language origin, language acquisition has to be incorporated, either in some abstract integrated form or in some detailed implementation. In the following we will give a brief review of the modeling approaches in these three areas.

4.1.1 Language acquisition

It has been well accepted that the linguistic environment plays a crucial role in children's learning language. However, it is very difficult to keep track of the linguistic input to children, and cross-individual comparisons are even much harder. Researchers used to construct their theory based on a small amount of data from one or two children. In recent decades, the methodology and facilities have been improved and the corpora from fieldwork are continuously expanding. Some large databases, such as the noted CHILDES project in Carnegie Mellon University (MacWhinney, 1995; Sokolov and Snow, 1994) have been made available for public use. However, the data are still far from sufficient, especially for the study of inter-individual and intra-individual variance for language learning. Moreover, unlike experiments with animals and plants, we cannot conduct controlled experiments on children, such as changing their learning environment. To address these problems, computational modeling provides an effective methodology, which allow to use artificially generated data and run control experiments.

In fact, computational modeling has received attention in studying language acquisition much earlier than language change and language origin. Since the revitalized development of connectionism in the 1980s (Rumelhart et al., 1986), scholars have simulated various problems of language learning with different kinds of artificial neural networks. For example, Rumelhart and McClelland (1986) report a model which is one of the first computational modeling works on the acquisition of morphology, to address the debate whether there are two routes of learning, one rule-based and one memorization-based. Their model uses a single layer network trained by the perceptron learning algorithm to learn English past tense formation. The network takes input of root forms of English verbs represented by 460-digit "wickelfeature" which encodes triples of consecutive phonemes in a verb, and output the past tense form. The network was claimed to demonstrate the U-shaped pattern found in empirical studies: initially a small amount of regular and irregular words are used correctly, but later the irregular words are wrongly regularized for a period of time, before finally being used correctly again.

However, this model has received criticisms such as its wickelfeature representation of the inputs; the sudden increase of training data to artificially produce a U-shape learning pattern (Pinker and Prince, 1988), etc. Triggered by this study, a large number of studies have been reported both from both connectionists' and the nativists' perspective. The nativists insist the existence of the word-and-rule dual-mechanism (c.f. a recent view Pinker and Ullman, 2002), by the evidence from native speakers[3], from other languages such as German and Hebrew, from brain-damaged patients who show two different types of pathology, and from neurocognitive studies on memories (Ullman, 2001). At the same time, connectionist have been improving their models, aiming to enable the models to produce behaviors closer to what have been observed empirically. More complex networks have been proposed, such as multi-layer feed-forward networks (Plunkett and Marchman, 1991), modular networks (Pulvermüller, 1998), etc. The representation of the input has also been changed to become more realistic than the highly specified wickelfeatures. The training procedures have been changed from two abrupt phases of input to a gradual increase of the training data, and also take into account the frequency effect, and so on (see reviews Elman et al. 1998, Christiansen and Chater 1999).

We note that on the other side of the debate, computer models have also been

[3]Native English speaker show apparent knowledge of the rule and can apply it to unusual novel words such as "oink", and treat different compound words such as "over-eat" and "fly out" in two different ways (i.e., "over ate" as opposed to "flied out").

utilized for the rule-based framework as well. Brent (1996) reviews some computational models for language acquisition, including learning meanings of words using semantic bootstrapping (Siskind, 1996), segmentation of utterances using distributional regularities and phonotactic constraints (Brent and Cartwright, 1996), and a few others. A recent work by Yang (2003) presents a new proposal suggesting that the irregular patterns are also rule-based, and that the child's task is not to memorize on a word-by-word basis, but to figure out which rule applies to which word. The simulation based on his variational learning model is reported to make predictions highly consistent with the corpus data.

The studies on English past tense have triggered a rapid growth of applications of connectionist models in simulating various aspects of language acquisition. Some examples are: learning the distinction between grammatical categories (Elman, 1990; Li, 1999), segmentation with multiple prosodic cues (Christiansen et al., 1998), or simulating infants' detection of syllabic patterns such as AAB as in "wi wi ki" (McClelland and Plaut, 1999), and so on. Most of these studies attempt to demonstrate that many symbolic rules assumed by linguists may not be present explicitly in language processing, but instead exist as distributed representation and associations in neural networks. Moreover, the linguistic data contain abundant statistical information, from which children can make generalizations after learning enough items. The linguistic regularities, which appear very much like rules, such as the past tense formation rules, may in fact be the recurrent distributed activation patterns in the neural networks. This process may be governed by some general learning mechanisms, as represented by neural networks which do not have any linguistic biased construction built-in. Different types of neural networks have been proposed after the early multi-layer feedforward network, including Elman's celebrated recurrent neural network (Elman, 1990) which can incorporate temporal information into the network, and the self-organization map (Kohonen, 1995) which uses unsupervised learning to avoid the unrealistic backpropagation training in earlier networks.

As the models have to make their assumptions and implementations very explicit, it is much easier to challenge and falsify them, such as what has been shown in the example of past tense learning. The challenge and criticisms often provide the modelers with clearer directions to modify and improve the implementation in their models. Through such spiral turns of criticism and improvement, models tend to become more realistic and at the same time more helpful in identifying the crucial factors in learning. The following example will illustrate such a point.

In an empirical experiment reported by Marcus and colleagues (Marcus et al., 1999), infants are exposed to several examples of sequences of three syllables conforming to a simple general pattern, for example ABA sequences such as "ga ti ga" and "li na li". In the subsequent test phase, it is found that infants listened less to test sequences that obey the pattern to which they had been previously exposed. On the contrary, they pay more attention to those sequences which constitute novel syllables, has not presented in the training phase but with the same sequential ABA pattern (such as sequences with entirely new syllables, such as "wo fe wo"). The authors argue that the experiment results suggest that infants extract abstract algebra-like rules that represent relationships between placeholders (variables), such as 'the first item X is the same as the third item Y,' or more generally, such as 'item I is the same as item J' ". They consider that simple statistical learning cannot account for the results.

This rule-like proposal has been challenged. McClelland and Plaut (1999) argue that "it seems very possible to us that seven-month-old infants possess mechanisms that provide powerful support for generalization. But we don't really see how experiments of this general sort can tell us whether they use rules *per se*; the powerful mechanisms might simply be ones that help statistical learning procedures generalize in powerful ways. Furthermore, these mechanisms might themselves be learned" (p166).

Various connectionist models have been used to simulate the rule-like behavior, but reached different conclusions. While Marcus et al. (1999) themselves show that the popular recurrent neural network cannot succeed in producing the result found in empirical studies, Seidenberg and Elman (1999) propose that a simple recurrent neural network can replicate children's discrimination behavior successfully. In their network, the input is a string which encodes the pattern whether two adjacent syllables are the same or not. The pattern is assumed to be extracted by an external feature detector. Marcus challenges this model by arguing that Seidenberg and Elman's model in fact has some hidden rules, which govern the external feature detector. However, Seidenberg & Elman suggest that the detection of the similarity of two subsequent syllables, which concerns only some local regularities, is a more basic capacity in infants (Jusczyk, 1997). If Seidenberg and Elman's model works, then at least the rule hypothesized by Marcus et al. would seem unnecessary. The algebraic rule can be explained as the network's statistical generalization from local regularities.

The above studies and arguments illustrate a point which has been highlighted in the self-organization perspective of language evolution: it is very likely that the global pattern is an emergent property. We need to seek explanations from lower levels. In the case of the rule behavior, it is important to know what kind of features that infants encode the input data with and utilize for generalization.

4.1.2 Language change

In studies of language change, two types of data have accumulated abundantly in empirical studies - synchronic on-going changes by sociolinguists, and diachronic changes by historical linguists. On the one hand, the sociolinguists have carried out in-depth microscopic analyses for the distribution of linguistic variants in language communities. It is found that the variation is determined by both language-internal factors (such as the functional criteria including production cost, perceptual distinctiveness, etc. (Labov, 1994)) and language-external factors (such as social network, social identity, social class, gender, age, etc. (Labov, 2001)). However, the various mechanisms that sociolinguists have identified for language change are concluded from data over a short period of time and it is not clear how to apply these arguments from synchronic studies to historical data.

On the other hand, historical linguists' analyses of language change usually compare languages at distinct points in history only. Very often there is no elaboration on what are the intermediate stages for the change, and how the synchronic variations can affect long term changes.

Computational models are especially useful for building up links between these two areas, by incorporating and manipulating various parameters in a controllable manner, and by simulating a continuous process along time. By taking into account synchronic factors explicitly, the implementation of change in history can be examined and analyzed quantitatively. Theoretical issues are addressed with the benefit of explicitness and falsifiability from computational modeling. The modeling approach has become more and more used in explaining historical changes in recent years (see a recent brief review in Science (Bhattacharjee, 2003).

For example, Nettle (1999b,c) uses computational models to discuss one of the main theoretical problems in language change, that is the "threshold problem": why a new linguistic variant initially being rare can win over the previous linguistic norm. In biological and cultural evolution, when a new mutant trait arises in an individual, it has a good chance to be passed on to that individual's offspring, as long as the mutant is not severely deleterious or actually lethal. Some genetic mutations can successfully diffuse into the whole population without natural selection (Kimura, 1983). But linguistic transmission is different from genetic transmission. Instead of inheriting genes from *one or two parents*, a language learner samples at least a proportion of the language community, which may include *a fairly large number of people* in the generations above him as well as in his peer group. The mutant which has arisen in the last generation will be in the minority for the next generation to learn. "All plausible learning algorithms lead, other things being equal, to the adoption of the most common variant in the sample for a given item, which will never be the new mutant" (Nettle, 1999c, p98). Therefore, new mutants cannot become the fixed norm in a language community unless "they can pass a threshold of frequency which in the early stages they never have" (ibid). However, if the learner is at least sometimes biased toward the new variants for some reason or another, then the variants in the minority would have a chance of overcoming the threshold of rarity.

Nettle (1999a,b) suggests that there are two possibilities for the innovation to overcome the threshold. One is functional selection, i.e., there is a functional bias toward the innovation over the original norm. Studies on language universals and language evolution have proposed various functional accounts, such as perceptual salience, production economy, markedness, iconicity, etc. (Croft, 1990; Kirby, 1999a) (see also Section 1.1.2.1). The other possibility to cross the threshold for change is "social selection", in which the innovation originates from some influential speakers who have higher influence, or "social impact", than others and learners may favor learning from them.

Nettle's model is an adapted version of Social Impact Theory in simulating attitude change in social groups (Nowak et al., 1990). The population is structured with age and social status. The language learner chooses one of the competiting linguistic variants by evaluating their impact after sampling the speech of individuals in the community. Individuals within shorter social distance or with higher social status have a higher impact on the learner.

Nettle's simulation models demonstrate that in a community homogeneous in social status, the functional bias need to be unrealistically high for the innovation to spread successfully; but with social selection, for a population with large differences in social status, an innovation with a very small functional advantage is very likely to spread. Concluding from these results from the simulation, Nettle (1999c) suggests that functional biases may affect the direction of language change, but cannot provide a sufficient condition for change to happen. Meanwhile, many variants causing language change may not have any functional advantage. He remarks that "without the potential for change provided by differences in social influence, functionally favored variants might never overcome the threshold required to displace prior norms" (ibid, p116).

Nettle further studies the effect of population size on the rate of change with his simulation models (Nettle, 1999a), and shows that the rate of change decreases as the community size increases, i.e., larger population size results in a slower change rate, and vice versa. Meanwhile, if languages change fast internally (i.e., the change is not due to language contact), then when a group of languages split from a common ancestor, identifiable relationships between the descendants will quickly reduce, and so after a given period of time, there will appear to be many language families[4], each

[4]Here the usage of "language family" is not strictly consistent with the conventional meaning in historical linguistics, but roughly refers to groups of the extant languages. Nettle uses "stock" following Nichols (1992)'s terminology, which refers to the deepest phylogenetic grouping of languages which can be identified using the traditional comparative method.

with small numbers of members. On the contrary, if languages change slowly, then relationship between languages which split from the same origin will be maintained longer with more identifiable traces, and therefore the reconstructable language families will be broader and include more languages. Therefore, based on the result from simulation and the above reasoning, Nettle predicts that a region with larger language communities would should have a smaller linguistic diversity, which is measured by how many language families there are in that region. Nettle finds that this prediction is consistent with his analysis of the linguistic diversity of three main regions in the world: in the Old World, large-sized language communities correspond to smaller language stocks, each with a large number of languages, while in the New World, language communities are often of small size, and there is a large number of language stocks, each with a small number of languages. And the Australia/Pacific region is in the middle between the Old World and the New World regions.

There are two problems in Nettle's analysis of group size and linguistic diversity. First the determination of language stocks, groups or families is very controversial. For example, Nichols believes that in the New World, there are 157 language stocks, while Greenberg (1987) only identifies three groups. Second, the group size of the extant languages may not reflect the situation when they split from their ancestral languages several thousands years ago. In fact language communities may grow and shrink in different historical periods, affected by demographic change as well as political influence. For example, English has increased its speakers dramatically from only a small population, since the colonial expansion of the British around the world. The national language in China, Putonghua, has been expanding its coverage in the country at the cost of many local dialects in recent decades. Nettle has been aware of this problem, but he assumes that "the figures are probably of the right order of magnitude". A more careful examination of the relation among group size, change rate and linguistic diversity should be necessary for both empirical and computational modeling studies. In this thesis, we examine the relation between group size and change rate with similar computational models by taking into more realistic social structures, and our results do not support Nettle's hypothesis. These models will be reported in Chapter 5.

While Nettle's discussion on linguistic diversity is from the language change perspective, there have been a number of computational models which study language diversity from the emergence and convergence point of view (e.g., Arita and Koyama, 1998; Livingstone, 2001; Livingstone and Fyfe, 1999). They examine when a group or groups of people converge to the same language, or converge to different languages, which means diversity arises, under different conditions of neighborhood size between groups, the level of noise in the communication, etc. Livingstone (2000) challenges Nettle's arguments against the neutral theory of language evolution. He uses simulation models to show that no functional or adaptive benefits are required to create linguistic diversity as Nettle suggests, and that diversity could arise naturally from the imperfect transmission of language from users to learners. Livingstone's model specifically address language diversity by comparing language among different communities in a continuous , address continuous linguistic domain and it needs more examination of the

Similar to Nettle, Partha Niyogi and colleagues are interested in studying language change as the result of vertical transmission through language acquisition; but they adopt a formal mathematical modeling approach rather than simulation models (Niyogi, 2002; Niyogi and Berwick, 1997). To achieve analytical results, their models deal with discrete and non-overlapping generations based on the cultural transmission model proposed by Cavalli-Sforza and Feldman (1981). The learning of a grammatical system is abstracted as a Triggering Learning Algorithm (TLA) within the Principle and Parameter framework (Haegeman, 1991). It is assumed that different grammars are already present in the population, and that some sentences are parsable by more than one grammar. They examine the dynamics of these grammars in the population under different parameters, such as the probability of producing "ambiguous"[5] sentences using particular grammars, the learning duration (the number of sentences a child would be exposed to before maturation), etc. They apply this

[5]Though these sentences are called "ambiguous" sentences in Niyogi's model (Niyogi, 2002, p215), they are not actually ambiguous in the sense of one-form-multiple-meaning mapping. These sentences are rather parsable by, or compatible with, more than one grammar.

model to address the historical changes from Old English to Modern English. Old English is largely head-final and +V2 (the finite verb is moved to the second position, such as in modern German), while Modern English is head-first and -V2. These models have produced some analytical predictions, for example, the percentage of +V2 grammars increases or decreases in the population over time under different conditions. However, there have been various idealizations assumed in the model, such as non-overlapping populations, the homogeneity of children's learning, including equal exposure to the linguistic environment in the whole population, and learning grammar using a TLA. It is not clear yet how changes in these assumptions would affect such analytical models.

Nettle and Niyogi's studies share the same underlying assumption that language change is the consequence of language learning. Nettle's model only considers the case of an abstract linguistic innovation and Niyogi's model represents language as a set of parameters. Connectionist models discussed in the last section which represent language as distributed associations have also been applied to the study of language change, similarly assuming that learning is the cause of language change. For example, Hare and Elman (1995) report a model which attempts to explain the development of English verb inflection from the highly complex past tense system of Old English towards that of the modern language which has one predominant "regular" inflection and a small number of irregular forms. A network is taught with a data set representative of the verb classes of Old English, but learning is stopped before reaching asymptote by the experimenter in order to simulate the end of the critical period, and this network is then used as the teacher of a new network. For a network, highly frequent patterns, or those that share phonological regularities with a number of others, are learned more quickly and with less errors than low-frequency, highly irregular patterns. As a result, the errors in the first network are passed on to become part of the data set of the second. Those patterns that are the hardest to learn lead to the largest numbers of errors, and over time are "regularized" to fit a more dominant pattern. It is shown that the results of the network simulations are highly consistent with the major historical developments.

Most of the models studying language change presume the existence of competing forms, and are interested in identifying different conditions for particular dynamics of change. The source of the linguistic variants is very often neglected, though some underlying assumptions are made: some variants may come from imperfect learning as a kind of internal factor, while some others may come from language contact. Thomason (2000) points out that the result of language change induced by language contact is to a large extent unpredictable. The argument for this claim is that the factors for the contact situations are very complex. Modeling could serve as a promising approach for us to scrutinize these factors separately or combinatorially and see how much predictability can be attained.

4.1.3 Language origin

There is no fossil as in biology which could allow us to trace the development of language from its prehistoric states. Historical linguistics has been confined to a time limit beyond which it is believed that very little information can be retrieved: "we cannot hope to identify any ancestral languages which were spoken more than a few thousands years ago - perhaps 6000-8000 years ago in a few particularly favorable cases, probably not more than 3000-4000 years ago in most cases. Older genetic links than this undoubtedly exist, but they will remain forever beyond our reach" (Trask, 1996, Section 13.1). However, the inquiry of human language history should certainly go further back in time.

Before the emergence of language, the ancestors of modern humans had their own effective communication system, as we observe in other animals. Therefore it is hard to identify an exact starting point of the process. However, we can take the starting point as the emergence of the earliest anatomically modern humans at least 160,000 years ago (White et al., 2003). Though the anatomy is not the sufficient condition for the origin of modern human language, it at least provides a necessary condition and suggests a lower bound for the time of origin.

How did the full-fledged human language evolve from its proto-form. In the absence of concrete evidence, various speculative theories have been proposed conjecturing different origins of language (Aitchison, 1996). In recent decades, the enquiry has gained more scientific ground. Scholars have tried to find hints and supporting arguments from our relatives the primates, from acquisition, i.e., ontogenetic emergence, from sign languages, from pidgins and creoles, and so on. Computational modeling joins this endeavor in recent decades with encouraging results.

It is unlikely that we can recover what the first early human language would have looked like in every detail, but the computational models provide us with a viable tool to reconstruct the various possible scenarios. Recent empirical studies have identified various crucial factors, including biological factors such as physiological bases (e.g., Nishimura et al., 2003), neural bases (e.g., Ramachandran and Hubbard, 2001; Rizzolatti et al., 1996), high-level cognition capacity (e.g., Tomasello, 2002), and cultural fact such as social structure (e.g., Dunbar, 1993). While these could be crucial prerequisites, it is important to see what the necessary and sufficient conditions are, and how they interplay with each other. Computational models can contribute a lot in this aspect. They serve as a framework to allow manipulation of these factors in a controllable manner, so that their effect can be examined in a systematic way.

When speculating on the process of language emergence, few will believe that a complex language system with elaborate lexicon, morphology and syntax could have sprung up as a whole all at a sudden from scratch. Language must have emerged and evolved gradually and incrementally to reach its modern form through cultural evolution. Symbolization, i.e., the use of a set of conventionalized symbolic signs in a non-situation-specific fashion, is considered as the first stage of language evolution (Deacon, 1997; Jackendoff, 1999; Wang, 1999). The subsequent steps for language origin may involve the emergence of segmental phonology, and of syntactic and morphological structure (Wang, 1999). Jackendoff (1999) hypothesizes seven partially ordered steps in the evolution of language starting from the stage of symbolization. While Jackendoff labels these steps under the title of the "evolution of the *language capacity*", it may be more appropriate to describe them as the steps of evolution of *language itself*. The increase of complexity of language in its form is the result of cumulative inventions of ways to communicate, and learning through generations across a cultural evolution process, rather than the result of a sequence of genetic mutations and the emergence of new language-specific capacities.

Computational models study the process of development between different stages. We can classify the models into two types, according to the modeling approach taken: either functional or emergent. The functional approach usually presupposes some functional criteria to evaluate the fitness of certain *innate learning mechanisms* or certain *linguistic structures*, and uses computational models to demonstrate that some have higher fitness than others, e.g., giving better communicative and/or learning performance. These adaptive language capacities or linguistic structures are shown to be more successful in reproduction, and thus become fixed in the population through natural or cultural selection. These studies usually presume the existence of the learning mechanisms or linguistic structures, and are not concerned with how they have come into being, similar to the situation in some studies of language change where the existence of innovations is presumed.

While the functional approach applies to the development of both language capacity and linguistic structures, the emergent approach is adopted mostly in the scenarios of cultural evolution. The models adopt a bottom-up perspective, simulating the process according to which linguistic forms evolve from pre-structured to structured stages, or from simple to complex structures, without recruiting new cognitive capacities. In these models, the interactions between agents in a population have to be taken into account; these agents modify their internal representations of meaning-form mappings or linguistic rules through language use. One important difference between the functional and emergent approaches is that the latter does not consider fitness of the linguistic structures, or at least no overall explicit fitness is taken into account. The underlying principles in these models share the essence of the theory of self-organization, i.e., the structured language system is an emergent phenomenon from the interactions of language users and the iterative transmission across generations.

4.1.3.1 Functional approach

The first example of this approach is a study by Hurford (1989), which hypothesizes that the Saussurean sign may have evolved as an innate language capacity in humans owing to its adaptive function. "Saussurean sign" refers to the arbitrary association between forms and meanings in human language, each association going equally in two directions: from "meaning" to "form" as used for speaking, and also from "form" to "meaning" as used for listening. For a number of meanings and forms, the two directions of associations form two separate matrices[6], which determine the agent's speaking and listening behaviors[7]. Hurford conceives three possible ways of constructing these two matrices, including Imitator, Calculator and Saussurean, as illustrated in Figure 4.1.

Strategy	Sampling of input from adult population		Learner's acquired behavior
Imitator	Transmission	⟶	Transmission'
	Reception	⟶	Reception'
Calculator	Transmission	⟶	Reception'
	Reception	⟶	Transmission'
Saussurean	Transmission	⟶	Transmission'
			↓
			Reception'

Figure 4.1: Three hypothesized strategies for acquiring the basis of communicative behavior, reproduced from Hurford (1989).

In the model, each new born agent is endowed with one of the three different learning strategies, and acquire his speaking and listening behaviors accordingly by sampling once the population of the last generation. The fitness of an agent is evaluated by computing his "communicative potential" (how well his speaking behavior is understood by other agents), and "interpretation potential" (how well his listening behavior can under other agents' speaking behaviors). The agents are selected as parents of the next generation with a probability proportional to their fitness. Then a new generation of agents are generated and replaces the old one.

In a set of simulations, the initial population is a mixture of agents with equal proportions of agents for each of the three different strategies. After a number of generations, the population is taken up by learners following a Saussurean strategy.

[6]Whether there are two matrices separately for speaking and listening is controversial, which will be discussed in Section 4.4 later.

[7]In the literature, "transmission", "active", or "production" are equivalent terms, all referring to the "speaking" behavior, while "reception", "passive", or "perception"/"comprehension" refer to the "listening" behavior.

Hurford therefore hypothesize that the innate Saussurean strategy would have been selected at an early stage of human evolution as part of the language acquisition device.

Oliphant and Batali (1997) propose an "obverter" strategy, which is actually similar to the "calculator" strategy and they show that under various conditions that a population with obverters can converge to a consistent communication system, unlike what was shown by Hurford (1989) that calculators cannot converge. In their model, the learner constructs his listening matrix by following the majority in the population. Then he determines his speaking behavior as if inverting his listening behavior, that is, he decides to speak in a way according to which he would give the best response if he were the listener himself. It is shown that a coordinated system can be achieved if the learner can sample the whole population. However, the convergence slows down and deteriorates much if learners only have only a limited number of observations, which is more realistic.

Strategy	Sampling of input from adult population		Learner's acquired behavior
Obverter	Transmission	⟶	Reception'
			↓
			Transmission'

Later Batali (1998) uses neural network model instead of two matrices to simulate the interaction, and the strategy designed there was also called as "obverter", though it is different from the original scenario. We will introduce this scenario in a later section.

Hurford (1989) exemplifies the functional approach from a biological evolution point of view, which assumes that the agents's survival and reproduction potential is determined by its language competence, and the language competence is in turned determined by certain genetically inherited traits, such as the three learning strategies. He applies the approach in a recent modeling study explaining why synonyms is rare and homophony is abundant in human languages (Hurford, 2003), in which synonym-avoidance and/or homonym avoidance are considered as innate mechanisms determining the language competence and are subject to natural selection. Similar ideas have also been applied to study the existence of a critical period for learning from natural selection (Batali, 1994; Hurford, 1999b).

However, as we have discussed about "subjacency principle" in Chapter 1, it is necessary to be cautious about what kind of language competence will determine the individuals' survival and reproduction potential. It is unlikely that some language competence related to processing of high level linguistic structure, for example, past tense formation, will affect the biological fitness of the individuals. Instead, the selection force is more likely to exert on language as an evolving cultural system. Some of the modeling studies applying the functional approach adopt this "linguistic selection" scenario. In these models, it is assumed that the structure observed in a communal language at some time instant is the result of a long term competition among a number of variant structures. At the beginning, several competing structures co-exist in a population, but they have different biases of being used or learned as they have different fitness determined by production, perception or cognitive processing criteria. As a result of the repetitive selection through interactions between agents and learning of the next generation, certain variant becomes dominant in the population. Kirby (1999a) reports several models in this vein. He takes parsing complexity and production complexity as the forces of linguistic selection, and simulates the emergence of several language universals.

Christiansen and colleagues (Christiansen et al., 2002; Christiansen and Devlin, 1997) apply the functional approach to argue that center embedding structure is rare because it is harder to learn, instead of being forbidden by some innate syntactic constraint. Grammars with recursive inconsistency, e.g., head-initial structures mixed with head-final ones, can create sentences with center embedding. In one of their simulations, several sets of artificial sentences are generated based on 32 types of grammars which have different degrees of violation of the head-order consistency. Each set of sentences from a grammar is used to train a number of Simple Recurrent Networks (SRNs) to learn the regularities underlying that grammar. In a simulation, the input to the network is the grammatical categories of words in a sentence or a sentence-final marker, and the output is the prediction of the grammatical category of the next word in the sentence. The simulation shows that the learning performance of the networks is highly correlated with the head-order consistency: the higher the inconsistency in the sentences, the more errors the trained networks produce. Similar models are used to show that the learning performance of languages which violate the so-called subjacency constraint is worse than those with no such violations (Ellefson and Christiansen, 2000). As SRNs represent a general purpose learning mechanism without any built-in linguistic bias, these simulations suggest that the non-linguistic constraint of learning and processing sequential structures as present in an SRN can account for the structural characteristics observed in languages, and the linguistic specific constraints proposed as part of the universal grammar (UG) may be obviated.

4.1.3.2 Emergent approach

Various models adopting the emergent approach have addressed the different hypothetical steps between today's language and its origins. How a shared lexicon emerges in the population has been studied extensively, and various possible scenarios have been proposed. Some of these studies are interested in the global convergent dynamics. The evolution of lexicon is simulated variously as imitations between agents (Steels, 1996b), observational learning (Oliphant, 1997), self-organization in individuals by modifying the speaking and listening mappings after each interaction (Ke et al., 2002). In these models, a closed set of meanings are assumed to pre-exist, and these meanings are indivisible and unrelated to each other. During interactions, the meanings intended by the speaker are assumed to be fully accessible to the listener. The agents just learn a set of arbitrary meaning-form mappings, while no regularity within the set of forms, meanings or form-meaning mappings can be learned or exploited. There is no generalization that the agents can make from learning the set of mappings.

Another group of studies pay more attention to the origin and internal representation of meanings in the agents. Instead of a set of meaning-form mappings, lexicon is viewed as a set of form-meaning-referent triples (Hutchins and Hazlehurst, 2001). Only referents are external to agents, and meanings are internal representations inside the agent and not accessible to others. The existence of a pre-defined set of meanings is discredited (Hutchins and Hazlehurst, 1995) and meanings are shown to be grounded with external stimuli and gradually emerging from communication (Steels, 2001). Not only the form-meaning mappings evolve, but the meanings evolve as well. Hutchins and Hazlehurst (1995) and Steels (2001) develop models to simulate how the meaning and meaning-form mappings co-evolve in a group of agents exposed to the same set of scenes they try to distinguish or categorize by verbal communication. Agents are represented by auto-associator networks (Hutchins and Hazlehurst, 1995) or association rules (Steels, 2001). The external structure present in the referents and the coordination through interaction between agents lead to the emergence of the intra-agent structured referent-meaning-form mappings as well as inter-agent referent-form mappings. Luc Steels and colleagues have developed robots in the Talking Head experiment and subsequent projects to address empirically the meaning grounding problem (e.g., Steels, 2001; Steels and Kaplan, 2002; Steels and Vogt, 1997). Cangelosi and Harnad (2000) and Cangelosi et al. (2002) have a similar emphasis on the grounding problem by simulating with neural networks to categorize physical objects.

The above studies mainly deal with the emergence of a set of shared symbols. Studies on other steps of language evolution have also been reported. Regarding the emergence of segmental phonology, it has been addressed recently in de Boer (2001) and Oudeyer (2003). Now we will introduce two studies to show how the later steps of the emergence of syntactic structures can be studied by modeling.

Compositionality and hierarchical structures in language are two main defining features of human language. While nativists ascribe the emergence of these features to an innate language acquisition device, modelers taking the emergent approach aim to show that such complex structures in language can emerge from some simple unstructured or holistic communication system in a group of interacting agents. The agents do not have the intention nor the global view to create a language with specific structures, but are simply equipped with the intention to communicate, the readiness of using existing available resources to express themselves, and some general learning mechanisms. Batali (1998) and Kirby (2002a) are two such representative studies which can illustrate the general practice in this line of research.

Batali (1998) calls his model a "negotiation model", in which each agent alternates between learning to interpret the sequences sent by others, and sending sequences for others to emulate. There are iterative interactions, each of which involves one speaker and one listener randomly selected from the population. An interaction takes place as follows: the speaker selects a meaning from a given fixed inventory, and expresses this meaning by an utterance and sends it to the listener. The listener interprets the received utterance according to his own experience, compares the inferred meaning to the speaker's intended meaning, and adjusts his own internal representation of the meaning-form mappings. In the model, it is assumed that the agents all have a number of pre-existing simple meanings, such as (me happy) or (all sad). There are totally 100 meanings from a full combination of 10 referents and 10 predicates. What agents actually do is to develop a shared consistent set of utterances to represent these meanings.

The speaking and listening behaviors are both represented by a simple recurrent neural network (SRN). The input of the SRN is the form (also called signal or utterance), and the output is the meaning. The speaker tries different utterances and uses the one which gives the best match with the intended meaning according to his own listening behavior. The listener inputs the received utterance to its SRN and obtains a meaning. Next the SRN is trained consequently after the listener compares his inferred meaning with the intended meaning. This is close to the "obverter strategy" discussed earlier.

The simulation shows that at an early stage, the agents communicate with each other totally uncoordinatedly and with a low rate of communication success. Few, if any, agents send the same sequence for the same meaning, and they are rarely able to interpret sequences sent by others correctly. However, after a large number of interactions, the situation changes: agents represent the meanings with approximately the same utterances, and the communications reach a high success rate. More interestingly, the mappings between meanings and forms exhibit certain structures, such as the same referents combining with different predicates sharing some similar parts in the forms, as shown in Table 4.1. Furthermore, in a second experiment, some meanings are deliberately not used during early communication. It is found

Table 4.1: A simulation result from Batali (1998)'s model. The majority of the population shares a set of meaning-form mappings which show certain combinatorial characteristics.

	one	they	you	yall	yup	me	we	mip	yumi	all	"root"
tired	cda	cdab	cdc	cdcb	cdba	cd	cdd	cddb	cdcd	cdb	cd-
scared	caa	caab	cac	cacb	caba	ca	cad	cadb	cacd	cab	ca-
sick	daa	daab	dac	dacb	daba	da	dad	dadb	dacd	dab	da-
happy	baa	baab	bca	bcab	baac	ba	badc	bab	bac	babc	ba-
sad	aba	abab	ac	acb	abac	a	abdc	abb	abc	abbc	a(b)-
excited	cba	cbab	cca	ccab	cbca	c	ccdc	cb	ccb	cbc	cb-
angry	bb	bbb	bc	bcb	bbc	b	bddc	bdb	bdc	bdbc	b-/bd-
silly	aa	aaab	aca	acab	adba	add	addc	adad	adc	adbc	a(d)-
thirsty	dbaa	dbab	dca	dcba	dbca	dda	ddac	dbad	dcad	dbacd	d(b)-
hungry	dbb	dbbd	dc	dcb	dbc	dd	ddc	dbd	dcd	dbcd	d(b)-
"root"	-a	-(a)b	-c/-c-	-cb	-ba	(-a)	-d/-dc	-db/-b	-cd/-c	-b/-bc	

that after the convergence of a shared consistent system, these novel meanings can be expressed with forms sharing similar parts of the existing ones. This suggests that the agents have developed some combinatorial systems which can generate novel meaning combinations with existing forms.

Kirby (2001) reports another type of simulation models to demonstrate the emergence of a compositional language from an early holistic communication system. This work is one of the series based on the Iterative Learning Model (ILM) (see a review by Kirby, 2002a). A significant difference between Kirby's and Batali's model is that Kirby emphasizes the role of language learning as a bottleneck, while Batali does not treat acquisition specially and simply considers the long term effect of general interactions without distinguishing adults and children. Another difference lies in the implementation. While Batali uses neural network to simulate the representation and processing of language, Kirby uses symbolic rule-based representations. Either holistic utterances, words or grammatical structures are represented by rules, such as

$$S/eats(tiger, john) \rightarrow tigereastsjohn$$

$$S/p(x, y) \rightarrow N/x \; V/p \; N/y$$

$$V/eat \rightarrow eats$$

The symbolic representation provides a convenient way to study complex meanings. While Batali's connectionist model only deals with one-place predicate meanings, Kirby's model deals with two-place predicates, as well as one- and two-degree embedding, such as

likes (gavin,mary), says (mary,admires(gavin,mary))

says (mary, thinks (mary, praises (john,gavin)))

The basic idea of Kirby's model is that at the beginning, agents randomly create holistic utterances to convey a set of meanings. When some recurrent patterns appear in these holistic utterances, i.e., two holistic utterances share some subset of strings and the corresponding complex meanings share some atomic meanings elements as well, then the agent can extract the patterns. It is assumed that the agent can realize the relation between the recurrent sub-string and meanings, and construct the association between an utterance and a meaning as a word. When a word is extracted, the remainder of the utterance is decomposable as well. Subsequently, the original holistic utterance becomes combinatorial, and certain combinatorial rules of word order emerge. Once such order rules are available, the agents can convey novel meanings by combing words using these order rules. Kirby's model shows that through various random creations, and extraction of words and order rules, the population starting from a random holistic communication system will always develop a highly efficient combinatorial communication system.

From the above two studies, we can see several common features of these works. One is that there is a pre-existing meaning space. This has been criticized as unrealistic (Hutchins and Hazlehurst, 2001). However, the preexistence of a complex structured meaning (or semantic) system in humans before the emergence of a complex syntax has been strongly argued for (Schoenemann, 1999). Moreover, the choice of simplification in this aspect may be necessary as the initial attempts to model the complex process of language origin.

These two studies are faced with a similar challenge: in each interaction, the listener is assumed to know what the intended meaning of the speaker is. This assumption is known as the "explicit meaning transfer" problem (Gong et al., 2004; Smith, 2003). How likely is it that the listener understands the meaning in exactly the same way as the speaker intends? Meaning in use is very flexible and fluid (Croft, 2000, Chapter 4). In reality such meaning transference is controversial. The well-know "gavagi" example given by Quine (1960) illustrates well this problem: a speaker pointing to a rabbit saying "gavagi" could mean the rabbit itself as a whole, or its head, ears, legs, color, shape, or even the running action of the animal. In the field of language acquisition, there are two different perspectives concerning this problem. One school of thought argues that very often children do not get the exact meaning that the caretaker refers to, but instead produce over-extension or under-extension of words' meanings (Barrett, 1978). On the other hand, scholars find that children show high degree of attention to the social context, and that they can reach a high degree of understanding with the adults (Tomasello, 1999). The importance of such social interactional context can be shown in early babbling not only in babies but also in birds (Goldstein et al., 2003).

modelers may have two possible reactions to this charge against "meaning transference". One is to justify that such meaning transference is plausible in real situations. Primatologists have found that chimpanzees' social behaviors are very complex. They can understand, predict and manipulate others' behaviors in the group. These complex behaviors suggest that chimpanzees have powerful cognitive abilities in terms of social intelligence (de Waal, 1998; Tomasello, 2000). It has been shown that children at a very early age also exhibit such social intelligence (Tomasello, 1999). Children have a remarkable ability to read the intention of others (Tomasello, 2003); they can in particular attend to the object that the adults are drawing their attention to. Baron-Cohen (1995) even suggests that mindreading is an innate ability in humans, and that the autistic patients are those who are unable to read others' mind. Based on this evidence, it may not be unreasonable to assume that the listener is able to decipher the intended meaning of the speaker.

Though the listener may not get the exact meaning, at least he can guess a related meaning from the contextual information in the environment. Some models have investigated this issue – how will the reliability of the meaning transference affect the performance for child's language learning or language convergence in a population. It has been shown that there exists a high degree of robustness. For example, Schoenemann (2002) reports a neural network model of lexical learning, in which some wrong meaning-form mappings, in addition to the correct ones, are used to train the network. It is shown that the neural network can still learn with a high accuracy, provided that the correct mappings are frequent enough in the input. Gong et al. (2004) adopt a similar idea, where the meaning available to the listener, called "environmental cues", may incorporate one or more meanings, either the exact meaning intended by the speaker and/or some distracting meanings from the environment. The availability of the correct meaning is dependent on the "reliability of cue(s)", which is a parameter of the model. It is shown that when the reliability of cues is higher than 0.7 (a value of 1.0 means a perfect meaning transference), the group of agents is able to converge efficiently to use a shared language. Smith (2003)'s "cross-situational statistical learning" model employs a similar idea to Schoenemann and Gong, in which the agent builds up all associations between the received signal and the possible meanings inferred from the context, and the more frequently recurrent associations will get recognized. He shows that the larger the context size for agents to infer meaning, the longer time for a shared language to emerge, and suggests that the ability to narrow down the size of the context, i.e., to share focused attention with others during communication, may be selected during evolution.

The second response to the charge against "meaning transfer" is to modify the models by grounding the representation of meanings, as shown in the work discussed earlier about simulating the emergence of meanings. Steels and Kaplan (2002) simulate the intended meaning as a description of an object which is in the visual focus of both the speaker and the listener, and this object is represented by its physical properties, such as its color, shape, and size, etc. Smith (2003, 2005) show that if the external world is structured rather than random, and if the agents are able to create meanings with some intelligent strategy, such as the "principle of contrast" (Clark, 1987), or "mutual exclusivity assumption" (Markman, 1989) which are found to be present in children's language acquisition, the agents reach high mutual understanding, while their internal representations of meanings are only with low similarity.

So far the models can only deal with a limited set of meanings within such embodied representation, or only employ some abstract representations which are mainly used for proof-of-existence. However, due to the limited knowledge regarding the internal representations of meanings so far, and the fact that language learning is based on rich world knowledge which is hardly incorporated in the models, the current models are still constrained with limited semantic domains and deal with only simple static objects (e.g., Steels and Kaplan, 2002) or simple object-action meanings (e.g., Munroe and Cangelosi, 2003). When meanings become more complex, such as predicate of degree-1 or higher of embedding meanings (Kirby, 2000), meaning transference is either explicitly or implicitly assumed in the models.

4.2 Modeling at different levels of resolution

François Jacob in his seminal article "Evolution and tinkering" (1977) remarks that the different sciences can be arranged into a hierarchy - physics, chemistry, biology, psychosociology, according to an order which corresponds to the hierarchy of complexity of the object being studied, from atoms to molecules, to organisms, to animal species. As an analogy, the computational models in language evolution can be classified as a similar hierarchy. Based on the degree of resolution with which "language" is represented, we classify the models into four levels, arranged in an order from "synthetic high level" to "grounded low level".

4.2.1 The first level – language as a synthetic whole

At the first level, language is taken as a complex but synthetic whole. The models do not consider the internal structure of the language, but assume this language has an overall fitness and evolves as a whole. Distances between languages can however be measured, and the accuracy of acquisition of a language can be quantified by a variable. Based on this formulation with the greatest simplification, the evolution process can be studied from a macro perspective.

Nowak, Komarova and Niyogi (2001) can be taken as an example of modeling language as a whole. They develop analytical models similar to those in population dynamics in biology to examine the conditions of evolution of the universal grammar. An initial population is created with a mixture of agents speaking different languages, each of which is assumed as a possible universal grammar. Each language has a fitness (or "payoff", as used in their article), which is measured by its communication success, in terms of its distance to other languages weighted by its frequency in the population. The reproductive rate of a language is determined by its probability of acquisition and its fitness. The evolution dynamics of the population is studied by a set of differential equations. The authors investigate the stability and convergence of the system, and the effect of various parameters such as the accuracy of acquisition, the number of learning samples, etc. It is shown that the population does not always converge to a single language. Under the simplified condition that all languages are equidistant, the population stabilizes at one dominant language only when the accuracy of language acquisition is higher than a certain threshold; otherwise the population remains with several grammars co-existing. This model provides a framework to study analytically the dynamics of language evolution.

Coupé (2003) develops a computational model to investigate the dynamics of language evolution in a fitness landscape. This is another example of studying language evolution by taking language as a synthetic whole. A language, whether an idiolect or a communal language, is considered as a point in a fitness landscape, which is a high dimensional space, determined by various constraints on language, either physiological, cognitive, communicational or social. Different languages have different fitness values according to the degree they fulfill these constraints. There are several attractors in the landscape, representing optimal configurations based on the constraints. The evolution of a language is illustrated as a trajectory of movement in the landscape[8].

[8] Nettle (1999b) hypothesizes a very similar picture of a high-dimension fitness space. He considers that the existence of multiple optima with different basins of attraction in the space may be an explanation how language diversity arises. Small variations are amplified by social or functional selections and languages evolve to approach different optima. But Nettle (1999b) only provides a description of the idea without further elaboration.

Coupé (2003) compares three conceivable types of evolution in the landscape. The first type is a deterministic type of evolution. A language always evolves according to a gradient descent in the landscape and therefore always approaches the nearest attractor, i.e., the local optimal configuration. In this situation, the language always stabilizes at one attractor after a given number of iterations. A second type of evolution does not depend on fitness, and the language moves in the landscape randomly. No stable patterns can be observed and the language keeps changing constantly. The third type of evolution is that a language changes according to a stochastic Gaussian distribution function. A language is likely to approach the local optimal with a high probability, but there are also some small chances to move away from it. The injection of limited stochasticity in the model demonstrates how a continuous evolution is possible in the fitness landscape.

A language starts from a random position in the landscape; it keeps evolving within the basin of attraction of a certain attractor. But it does not stabilize at the position of the attractor. Occasionally some dramatic changes happen, as the language jumps from the basin of one attractor to that of another. While the first two types of evolution are both unrealistic, the dynamics generated in the third scenario are consistent with the real dynamics of biological evolution and language evolution in particular: languages keep changing all the time, following some universal tendencies of change. The model predicts that the dynamics may exhibit the "punctuated equilibrium" phenomenon which has been discussed in biological evolution (Gould and Eldredge, 1977) and recently in language evolution (Dixon, 1997).

The first level of analysis provides us with a way to investigate the mechanisms and dynamics of language evolution from a macro perspective, while language is represented as a holistic system without internal structure considered. However, this holistic representation of language cannot deal with many realistic considerations. First, language would not have emerged or changed as a whole. Different parts of a linguistic system are subject to different constraints and have different evolutionary dynamics, and therefore require different treatment. Especially for the study of language emergence, the various complex structures, such as the phonological structure, syntactic structure, or morphological structure, etc., should have appeared gradually in steps rather than emerged altogether at once. Therefore it is necessary to consider these individual steps diachronically and individual subsystems synchronically. There have been a number of models which deal with these subsystems specifically, which we classify as models taking the second level of resolution.

4.2.2 The second level – subsystem as an independent whole

Most of the models reviewed for language acquisition in Section 4.1.1 belongs to this level of analysis. Each acquisition model is dedicated to a certain learning task in a specific linguistic domain, such as English past tense (Rumelhart and McClelland, 1986), segmentation of continuous speech (Christiansen et al., 1998), distinction of grammatical categories (Elman, 1990), and so on. These models focus on isolated individual language learners at the synchronic stage.

Meanwhile, a few number of models take into account the diachronic aspect, examining the evolutionary dynamics through vertical transmission - language learning. These models adopt a selectionist assumption: systems with better fitness have a higher chance to reproduce, or a higher chance to be learned by the next generation in this case. In this type of models (e.g., Kirby, 1999a), several competing variants first co-exist in the population, being used by certain proportions of agents; a variant having a higher fitness will be preferred and increase its proportion in the population. Similar to natural selection in biological evolution, it is through differential reproduction of the competing variants that one or several optimal linguistic structures become dominant in the population. However, these models have little concern about where the linguistic variants come from; the origin of the competing linguistic variants still require further explanation. Moreover, the population is shown to converge to become homogeneous in using one linguistic variant while other non-optimal ones simply all die out. Such homogeneity, however, is very rare in real situations.

Some models apply the selectionist framework from another perspective; it is assumed that language performance affects biological fitness: parents with certain structures having a better communicative fitness will produce more offspring who will learn their language, e.g., Komarova and Nowak (2001); Nowak et al. (1999). Komarova and Nowak (2001) apply the population dynamics framework to examine the evolutionary dynamics of lexicon which is represented by a matrix, under two different conditions: incomplete and imperfect learning. They calculates the maximum error rate that is compatible with a population maintaining a coherent lexical matrix of a given size using an analytical model.

A number of models apply the selectionist framework from an optimization perspective. It is assumed that the linguistic subsystems are constrained by some optimization criteria concerning production and/or perception/comprehension, and the optimization process is carried out in an abstract sense or in individual language users, without considering language use in communication at all. Ke et al. (2003); Liljencrants and Lindblom (1972); Lindblom (1986); Redford et al. (2001) are some of the models on the optimization of phonological systems.

The models at the second level of analysis do not consider the actual language use. Especially, the optimization models assume that the optimization is carried out in the individuals, which is neither likely nor necessary in reality. As a comparison, models at the third level embody the evolution of language in language use, and in the population, which is pertinent to an evolutionary framework.

4.2.3 The third level – language embodied in use

Along with the convergent view from empirical studies, it has become more evident to modelers that it is necessary to situate language in *the process of use* in *the population of its users* in studying language evolution, no matter acquisition, change or emergence. Language use is modeled by simulating pair-wise interaction between agents, i.e., a speaker and a listener, which are either chosen randomly from the population, or selected according to various distance constraint (an agent only interacts with its neighbors or with those which have social connections with him). Interaction can also take place between a teacher and a learner in a situation of language acquisition situation. Each agent has his own representation of language, i.e., his idiolect, which is not an abstract synthetic whole, but consists of some components with internal structures. It is represented as either a look-up table of mappings between meanings and forms (Steels, 2001), or nonlinear mappings between input and output of a neural network (Batali, 1998), or a set of symbolic rules (Kirby, 2001). A very important feature of the models at the 3rd level is that an agent's idiolect is formed and changes as the result of the interactions with other agents. What is evolving is a set of idiolects.

Most models which belong to this level address the problem of language emergence, such as the emergence of vowel systems (de Boer, 1997, 2000), or the emergence of combinatorial syntactic structures such as Batali's and Kirby's models introduced in Section 4.1.3.2. In the latter two models, it is assumed that a set of meanings preexist. The population of agents first create arbitrary forms to express these meanings. Through learning or processing, the listener re-organizes his own language accordingly. Through such iterative process, the idiolects in the population converge to reach a high rate of mutual understanding, and exhibit combinatorial structures.

The "meaning transfer" problem discussed in early section is present in many models at the third level. It is often implicitly assumed that the understanding between the speakers and listeners poses no problem, or the speaker's intended meaning can be obtained by the listener. To tackle this problem, some computational models go further in refining the part how meanings are represented and then transmitted. Models of this type are classified as belonging to the fourth level.

4.2.4 The fourth level – meaning and/or form embodiment

At the fourth level of analysis, the representation for language, in terms of the mapping between meanings and forms, is further elaborated. Meanings neither pre-exist nor are independent from each other, but often emerge from interactions between agents. For example, in the Talking Head experiments with real robots and related simulated experiments by Luc Steels and his colleagues, the agents receive visual information as input from their sensory channels, for instance, the color, shape and size of an object which is within their visual focus, and encode or decode the information with a decision tree. The meaning is not explicitly shared by agents any more, but is inferred through their own perception of the shared environment and their analysis of the received utterance. Steven Harnad and his colleagues have discussed the "symbol grounding problem" which shares similar concern that the meanings associated with symbols should be grounded into the physical environment, at least for some initial steps of the development of the symbolic system.

In some models, forms are represented in more elaborated or realistic manners. There is no doubt that strings of sounds are not processed as wholistic entities in the brain, but rather through the distributed activation of the neurons and their connections in the neuronal network. Though artificial neural networks at this stage are still far from being realistic, they conceptually catch the essence of such reality of processing. In the models proposed by Oudeyer (2002; 2003), the processing of sounds is represented with neuronal maps which self-organize into various patterns. Though these models mostly still deal with vowel systems or some hypothetical simple consonant-vowel combinations, they are moving toward the ultimate aim of simulating with biological plausibility.

From the above summary of the models at four levels, it can be seen that computational models for language evolution vary a lot in their theoretical orientations and focus of investigation. Some are interested in the evolution of the biological basis of language competence, studying how an innate universal grammar may be possible (e.g., Briscoe, 2000b; Nowak et al., 2001); while some are interested in different aspects of cultural evolution, showing how language evolves in a similar way as other cultural systems. For the latter perspective, language itself evolves through iterative language use and language acquisition in a heterogeneous population (e.g., Batali, 1998; Kirby, 2002a; Steels, 2001).

To conclude the summary of this hierarchy of analysis at four levels, it is worthy to note that these four levels are not mutually exclusive. Some mobiles may fall into more than one level of resolution. For example, some models may adopt a holistic representation, but simulate the interaction in embodied situations. More importantly, the distinctions between these four levels of analysis do not imply that any one level is more pertinent to our investigations in language evolution. In fact the different levels of analysis provide mutual benefits to each other. The higher levels of analysis focus more on the general mechanisms governing evolution, which may influence the choice of principles in designing the models at lower levels. Conversely, the findings from the models at the lower levels may provide evidence and arguments to support or falsify the theories investigated by the models at a higher level.

4.3 Agent-based modeling

Among the models at the high level of representation, some are analytical models using mathematical formulation (e.g., Komarova and Nowak, 2001; Minett and Wang, 2003; Niyogi, 2002; Niyogi and Berwick, 1997; Nowak et al., 2001). These models are interested in the deterministic behavior of the system, by carrying out the equilibrium and stability analyses, and studying the average characteristics of agents in a population of infinite size. In comparison, models at the lower levels of representation often adopt the simulation framework, which examine the actual interactions between agents in a population of limited size. Such a framework is sometimes called agent-based modeling (ABM).

Agent-based models have been widely used in the study of *complex adaptive systems*, in which a number of individual components, i.e., the "agents", continuously interact with each other, leading to the emergence of some global structural patterns from such local interactions (Holland, 1998). The agents are autonomous, share similar basic characteristics, and act according to simple rules and local information. The simple interactions among the agents often lead to complex structures in the system. This research framework represents a "bottom-up" approach of studying real-world complex systems, instead of analyzing them from a top down perspective. It has been proven fruitful in offering new insights in the study of man-made systems such as stock markets and traffic jams, etc., and natural systems such as immune systems, ant colony, etc. (e.g., Bonabeau et al., 1999). The study of honeybee combs introduced in Chapter 1 is a successful and comprehensive example of ABM.

To view language evolution as a self-organization process fits well with the ABM framework. Many models discussed above, such as Batali (1998) and Kirby (2001), are good examples, though ABM is not explicitly stated in their work. The models developed in this thesis, which I will report in the next chapter, all adopt this framework.

One of the most important theoretical implication for adopting ABM framework is to emphasize the importance of interactions between agents in the system. As Briscoe remarks, "it is necessary to move from the study of individual (idealized) language learner and users, endowed with a LAD and acquiring an idiolect, to the study of populations of such generative language learners and users, parsing, learning and generating a set of idiolects constituting the language of a speech community" (Briscoe, 2002, p257). In the following, I will summarize some features shared by models adopting ABM in the study of language evolution.

In the ABM models, individual language users are the basic components of the system, i.e the agents. These agents share similar characteristics, for example, articulation and perception of sounds (de Boer, 1997; Oudeyer, 2002), or some general learning mechanisms such as imitation, association (Cangelosi and Harnad, 2000; Ke et al., 2002; Steels and Vogt, 1997), or recurrent pattern extraction (Gong et al., 2004; Kirby, 1999b, 2000). The representation of the language in the agents can mainly be divided into two types. One is neural networks, which is characterized by its distributed nature. The input of the network may be the meaning represented by some grounded features of physical objects, and the output the corresponding linguistic form or signal (Cangelosi and Harnad, 2000). Conversely, the input of the network may be the signal and the output the meaning (Batali, 1998). The other type of representations is symbolic. Language is represented by a set of rules, either lexical (Kirby, 1999b), syntactic (Briscoe, 2000b; Kirby, 1999b), or by exemplars composed of strings as forms associated with formula sets as meanings (Batali, 2002). While agents are assumed to be governed by similar underlying mechanisms to learn and use language, the actual representations of language in the agents may differ in various ways. For example, in the neural network models, though two agents appear to have the same meaning-form mappings, the internal weights of their individual network may be very different. In some models, more heterogeneity may be taken into account in the implementation. For example, agents may be classified as in different age groups or social classes, or have different linguistic characteristics in terms of learning capacity or social impact (Ke et al., 2004; Nettle, 1999c).

Parallel to the consideration of implementing individual agents, the model needs to address the conditions of interactions among agents. Agents interact with other agents in the population iteratively. The interactions may take place in a random way, i.e., for each instant two randomly selected agents interact (Batali, 1998). Or agents interact only with the nearest neighbor (Kirby, 1999b), or with a number of neighbors within a certain distance (Livingstone, 2001; Nettle, 1999c). While we know from socio-linguistic studies that social networks play an important role in determining individuals' linguistic behaviors, few studies based on computational models have taken various social networks into account. In Chapter 5 I will report some preliminary studies in this aspect.

4.4 Relationship between production and perception

4.4.1 Evidence of the disparity

The relationship between production and comprehension (or perception when focusing on speech sounds only) is an intriguing issue in various areas of linguistics. Various types of evidence have suggested that comprehension and production are separate processes, and that there is a split between the two (Straight, 1976).

This disparity has been examined extensively in children phonology acquisition (Clark, 1993; Jusczyk, 1997). For example, children can perceive more sound distinction than they produce as shown in the process of their lexical development (Matthei, 1989). The disparity is not surprising at all, once we are aware of the differences in the sensory and motor systems underlying perception and production. The large changes in shape and dimensions of the vocal tract and the complex control required for the coordination between articulators are both causes for the lag of production behind perception.

Not only in children does such disparity exist, it is also found in adult language users and language-learning chimpanzees. It is doubtless that we always understand more than we can say. We are able to understand not only people from different background, whose choices of words and style are very distinct from ours, but also people from other places who have a very different accent. The accommodation to the variation of others' production is a clear evidence of the flexibility of comprehension over production. Chimpanzees trained by human to learn language are able to understand a considerable amount of spoken language, while their active use of language remains poorer than 2-year old children. We may have more direct evidence from linguistic pathologies: in many aphasias, there is dissociation between production and comprehension. For example, Broca aphasia patients' comprehension of language is usually relatively intact and they can understand oral communications from others. However, their speech production is usually reduced to some stereotypic nonsense phrases. In a more specific linguistic pathology affecting the lexical domain, called anomia (Caramazza and Berndt, 1978), patients have a specific inability to name objects. They cannot utter the name of a given physical object, but when the name is provided, they can recognize it.

4.4.2 Role in language evolution

What has caught researchers' attention is the conflict between production and comprehension and its consequence in language evolution, usually from a functional perspective. It has been often suggested that the linguistic structures are the result of some balance between factors from these two aspects. For example, the universals of phonological structures are explained by the compromise between sufficient perceptual distinctiveness and minimal production cost (Lindblom et al., 1984); the persistence of homonyms and synonyms is considered as an unavoidable result of the conflict between production and comprehension (Steels and Kaplan, 2002). A recent study proposes that the many-to-many structure of the lexicon is the consequence of the minimization of two functions concerning these two aspects (Ferrer i Cancho and Solé, 2003).

Besides the conflict between production and comprehension, their mutual reinforcement has received less attention. However, concerning language origin, the relationship may be of great effect. Burling (2000), from a purely theoretical perspective, suggests that "communication does not begin when someone makes a sign, but when someone interprets another's behavior as a sign" (ibid, p30)[9].

The recognition of the disparity also provides a strong support for the importance of considering the heterogeneity in a language community as sociolinguists have been emphasizing it. It is related to the "Saussurean paradox", which states that "if a language is primarily an orderly system of relations, how is it that a language can change without disrupting that system?" (Trask, 1996, p267). In other words, "how can a language continue to be used effectively as a vehicle for expression and communication while it is in the middle of a change, or rather in the middle of a large number of changes?" (ibid).

The answer lies in language's heterogeneous nature. The paradox can indeed be easily resolved when one recognizes the fact that people can understand much more than they would speak, in other words, the comprehension is always a super set of the production. It is this larger comprehension that provides the capacity of individual language users to accommodate variations and changes in the language community. People may create their own way of speaking but they are understood most of the time (if not all the time), and can understand others who speaks differently from them.

Moreover, effective communication in an absolute sense may be an illusion for us. In fact language never serves us as a perfect communicative media. Communication does break down now and then in individual occasions, because of vagueness or ambiguity. However, the sporadical misunderstandings are either ignored or repaired, as illustrated by some examples of mis-communication caused by homophones which are shown in Appendix 1.2. People do have various strategies to repair the confusion, either by rephrasing the sentence or by adding new information. Language has a rich redundancy to allow many ways to say the same thing, by using different synonyms, different constructions, etc.

Besides repair, ambiguity avoidance is another result of communication breakdowns. Certain words or phrases which previously caused confusion may be deliberately avoided in later interactions. Taboo avoidance is a common phenomenon to

[9]This proposal is close to the Chinese saying 说者无心，听者有意 ('shuo zhe wu xin, ting zhe you yi'. "What is heard is not what is said".)

this respect. Hankamer (1973) shows that "structural ambiguity which might be expected to arise from deletion is in fact always avoided" (p17), which Straight (1976) considers as an example of speakers' self-monitoring and consequent self-revision in the production process.

Even when noting the existence of communication break-downs, language still appears as being sufficient in ensuring effective communication at the global level. The heterogeneity and redundancy are the two important characteristics of language which can explain the perpetual instability and its high degree of effectiveness. However, how to prove this claim and test it in different languages is beyond the reach of empirical studies. It is a notorious problem that it is hard to compare different languages in terms of their complexity and effectiveness as a whole, because we don't know how to combine the various parts of a language, such as its lexicon, morphological system and syntactic system, into an integrated whole and measure their effectiveness for communication. Moreover, the heterogeneity is so prominent and there may be no way to choose one single representative idiolect or any intersection of the idiolects to make the comparison. Fortunately, computer modeling may offer special contributions in this aspect. As models can simulate individual interactions over a long time, the communication effectiveness can be measured quantitatively by calculating the percentage of successful interactions, while having a control on the complexity of the language. We may compare different hypothetical languages first, and apply the models to real languages, once systematic measurements are developed.

4.4.3 Modeling the disparity

Empirical studies in language acquisition have been concerned with the disparity of comprehension and production. Some believe in a "two-lexicon" hypothesis, which suggests an input lexicon used for recognizing words, and a separate output lexicon for word production, derived from the input lexicon (Menn and Matthei, 1992). Some others propose a "two-entry" model (Matthei, 1989) in which there is only a single lexicon but separate access routes for perception and production; each route might develop independently, at least initially. For more details, one may refer to a brief review of these two models in Jusczyk (1997, 190ff).

Similar to these two distinctive proposals from empirical studies on mental lexicon, recent studies using computer models to simulate language evolution share the same division of assumptions. First, some existing models do not make the distinction between production and comprehension. For example, in the series of studies on the iterated learning model (ILM) (Kirby, 2001), the linguistic systems for production and comprehension are the same without differentiation. In a typical setting of the ILM, an agent first learns a language by listening to a neighboring teacher and generalizing a set of rules, and later speaks to a new learner by making use of this set of rules. Due to the absence of distinction between the linguistic system for production and comprehension, the model cannot simulate a situation where production is only a subset of comprehension.

Batali (1998) simulates production and comprehension as two different processes: while comprehension is simulated as the input-output mappings by a simple recurrent neural network, production is taken as an inverse function of the comprehension using the obverter strategy (Oliphant, 1997; Oliphant and Batali, 1997). However, as the same neural network is used in these two processes, the same set of meaning-signal mappings for both production and comprehension is used for one individual when the language stabilizes.

Komarova and Niyogi (2004) have some arguments against the distinction in their study of the evolutionary dynamics of the lexicon. They suggest using only one common association matrix to represent both production and comprehension mappings. They consider that it is natural to give symmetric considerations to forms and meanings, and treat language as a relation between forms and meanings rather than two separate functional mappings for production and comprehension. Moreover, they suggest from a computational standpoint that a common matrix provides a more compact representation for a language, from which production and comprehension modes may easily be derived.

It would be unreasonable to hypothesize totally dissociated production and comprehension; to simply equate the two would be unreasonable either. However, though we recognized the disparity of the two, it is far less clear what are the relationship between them and how they influence each other. This issue becomes especially important for simulation modeling because the models require all the assumptions to be clearly implementable. Though intuitively, the correlation between the two is very strong at the beginning of language acquisition and decreases after adulthood, it is not clear yet how they are exactly related. Obviously it is very difficult to make the comparison of the two quantitatively in empirical studies. To evaluate active and passive vocabularies is already very difficult, not to mention an overall comparison of production and comprehension, or tracing the development of the two.

Simulations may provide a solution to this problem. Modelers may test different hypothesized situations and possibly identify some plausible scenarios for empirical studies to justify. Typically, we can compare the evolutionary dynamics of a model implementing the distinction with the dynamics of a model without the distinction. If the models do not produce significantly different results, it may then be acceptable to make the assumption of equal production and comprehension. Otherwise, it will suggest that the simplification is not appropriate.

4.5 Pros and cons of computational modeling

Compared to empirical studies, computational modeling conveys a number of advantages for research in language evolution. It provides "virtual experimental laboratories" to "run realistic, impossible, and counterfactual experiments", and "test internal validity of theories" (Cangelosi and Parisi, 2001, p2-3).

Compared to traditional mathematical models which are considered as "passive", computational models are "executable" (Belew et al., 1996, p432). While mathematical models are often limited to describe the aggregate characteristics of a system and often require infinite populations, computer models can capture the various complications that are difficult to treat in mathematical models. Computational models provide the possibility to study the transient behaviors which are often of more interest than equilibrium states. More importantly, computational simulations provide a framework to study the various parameters, initial conditions and boundary conditions under well control. Experiments can be run with unlimited duration and repetitions under the same condition, so that statistical analyses can be performed for comparison with empirical data from language use and language change.

Furthermore, computational models may provide a more convenient way for cross-disciplinary communication and cross-fertilization. Fundamental scientific advances frequently emerge from the discovery of unsuspected connections between disciplines. As computational models always require the arguments, conditions and parameters being explicit and quantified, they can " 'export' results from one discipline to another in a comprehensible fashion" (Belew et al., 1996). Therefore knowledge from one scientific community becomes more accessible to researchers from other disciplines, and connections are more possible to be built. Researchers may further extend these models to investigate specific questions in their own areas. For example, inspired by Hinton and Nowlan (1987) which introduces a simple model demonstrating the Baldwin effect in general, Briscoe (2000b), Turkel (2002), and Munroe and Cangelosi (2003) have addressed similar questions but more specific to language evolution. The recent and rapid development in complex network studies in other areas (Barabási, 2002; Watts, 1999) have triggered a surge of studies of networks related to language (Ke, 2003). A more detailed introduction will be given in Chapter 5.

Critics of computer simulations often complain that the researcher has simply built into the model whatever he/she desires, and therefore the results provided by the models are neither unexpected nor interesting. It is often accused of being circular in some occasions. However, as Nettle (1999c) points out, in fact these are misunderstandings of what simulations are for. The interest does not lie in what the model can be made to do, but rather what assumptions and initial conditions have to be included to make the model produce the desired result. The beauty of modeling is not only in the demonstration of some results which are expected, but more importantly in the process of building the model. What we learn from the model is in building it and in using it (Morgan, 1999). First, the modelers have to identify a set of explicit and implementable assumptions based on the chosen theoretical basis, and choose some arbitrary values for the parameters. When the simulation leads to dead-ends, or unexpected outcomes, the modelers will have to modify the existing assumptions or parameters. Through this process of trials and errors, the internal validity of the theories can be tested. And very often modelers can identify some new directions for empirical studies in order to answer problems arising from the failure of the models. Therefore the seeming "circularity" may not be a problem for modeling studies, as long as the conditions for such circularity are made explicit.

Having highlighted the advantages of computational modeling in studying language evolution, we note that modeling in this research area is still at its rudimentary stage, especially for the studies on language change and language origin. There is a lot of room for improvement, as there are many difficulties and problems for modelers to address.

One of the problems for modeling studies in language evolution is that most modelers do not start from linguistics, and most of the time the modelers are unfamiliar with the intricate characteristics and complexity of language, especially when going into higher level of grammatical structures. As suggested by Holland, one of the pioneers in computer modeling for adaptation, modelers have to "become intimately familiar with the problem (no 'I could look that up'), learn the related problems and the tricks and the oral tradition which go with them, be on the lookout for analogies and exploit them - but ultimately it comes down to trial and error, and so to luck" (Holland, 1998).

Furthermore, similar to modeling in any areas, computational models can only be very much idealized and seem far away from reality. There exists a dilemma, which is how to achieve the balance of simplicity and reality. On the one hand, if a model is too simplified without retaining the essence of the problem, it is just a self-contained toy model which has little contribution to give to real problems. As de Boer (2001) argues, "computer models should stay as close as possible to actual human linguistic behavior" (p122). He considers that the experiments on phonology and phonetics should be less controversial than those on syntax or semantics. We have reached a considerably good understanding in the physiological basis of language, which enables us to construct models with low-level representation and processing mechanisms (e.g., Au and Coupé, 2003; Oudeyer, 2002). Syntax and semantics are far more complex, and unlike phonetics, we know much less about the low-level representations in these two areas. Most of the studies on modeling the evolution of syntax have been based on some high level representations, such as the principle-and-parameter framework (e.g., Briscoe, 2000b; Niyogi and Berwick, 1997), although this theory is still very controversial in its plausibility.

On the other hand, if the models include too much complexity, they will face the same difficulties as in empirical studies, as there will be too many interwoven parameters whose individual effects will be hard to sort out. How to identify the most significant factors of the complex situation is the crucial challenge for modelers, and it requires them to have in-depth knowledge of the problem under investigation. This is often the weak aspect of existing modeling studies, as the constructed models are often too abstract or idealized and have little relevance to real languages.

Most of the models only focus on individual subsystems of language, such as phonology, lexicon or syntax. However, these subsystems are dependent from each other, and each part exerts pressure on the others. Therefore, "modeling any part of language in isolation is in fact a large simplification. Whenever one wants to build a model of the whole language, all aspects of language have to be taken into account" (de Boer, 2001, p123). So far for models which adopt an emergent perspective, only simple forms of language, such as simple syntactic word order rules (Kirby, 1999b) have been simulated. There have been few models to address the emergence of complex morphological structures, or the dynamical process of grammaticalization.

As a final comment, we would like to highlight the central role of words or lexicon in modeling language evolution. Contrary to the generative linguistic tradition which is often too syntactocentric, more and more students of language evolution agree that words are of more importance, either in language acquisition (Clark, 1993), language change (Wang, 1977), or language emergence (Jackendoff, 2002). Changes in phonology or syntax are always implemented through changes of words. The computational models of language evolution should also take into account this aspect.

Chapter 5

Some simulation models of language orgin and change

In the last chapter, I have reviewed briefly the application of computational modeling in the study of language evolution. In this chapter, I will present some simulation models which have been developed to address some issues related to the empirical studies discussed in early chapters on homophony and sound change. There are mainly two sets of models grouped according to the time scale they address, one dealing with language origin, i.e., the phylogenetic emergence of language, and the other with language change.

These models all adopt the perspective of self-organization, which emphasizes the importance of interactions among agents in addition to the agents themselves. This is different from some of our early modeling studies which consider abstract language system only. For example, in a study of explaining some universal structures of vowel and tone systems (Ke et al., 2003), we used an optimization model using Genetic Algorithms, assuming the vowel or tone system evolves as if to achieve an optimal state governed by a set of articulatory and/or perceptual constraints. These models lack realistic basis of the existence of a language or a subsystem of a language. Upon realizing the limitation of this type of models, we turn to models in which the evolution of language is situated with its agents and the interaction between the agents.

5.1 Some models of language origin

In Chapter 4, we have reviewed several models which postulate the very early stage of language phylogeny. A shared consistent vocabulary is often considered as the first step of language origin. Several simulation models have been developed to show how such a vocabulary may emerge through self-organization within the agents and within the population. This section will introduce these models briefly. We also develop a model to demonstrate that homophones can persist without causing communication problem because ambiguities can be solved with the help of contexts, as discussed in Chapter 2.

5.1.1 The emergence of vocabulary

The earliest form of human language must be communication systems consisting of a number of holistic signals like those found in primates and other animals such as bees and birds (Hauser, 1996). The signals from animal calls are not equivalent to the "names" or "words" in human language, since these signals are holistic (non-decomposable) and each specific to a situation. A food call is used when food is discovered, and a snake alarm call only report the sighting of a snake (Hauser, 1996). At some later stage, such signals were used as symbols to refer objects out of sight, and used in a non-situation-specific fashion (Jackendoff, 1999). Similar to what we observe in children's use of word, "kitty" may be uttered by a baby to draw attention to a cat, to inquire about the whereabouts of the cat, to summon the cat, to remark that something resembles a cat, and so forth (ibid). The realization of signals as symbols representing objects or events is referred as the "naming insight" (McShane, 1980). In the language development of a normal child, the naming insight comes so naturally that most parents do not notice the exact moment of this event without special attention. On the contrary, chimpanzees need intensive teaching to learn to name (Savage-Rumbaugh and Lewin, 1994).

There is also a striking difference between the signals used by animals and words by humans: while the repertoire of the holistic signals for animals usually is no more than several dozen, be they vocal calls, facial expressions or body gestures (Hauser, 1996; Wilson, 1972), the reading vocabulary of the average high school graduate is about 40,000 words (Miller and Gildea, 1987). The ability to manipulate a large number of symbols for communication is considered as a special trait in humans (Deacon, 1997; Hauser et al., 2002b).

How the naming insight occurred in the phylogeny of the hominid line is still a mystery. It could be due to some genetic mutation. Here we take this stage as the starting point. What interests us here is that once the naming insight is present in human species, each agent might have his own way to name in an arbitrary manner, how did the actual naming of objects or events become consistent across the entire population? While animals' communicative signals have been considered as being imprinted innately or learned by instinct (Gould and Marler, 1987), the words used in a linguistic community are mostly established by convention. Over twenty centuries ago, two great philosophers, continents apart, arrived at a similar observation that names are mostly arbitrary and established by convention (Wang, 1989a). In Greece, Plato (427-347 BCE) wrote that "any name which you give is the right one, and if you change that and give another, the new name is as correct as the old" (translation by Jowett (1953), quoted in Wang (1989a)). At about the same time, Xunzi (around 310 BCE) in China taught that '名无固宜' ("words have no intrinsic correctness") and '名无固实' ("words have no intrinsic content") (quoted in Wang, 1989a).

We have designed several models to simulate the process of conventionalization leading to emergence of a shared vocabulary (Ke et al., 2002). In these models, we have made a number of assumptions which are plausible for early humans. First, the agents are assumed to have already possessed the ability of naming, or, more generally, are able to use symbolic signs. Second, there exists a set of objects or meanings that are particularly salient and agents need to use them frequently in their daily life. Third, the agents are all able to produce the same set of utterances[1]. The agents intentionally interact with each other to communicate using these utterances to convey meanings. The mappings between meanings and utterances can be represented in various ways,

[1]Here "utterance", "signal" and "word" are sometimes used interchangeably for the sake of convenience, particularly when discussing other studies, though we acknowledge that there are important differences among them in some context.

for example, by a look-up table (Steels, 1996b), an association matrix (Hurford, 1989; Nowak et al., 1999; Oliphant, 1997), or a neural network (Cangelosi and Parisi, 1998). The models reported here adopt two different forms, following earlier studies, including look-up tables and probabilistic association matrices.

The emergence of a shared vocabulary refers to a stage in which agents have the same set of mappings between meanings and utterances, for both speaking and listening. The question then is how these mappings are formed and how members of a population reach the same set of mappings. The answers to these questions lie in the form of interaction among agents during communication.

5.1.1.1 The imitation model

Imitation is one often-observed interaction scenario. The strong ability of humans to imitate, even from early infancy, has been extensively documented (Meltzoff, 1996). While other social animals, particularly the primates, also imitate (Dugatkin, 2000), it is argued that a strong imitation capacity may have only evolved in hominid line as one of the important stages leading to human language (Arbib, in press). The imitation in humans refers to "the ability to recognize another's performance as a set of familiar actions and then repeat them, or to recognize that such a performance combines novel actions which can be approximated by variants of actions already in the repertoire". We assume that imitation serve as the most explanatory mechanism for the formation of a common vocabulary. Before establishing a consistent way of naming things, early humans very likely made use of their propensity for imitation; the younger ones imitating their elders, the followers imitating the leaders or, just by chance, their neighbors.

In the simulation model, we assume that there are M meanings for a group of agents P to communicate, using U different utterances. Each agent's vocabulary consists of a set of mappings, or one-to-one mappings, between meanings and utterances. Each agent can create and change his own vocabulary by imitation, similar to the model proposed by Steels (1997). Table 5.1 shows two agents' initial vocabulary in the upper part of the table.

When two agents interact, one imitates the other according to some strategies, such as by random, or by following the majority in the population. An example of the imitation result is given in the lower part of Table 5.1. We use a Markov chain model to prove that the agents in the population always converge to a consistent vocabulary under the above simple imitation conditions (Ke et al., 2002). In this model, we highly simplify the conditions for imitation, such as assuming a perfect meaning transfer as discussed in the last chapter, and neglecting various social factors which may affect interactions. The effect of social factors will be examined in the model of language change reported in later sections of this chapter.

Table 5.1: The initial and final utterance-meaning mappings, Ψ, of two agents.

Initial	m_1	m_2	m_3	m_4	$\ldots$	m_M
$\Psi(A)$	u_2	u_4	u_1	u_4	$\ldots$	u_5
$\Psi(B)$	u_5	u_3	u_1	u_1	$\ldots$	u_3
Final	m_1	m_2	m_3	m_4	$\ldots$	m_M
$\Psi(A)$	u_2	u_3	u_1	u_4	$\ldots$	u_3
$\Psi(B)$	u_2	u_3	u_1	u_4	$\ldots$	u_3

5.1.1.2 The self-organization model

In the above imitation model, the vocabulary is simulated as one set of mappings for the mappings between meanings and utterances, without distinguishing speaking and listening. As we have discussed in Chapter 4, it is important and necessary to make the distinction between speaking and listening behaviors. In the following, we will present a model which simulates the active and passive vocabulary as two separate sets of mappings in the form of two association matrices. An example of the two matrices each with three meanings and three utterances is given in Table 5.2. Each element in the matrices represents the probability for the mapping between a certain meaning m_i and a certain utterance u_j. The two matrices are stochastic matrices, having the constraint that each row of the speaking matrix and each column of the listening matrix sum to one, i.e., $\sum_i p_{ij} = 1$ and $\sum_j q_{ij} = 1$, to meet the assumption that each meaning is expressible, and each utterance is interpretable.

In the early imitation model, the mappings between meanings and utterances are

Table 5.2: An example of the speaking and listening matrices of one agent in the self-organization model.

p_{ij}	u_1	u_2	u_3		q_{ij}	u_1	u_2	u_3
m_1	0.3	0.4	0.3		m_1	0.1	0.3	0.6
m_2	0.4	0.55	0.05		m_2	0.5	0.3	0.3
m_3	0.7	0.2	0.1		m_3	0.4	0.4	0.1

binary, an mapping being either present or absent. When an imitation event happens, the agent copies perfectly the other's meaning-utterance mapping, regardless of his original one. The assumptions in a scenario, such as that agents always get others' intended meaning and want to change their own readily, are too much idealized. To make the model a little bit more realistic, we change the interaction from a deterministic nature to a probabilistic one, i.e., the agents apply probabilistic changes to the mappings based on the success of each interaction.

At the beginning of the simulation, the speaking and listening matrices of each agent are both randomly initialized. When two agents interact, a successful interaction occurs when the listener interprets the received utterance as the meaning intended by the speaker, resulting in a reinforcement of the mapping used in the speaker's speaking matrix and the listener's listening matrix. Conversely, if the listener interprets a meaning that differs from the one intended by the speaker, such a failed interaction will lead to weakening of the corresponding mappings used by the two agents. Such a process of updating the internal mappings is one way of self-organization within the agents. Also through the iterative interactions, the population undergoes a self-organization process.

The simulation proceeds as follows. At each time instant, two agents are randomly selected, one as speaker and one as listener. The speaker updates his speaking matrix and the listener updates his listening matrix after the interaction based on the success or failure of the communication. We implement a special type of interaction in the model, that is, "self-talk": an agent can be both speaker and listener and he talks to himself. The simulation stops when a given number of interactions has completed, or a consistent vocabulary has emerged.

The model shows two convergent phenomena: 1) intra-agent convergence: the speaking and listening matrices of each agent become compatible: in the case when $U = M$, the two matrices are identical, as shown in Table 5.3; when $U > M$, agents' speaking matrix stabilizes as a subset of the listening matrix, as shown in Table 5.4; 2) inter-agent convergence: agents share the same vocabulary, i.e., the same speaking matrix and the same listening matrix. Table 5.3 only shows one of the possible convergent states. In fact there are in total six convergent states in the case of $U = M = 3$. And the experiments show that these six states occur in different runs, and their frequency is approximately the same. This is the "multistability" characteristic of self-organizing systems mentioned in Chapter 1.

Table 5.3: An example of a stable state: All agents have the same speaking and listening matrices.

p_{ij}	u_1	u_2	u_3
m_1	1	0	0
m_2	0	1	0
m_3	0	0	1

q_{ij}	u_1	u_2	u_3
m_1	1	0	0
m_2	0	1	0
m_3	0	0	1

Table 5.4: An example of convergent vocabulary when $U > M$: Two types of resultant speaking and listening matrices, speaking being a subset of listening vocabulary.

p^1_{ij}	u_1	u_2	u_3	u_4
m_1	1	0	0	0
m_2	0	1	0	0
m_3	0	0	1	0

p^2_{ij}	u_1	u_2	u_3	u_4
m_1	0	0	0	1
m_2	0	1	0	0
m_3	0	0	1	0

q^1_{ij}	u_1	u_2	u_3	u_4
m_1	1	0	0	1
m_2	0	1	0	0
m_3	0	0	1	0

q^2_{ij}	u_1	u_2	u_3	u_4
m_1	1	0	0	1
m_2	0	1	0	0
m_3	0	0	1	0

The two types of convergence are both emergent phenomena. There is no explicit or obligatory mechanism forcing the speaking and listening matrices to be compatible, as the two matrices in each agent do not influence each other, but are updated through the interactions with other agents. This way of updating the matrices is different from the strategies proposed by Hurford (1989), as discussed in Chapter 4. The three strategies, i.e., the imitator, calculator, and Saussurean, seem implausible when we attempt to dissect the actual interaction situation, which should go through such the following process:

$$\text{(speaker) } m_s \rightarrow u_s \Longrightarrow \text{(listener) } (u_s \rightarrow) u_l \rightarrow m_l$$

The listening behavior is internal to each agent, and one cannot access another agent's listening matrix. It is unjustified that the learner[2] can always "imitate" the listening behavior of the teacher's directly by copying his listening matrix. So Hurford's "imitator" strategy is implausible. Similarly, the "calculator" strategy also seems implausible, because the learner should not be able to copy the teacher's listening behavior for his own speaking. The "Saussurean" learner is assumed to construct his speaking matrix by learning the teacher's speaking behavior; however, in order to do so, the learner has to construct his listening matrix in order to understand the teacher's speaking before he himself learns how to speak. Therefore, the Saussurean suggested by Hurford seems unrealistic either. The "obverter" strategy implemented in Batali (1998) conforms better to the real interaction situation, provided the assumption that when an agent speaks, he actually talks to himself first, and decides his speaking behavior based on the understanding of his own speech. This assumption is more plausible, as empirical studies have shown that there exists auditory feedback (Houde and Jordan, 1998) and even somatosensory feedback during speech (Tremblay et al., 2003), for speakers to make corrections and compensations in achieving speech targets. If such internal feedback to adjust speaking exists at the lowest level of speech production, it should be expected that a feedback also exists at a high level of language production, as observed from studies of self-repair during conversation (Levelt, 1989). However, the existence of such feedback should not be interpreted as there is a mechanism in making speaking and listening identical instantaneously. The consistency between the two behaviors should be the product of a period of modification during learning. Therefore, we maintain that it is more realistic for the models to keep speaking and listening matrices separate and simulate the actual communication process, while adding the self-talk in the model to simulate the internal feedback.

[2]Here we adopt the term "learner" and "teacher", instead of "speaker" and "listener" following Hurford's work, as the scenario is a teacher-learner uni-directional interaction, and both speaking and listening matrices of the two agents may be involved.

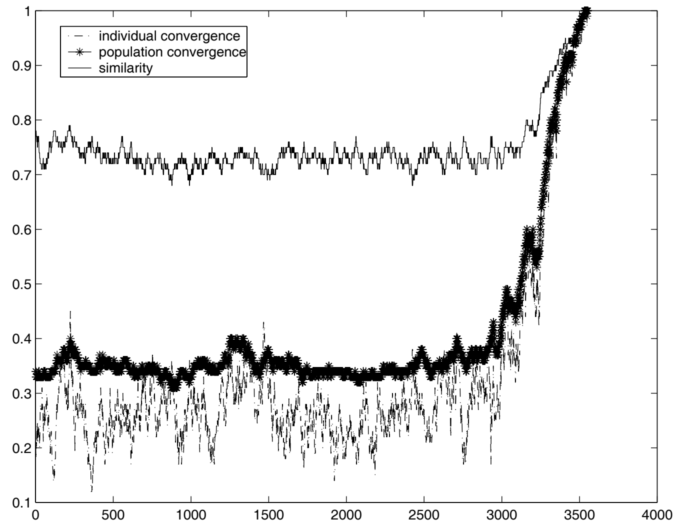


Figure 5.1: The convergent trends from a typical run of simulation of the interaction model (P_s=10, $M = U = 3$, $\Delta = 0.2$). Three measures of the convergence (SI, PC and IC) are shown. A consistent vocabulary emerges after 3553 interactions. An abrupt phase transition can be observed around 3000th interaction.

We design three measures to monitor the process of development: 1) similarity (SI, the average similarity of the two matrices between all pairs of agents), 2) population consistency (PC, the average understanding rate between all possible pairs of agents) and 3) individual consistency (IC, the average consistency between agent's speaking and listening matrices over all agents). Figure 5.1 shows the curves of three measures from a typical run of simulation.

In this run, the population consists of 10 agents. Each agent starts with a vocabulary which consists of a set of utterance-meaning mappings with randomly-initialized probabilities. The convergence is not gradual but rather abruptly arises after about 3000 interactions, exhibiting a "phase transition". For a long period of time, the interactions among agents only result in fluctuation, and there is little consistency in the agents' vocabularies. However, at some instant, there is an abrupt rise of the consistency, and the population converges quickly after that period. The conditions of the interactions have not changed at all in the process toward convergence. The abrupt emergence of order in the population is the result of a sequence of interactions which by accident lead the agents to change in the same direction, thus bringing about a momentum toward convergence.

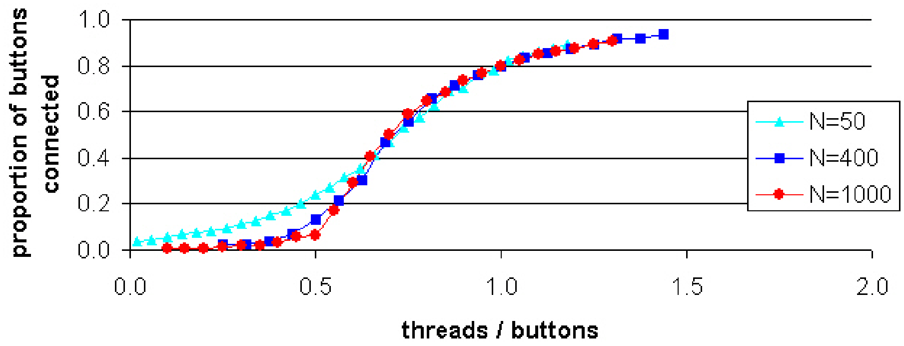


Figure 5.2: An example of phase transition in the button-and-thread model. The number of connected buttons increases sharply when the number of threads reaches a certain value. The transition is more obvious when the number of buttons is larger.

5.1.1.3 Phase transition

The phase transition observed from the simulation is a phenomenon often attested in self-organizing systems. The spontaneous emergence of structure at the global level often goes through a phase transition. A common example of phase transition is the abrupt transition from ice to water, and from water to steam. The heat is gradually added by a constant amount, but the state of the system undergoes a nonlinear and qualitative change, instead of a linear quantitative change. Kauffman (1993) gives an illustrative button-and-thread model, which is modified from a classic paper in random graph theory (Erdös and Rènyi, 1959), to show how phase transition may occur: given a pool of N buttons, each time randomly pick up two buttons and connect them with a thread and put them back to the pool; repeat this for L steps and monitor the number of the connected buttons (C). It is found that initially C increases slowly, but when L reaches a certain value (in the limit of N approaching to infinity, the value of L/N for the transition is 0.5), C increases sharply and a large proportion of buttons are connected. The larger the pool of buttons is, the sharper the transition appears, which can be seen by comparing the three conditions of N as shown in Figure 5.2. When N=1000, the sharp increase appears around L/N=0.5.

Wang (1999) and Wang et al. (2004) hypothesize several phase transitions during language evolution, "as in the emergence of segmental phonology, the invention of morphology and syntax, and the use of recursion in sentence construction, etc." (Wang et al., 2004)[3]. It is hardly possible for us to obtain any direct evidence for these phase transitions as language leaves no fossil until the emergence of writing. But we may infer these transitions from some indirect evidence, assuming that the development of language should have brought about changes in culture as fostered by the improvement of communication, even without any significant changes in the biological basis. The "cultural explosion" reported by anthropologists as mentioned in Chapter 1 may serve as one piece of indirect evidence, though it is unknown yet what levels of complexity the language was at that time period after the transition: is it a fully-fledged human language, or is it the pre-language with a limit set of symbols?

While empirically we have no way to identify or reconstruct the phase transitions in the past, computer models may serve as a useful tool in examining the conditions and situations of these phase transitions. Our simulation model has demonstrated an example of phase transition during the emergence of vocabulary in a highly simplified hypothetical scenario. Similar phase transitions have been demonstrated in some other computational models of language phylogenetic emergence.

For example, Kirby (2000) simulates the emergence of a compositional language from a holistic communication system. In his model, at the beginning, agents all use holistic signals to communicate. It is assumed that agents are able to detect recurrent patterns in these signals and extract these recurrent subparts as words. They will combine words to form composite meanings later, recruiting the general capacity of concatenating related entities in a sequence. Thus compositionality in a communication system emerges. This scenario has been proposed by Wray (1998, 2000) from an empirical perspective. It is different from Bickerton (1990) who hypothesizes that human language may have started from words as the early stage in pre-language, and later words are combined into phrases according to some innately given word orders, which was the result of abrupt mutation specific to language.

[3]We may need refine these proposed transitions, as the last three transitions, i.e., invention of morphology and syntax, and use of recursion may not fall in the right order. First, the definition of morphology and syntax need specified. Simple word order should emerge much earlier than inflection morphology. Recursion may be a general cognitive capacity in humans but exapted for language later. There are other versions in hypothesizing the possible phase transitions during language evolution. Here we combine Wang et al. (2004) and Jackendoff (1999) to suggest the following ordered transitions: symbolization → segmental phonology → compositionality (simple word order and recursion) → hierarchical structure (with embedding).

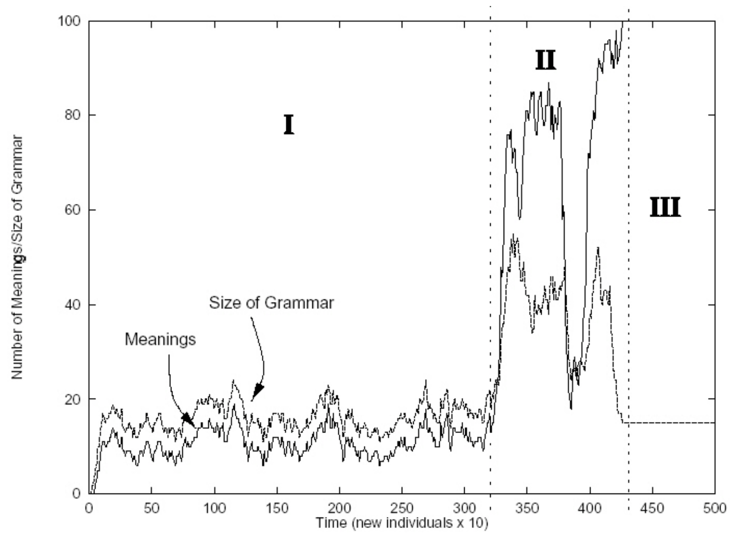


Figure 5.3: Three phase transitions in the emergence of a compositional language in a population, reproduced from Kirby (2000), Figure 18.3.

Kirby's model shows that the agents' expressitivity and the number of grammars (in terms of how many lexical categories and word order rules) undergo three emergent stages, as shown in Figure 5.3. At the first stage the agents can only express a small percentage of meanings, using a small number of grammars which basically consists of holistic utterances. Later there is a sudden increase in expressitivity and grammar size, owing to the emergence of a number of words, some lexical categories and partial word order rules. Shortly after the second stage, the population undergoes the third abrupt change and stabilizes afterwards. The agents in the population can express all the meanings by compositional rules, and therefore the number of grammar rules drops sharply compared to that of the second stage.

Gong et al. (2004) demonstrate a similar phenomenon of phase transition during the process of the co-evolution of lexicon and syntax. This model is a modification of Kirby (1999b, 2000). It removes the unrealistic meaning transference in Kirby's model, replaces Kirby's highly simplified teacher-learner pair-wise interaction with more realistic multi-agent interactions and considers various social structures, and allows competition among rules within agent's internal language system. These models with more and more refined realistic implementations suggest that it is highly possible that the observed phase transitions have occurred in the actual process of language phylogenetic emergence.

Unlike language emergence which occurred too far back in time, we may identify some evidence for such type of phase transitions in language change from the recent histories of languages. Languages are incessantly changing due to their heterogeneous nature at the communal level and almost unavoidably contact with neighboring languages[4]. Most of the time such changes may appear as gradual and can only be detected retrospectively when the changes are complete. However, in some cases, dramatic changes happened and can be scrutinized by linguists. For example, the transition from Old English to Middle English within a few hundred years starting from the early tenth century has brought to English dramatic changes in phonology, lexicon (around 50% of words of Germanic origin in the vocabulary have been replaced by Romance words (Trask, 1996)), and grammar (from case marking systems to fixed word order). Dixon (1997) suggests that punctuated divergence might happen in the history of Australian languages. Such phase transitions may be attested time and again in many languages which have been in heavy language contact with other languages. The input of heavy contact does not need to last long. After the external influence is injected, the language may respond through a self-organization process, including large-scale changes in adults and positive feedback from children's learning, and thus result in a phase transition in typological changes.

5.1.2 The evolution of homonyms

In the above models for the development of a shared vocabulary, it is assumed that there are a fixed set of meanings and utterances, and the internal relationship between utterances and meanings has not been taken into account. In this section, I will report another model which is specially designed to examine the emergence and persistence of homonyms[5] during the process of language evolution. The simulation model[6] is designed based on the "naming games" framework (Steels, 1996b, 1997), which seems

[4] Different languages may experience different degrees of influence from other languages. For example, some languages which are used in isolated areas or closed communities may have little contact with other language (Icelandic is such a case), and some languages with high prestige compared to neighbors are less likely to be influenced by contact.

[5] In Chapter 2 I distinguish homonymy from homophony (the former is a subset of the latter) due to the need of analyses. In this chapter I will narrow down to "homonymy", for the convenience of exposition, especially, to be parallel with the discussion of "synonymy".

to be in the middle ground between the discrete imitation model and self-organizing probabilistic models discussed above.

Agents in the model are assumed to be able to produce a number of distinctive utterances (U) and to make use of such utterances to name a set of meanings (M). At the beginning, the agents do not have any words, but they can create new words at random, as well as learn the words created by other agents through naming games. Each word (an meaning-name mapping) has a score in the agents' vocabulary and the scores are updated after interactions.

Naming games are interactions between two agents, a speaker and a listener. In one game, two agents are chosen; the speaker decides a meaning he wants to communicate, looks for or creates an utterance which is associated with the meaning, and transmits the utterance to the listener. The listener perceives the utterance and tries to interpret the meaning by searching his existing vocabulary. If the listener interprets the same meaning for the utterance, then this is considered to be a successful game, and therefore the score of the word is increased. Otherwise, the score is decreased. When the score of the word becomes too small, the word is removed from the vocabulary. Upon failure, the listener learns the word from the speaker by adding a mapping between the perceived name and the meaning referred to by the speaker. This process of interaction is very similar to our self-organizing model introduced above, except that in the early model the updating of a particular meaning-utterance mapping affects other mappings while in this model it is not so.

With this model, we compare the situation for different ratios between the number of meanings (M) and the number of utterances (U). Figure 5.4 shows simulation results for two different ratios. When $M : U = 1$, we can see that agents are able to acquire the same vocabulary, and their communications are successful 90% of the time, 20% of the words having homonyms. When we increase the number of meanings that are to be communicated by agents, for example, setting $M : U = 3$, the vocabularies of the agents no longer converge to achieve an effective communication - every word has at least one homonym - resulting in a low rate of communicative success (only about

[6] This model is a collaborated work with Dr. Christophe Coupé in Laboratoire Dynamique de Langage, Institut des Sciences de l'Homme, Lyon, France.

30%).

This condition, $M > U$, simulates a more realistic situation in which our lexical need far exceeds the number of forms we would like to utilize[7]. Cheng (1998) has shown that there exists a general limit to the size of the active vocabulary used by various writers[8]. If we assume that the number of meanings that humans want to manipulate is infinite, the limit on the vocabulary size suggests that there may exist a cognitive constraint on the number of forms which can be memorized and used effectively [9]. If it is true that there is only a limited number of forms for the making of words, the condition of $M > U$ then is not only realistic, but necessary as a condition for the models.

To meet the lexical need, it is obvious that the existence of homonymy is inevitable under this condition. However, in spite of the considerable ambiguity implied by homonyms, our daily communication does not seem to be much hampered by it, contradictory to what the above model shows. What are the explanations for the effective communication under the condition that $M > U$ which the current model seems unable to demonstrate.

In the above simulation, only one meaning is transmitted during each communication event. In a real situation, most of the time, we communicate with a phrase or a sentence. Words in the phrase or sentence most of the time are semantically related. To simulate this situation, we have designed a two-word communication model. In a communication event, the speaker chooses two meanings (m_1 and m_2) which are close to each other in the semantic space, and produces two utterances (u_1 and u_2) to communicate with another agent. The listener receives these two utterances. If u_1 has only one meaning, m_1, and u_2 has two meanings, say m_2 and m_3, then the listener

[7]Recall the discussion in Chapter 2 that the potential usable forms are actually not limited, provided that forms can be of infinite length. However, it has been shown that the actual exploitation of this potential is very limited, in terms of the used combinations of segments, the uneven distribution of the syllables, the preference for shorter words, etc.

[8]Note that the vocabulary here refers to the set of word forms, disregarding the presence of homonyms and polysemes.

[9]There are arguments against this hypothesis of the limited size of memory for words. For example, as bilingual people appear to have words twice as that of the monolinguals. However, we are not clear yet whether the bilingual lexicon requires twice the memory as the monolingual. A more reasonable argument for the limited vocabulary size in normal people may be that as long as the vocabulary can serve the need, people will not bother to learn more. And the decay of memory due to lack of activation may be another factor for the limited vocabulary size. Words may fade away if not used frequently enough.

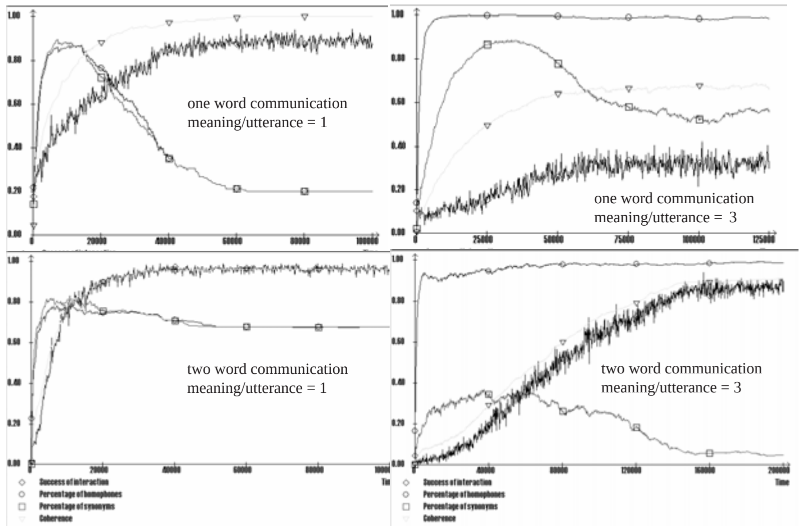


Figure 5.4: Homonym evolution under four conditions (upper left: M = U, one-word communication; upper right: M = 3U, one-word communication; lower left: M = U, two-word communication; lower right: M = 3U, two-word communication. (The thick solid line: success of communication; line with circle mark: percentage of words having homophones.)

will choose between m_2 and m_3 the one which is closer to m_1 in the semantic space. If neither u_1 nor u_2 has a unique meaning, the same principle of disambiguation can be applied: the listener will choose the pair of meanings which are the closest in the semantic space.

In this formulation, though the semantic proximity helps to disambiguate homophonous utterances, it is not trivial for the agents to converge to the same set of mappings. Nevertheless, we observe a gradual increase in the rate of communicative success through successive interactions. When $M = U$, we see that the communicative success reaches 100%, much better than the earlier case of one word communication, even with a degree of homonymy as high as 70%. When we increase the lexical demand, we can see a much clearer improvement of the system owing to the two-word communication scenario. Homonymy can be tolerated up to 100%, while the rate of communicative success still rises to more than 80%. This simulation demonstrates clearly that, with the help of context (the two words serve as context for each other), the vocabulary can tolerate a high degree of homophony, even when the number of meanings greatly exceeds that of utterances.

This simulation model illustrates that homonymy can persist in the vocabulary and still maintain a high communication effectiveness, given a realistic condition of contextual communication. As mentioned earlier, this model is undertaken from the perspective of emergence, i.e., vocabulary starting from scratch. However, as discussed in Chapter 2, we know that homonyms constantly emerge as the result of sound merger from language change. Also pairs of homonyms exhibit various differentiation characteristics such as in frequency, in part of speech, etc., exhibiting self-organization in the language system. It will be interesting for future modeling studies to simulate how such self-organization takes place regarding homonymy.

5.2 Simulation models of language change

The above models mostly focus on how language would have emerged from scratch. This line of modeling is mostly centered on making various necessary and realistic assumptions on the agents, and testing the consequences of these assumptions on language emergence. In this section I will present the other line of modeling for language evolution, which is to study how an existing language changes in the diachronic dimension. In the former line of models of emergence, we can also study language change by tracing the rise and fall of the various elements in the emerging system, such as synonyms and homonyms. However, the first line of models focuses more on the properties of the agents, or in other words, on the language-internal factors. In contrast, the latter line of models pays more attention to the language-external factors which affect the interactions and learning between agents, such as different social structures, age distribution, the intensity of interactions, the degree of conventionalization induced by education, etc.

These two lines of models exemplify well the two basic components of a self-organization framework in studying language evolution, i.e., the agents and the interaction between agents. Though all models have to incorporate these two components, there are different emphases and aims in the actual consideration and implementation. In addition to the differences mentioned above, the first line of models is more interested in the emergence of a shared language, while the other has more interest in the dynamics of the evolutionary process.

In the following, I will first present some discussions on the theoretical issues in viewing language change as a diffusion of linguistic innovations, such as the threshold problem. Various factors determining the diffusion process are examined, including the social structure and the characteristics of learning in agents. The features of the dynamics of language change, such as the stochasticity, the S-curve shape, etc. will be discussed.

5.2.1 Language change as a diffusion process

Language change can be viewed as a process of innovation diffusion (Croft, 2000; Shen, 1997). A new form of a certain linguistic function, be it a novel lexical item, or a new grammatical structure, always originates from only a number of individuals. Some innovations may be brand new expressions which do not have any competitor in the current language, though very often these first names do not survive and get replaced quickly by other "better" names invented later. Some innovations may have existing counterparts in the current language system, for example, before the word '手机' for "mobile phone" in Chinese emerged in Shanghai dialect, there were several existing terms such as '大哥大', '移动电话', etc. (Tsou and You, 2003). Speakers in the speech community adopt the linguistic innovation through interacting with and learning from the innovator(s). A language change is complete when the innovation spreads to the whole speech community[10].

There have been similar ideas of dividing the process of language change into two sub-processes, such as actuation and transmission discussed in Weinreich et al. (1968), actuation and implementation suggested in Chen and Wang (1975), or innovation and diffusion (or propagation) proposed by Croft (2000). Croft also terms these two processes as "altered replication" and "differential replication" following the analogy from biological evolution. He remarks that "both processes are necessary components

[10] Actually this is rarely the case. It is hard to show that the whole community is homogeneous at any stage. However, we make such a simplification for the sake of simplicity. It is less problematic if we only consider the homogeneity in some restricted aspects, for example, the pronunciation of a lexical item. When the change is complete, we can assume that the pronunciation of the lexical item is homogeneous in the speech community.

of the process of language change" but this fact is "rarely recognized in models of language change" (ibid, p5). Also, Croft suggests that "a comprehensive framework for understanding language change must subsume structural, functional and social dimensions of language change or their equivalents" (ibid). He further correlates these three dimensions with the two processes: "the functional factors - the phonetic and conceptual factors appealed to by functionalist linguists - are responsible only for innovation, and social factors provide a selection mechanism for propagation" (ibid, p38).

In this study, the proposed computational model will address some factors in the two processes and the three dimensions. In the model, the whole population uses the original form (denoted as the unchanged form, U) in the beginning. There are different possible sources of the innovation (denoted as the changed form, C), for example, different adults' random creation, or children's imperfect learning. The model will compare these different origins. For the three dimensions, the structural dimension is ignored, as only one linguistic item is considered in the model. Or the structural factor can be integrated with the functional factor, together determined by the functional bias of the linguistic item, which is an abstract value integrating cognitive and/or physiological biases toward learning the linguistic item. It is assumed that U and C both have their own functional biases, $f(U)$ and $f(C)$. If the ratio β between $f(C)$ and $f(U)$ is greater than one, i.e., $\beta = f(C)/f(U) > 1$, it means that form C has a certain functional advantage over U. The social factors include different social network structures. The model compares the diffusion dynamics under these different conditions.

5.2.2 The threshold problem of language change

Linguistic innovations arise frequently in the speech community. However, not all innovations are successful in leading to a change. There is a "threshold problem" for successful changes, which has been introduced in Section 4.1.2.

Nettle's model requires the existence of super-influential agents for the change to overcome the threshold. This condition for change is too stringent and it may find

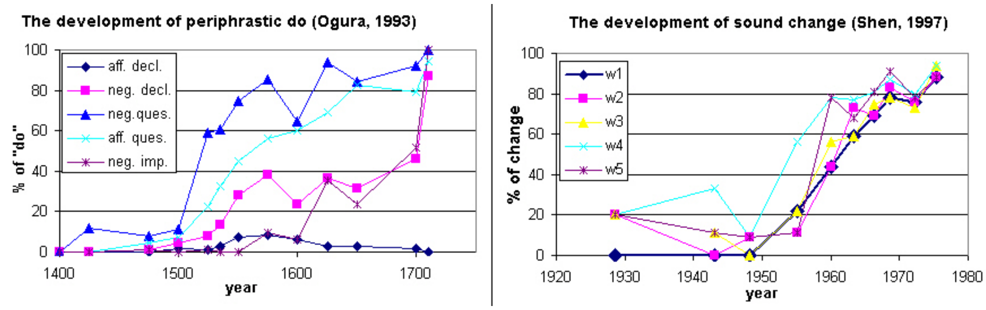


Figure 5.5: Examples of S-curve from two empirical studies. (a) development of the periphrastic *do*, reproduced from Ogura (1993); (b) diffusion in the apparent time of a sound change in Shanghai, produced using data from Shen (1997).

difficulty in explaining changes from below, where the change is often found, not from the highest social class, but from upper working class or lower middle class (Labov, 2001, p31). The models to be reported in the following will consider various types of social structures which are different from Nettle's regular network and add some realistic considerations in learning strategies. We want to demonstrate that there is still a high chance for changes to happen without requiring innovation from agents with specially high social status.

5.2.3 The S-curve dynamical pattern of language change

In addition to the threshold problem in language change, the dynamics of language change is another interesting issue. Traditionally the dynamics of a language change is idealized as an S-curve pattern (e.g., Altmann, 1983; Bailey, 1973; Kroch, 1989; Ogura and Wang, 1996; Weinreich et al., 1968), i.e., a gradual slow-fast-slow dynamics. Many studies have shown that the empirical data can be fitted with an idealized S-curve, for example, sound changes (Shen, 1997), syntactic changes (Kroch, 1989; Ogura, 1993), though there are irregular patterns with many ripples. Figure 5.5 reproduces the change dynamics reported in Shen (1997) and Ogura (1993).

Various mathematical models have been proposed to reproduce or fit the observed S-curve diffusion dynamics, such as the cumulative frequencies of the binomial distribution (Labov, 1994), logistic function (Shen, 1997). Cavalli-Sforza and Feldman

(1981) give three types of models, all of which can derive a dynamics following a logistic function. Equation 5.1 gives the one which is used to model an epidemic, i.e., the spread of an infectious disease within a generation. It is taken by Shen (1997) to represent a linguistic diffusion process. The rate of increase of the number of people adopting C is dependent on the number of people adopting C and that adopting U in the population, and the rate of effective contact, α, which indicates the probability for an individual to change.

$$\frac{dC}{dt} = \alpha C(t)U(t) \tag{5.1}$$

Several computational models, mostly concerned about grammatical changes, have been proposed to simulate the diffusion process. Niyogi and Berwick (1997) use a deterministic learning model based on the parameter-setting framework to analytically derive a logistic S-shaped spread dynamics. Briscoe (2000a) points out some of the unrealistic assumptions in N&B model, such as the infinite and non-overlapping population. He also criticizes that the Niyogi and Berwick model actually examines the gross statistical behavior of learners, instead of behaviors of individual learners. Using a stochastic learning model with more realistic demographic assumptions, he shows that language change can proceed in S-curve only under certain conditions.

Some reservations and criticisms of the S-curve have been raised in the literature. Croft (2000) points out the difficulty in determining the beginning and end point of the diffusion (p185). Denison (2003) mentions the necessity to clarify the variable of being measured for an S-curve, whether it is the percentage of speakers in a community who use the innovative form, or it is the percentage of words which adopt the changed form. Also there are often multiple competitors or continuous variants, rather than discrete and binary changes. While these criticisms present serious challenges for empirical studies in their description of the S-curve pattern in language change, it is relatively easy for computational models to handle these problems, as all variables and parameters can be under rigorous control.

Cavalli-Sforza and Feldman (1981) point out some more problems in the assumptions of the logistic model to study innovation diffusion. For instance, it is assumed that the probability of adoption is proportional to the fraction of individuals who have already adopted the innovation, and the residual fraction who have not yet adopted it. But the increase of adoption may change the social environment and therefore put extra pressure to increase the probability of adoption. In other words, the probability of change, α, should be dynamic rather than fixed. Also, the adoption of the innovation is only due to the knowledge gained by observation or contact with people who have already adopted the innovation, but actually individuals could have their own change for adoption without the influence of the early adopters, which is similar to the imperfect learning in children. Third, it is assumed that the individuals are homogeneous with the same rate of change. However, it seems very likely that there is variation between individuals in their capacity in learning or adopting an innovation. Fourth, the population is assumed to be spatially homogeneous, i.e., perfect mixing. However, contacts are ordinarily more probable between individuals closer in space (ibid, p34-38).

The above criticisms reveal some common problems in analytical models. For example, they often have to make unrealistic assumptions, and lack of flexibility in incorporating realistic conditions. Simulation models usually can deal with these problems effectively. Some conclusions from the analytical models based on the unrealistic assumptions may become invalid and need revision, as illustrated in the following where different types of social networks are compared.

5.2.4 Importance of social networks

In most of the models of language change or language emergence, the social structure of the population is often assumed as a regular network: the learner only learns from one or several teachers next to him in a ring-like population structure (Kirby, 1999b, 2000; Tonkes, 2001), or from the whole population but with weighted connections (Nettle, 1999c). However, it has been shown that the linguistic environment for language learning is larger than the domain of the family, and involves peers and public media, etc. (Weinreich et al., 1968). It is found that not only caretakers' language input to children is important (Snow and Ferguson, 1977), but also the language data the children observed from the interactions between adults (Blum-Kulka and Snow, 2002).

The characteristics of social networks should play an important role in affecting children's language learning. In the past, social networks analyzed by linguists are usually of small size (Labov, 2001, p328). For example, in a classic sociolinguistic study by Milroy (1980/1987) on three lower working-class communities in Belfast, Ireland, the social networks in these communities examined all include less than 20 people. In Eckert (2000)'s study on social identities in a high school in Detroit, only 69 speakers were examined though several hundred students were involved in the social network. Compared to the size of the usual communities, these numbers are very small. Data on linguistic interaction networks of larger groups seem still be lacking empirically. The long term effect of different social networks on language change is even further beyond the reach of empirical studies done so far.

In recent years, studies on networks in general have seen a surge of new development in a broad range of areas. Owing to the availability of powerful computational equipment and techniques, researchers are able to investigate huge complex networks in the real world, such as internet, world wide web, scientists and actors collaboration networks, foodwebs, airline networks, and so on. While early studies mostly focus on the analyses of structures of small networks, properties of individual nodes or subnets of networks, recent studies are more interested in examining the global statistical properties of large-scale networks. Several popular books published recently for general audiences provide a good coverage of the background and development of the field (Barabási, 2002; Buchanan, 2002; Watts, 1999, 2003) (see a review of the first three books by Frommer and Pundoor, 2003). Extensive review of the technical aspects of complex network studies can be found in Albert and Barabási (2002) and Newman (2003).

Previously, random networks have long been taken as the model of complex networks in the real world, as it is obvious that regular networks are rare. Extensive mathematical analyses on the properties of random networks have been done since the pioneering works by Erdös and Rènyi (1959). However, recent empirical studies have found that real-world networks exhibit features which cannot be captured by random networks. The rapid growth of the area has mainly been triggered by the discovery of two new types of network, i.e., small-world network (Watts and Strogatz, 1998) and scale-free network (Barabási and Albert, 1999).

To explain the differences of the various types of network, I will first introduce some of the quantitative measures of the network properties which have been widely used. There are two basic measures at the level of individual nodes: for a node i,

1. Degree (k_i): the number of links the node i has.

2. Distance (d_{ij}): the number of links along the shortest, i.e., geodesic, path to connect node i to node j.

The aim of network analysis is to go beyond the individual nodes and obtain the global statistical properties of the whole network. The following gives some of the common concepts in describing a network structure:

1. Density, D: the percentage of actual links among the maximum possible links in the network; it equals to the ratio between average degree ($<k>$) and the size of the network. Usually real-world networks are sparse, which means $\frac{<k>}{N} \ll 1$.

2. Degree distribution, $P(k)$: the statistical distribution of the degrees of the nodes. Usually it is plotted as a histogram or extrapolated as a distribution function. It gives an idea of the homogeneity and scaling properties of a network. A regular network is homogeneous as the degrees of all nodes are the same, and its $P(k)$ is a delta function shown as a single spike in the histogram. Typical random networks are found to follow a Poisson distribution.

3. Average path length (also called characteristic length), L: the mean or average shortest distance between any two nodes. This gives an idea of the effective size of the network. For regular networks, L increases linearly with respect to the size of the network. In random networks, the distance between any two nodes is small even when the network is huge; L increases logarithmically rather than linearly when the network size increases.

4. Clustering coefficient, C: the average percentage of pairs of neighbors of a node that are also neighbors of each other. It provides a measure of the independence of neighboring links. Random networks have a small C which tends to zero in the limit of large network size. In contrast, regular networks are highly clustered compared to random networks.

The actual implementation of the measure for these properties involves many detailed considerations. Rigorous definitions and various algorithms for calculation can be found in Newman (2003).

The discovery of small-world networks (SWN) is renowned as it incorporates of several interesting characteristics of real-world networks, including a short average path length (L) as shown in random networks, and a high clustering coefficient (C) as shown in regular networks. Watts and Strogatz (1998) propose a very simple way to construct a small-world network to have these characteristics. First form a one-dimensional ring; each node has the same number of connections with its close neighbors in the ring. Then a proportion of these regular connections are chosen to be rewired randomly. The rewiring procedure involves going through each node and, with a probability p, changing each of its existing regular connection from a close neighbor to another randomly chosen node, provided that no double connections or self-connection are ever created. The parameter p indicates the degree of randomness in the small-world network. It has been found that for a large range of p that the network exhibits the two small-world characteristics, i.e., small L and large C. Later, other refined methods are proposed for building networks with similar small-world characteristics but easier for analysis, such as adding random connections instead of re-wiring (Newman, 2003).

Later, Barabási and Albert (1999) find that many real-world networks, such as Internet, WWW, etc., exhibit a scale-free characteristic, that is, the degree distribution of the network follows a power-law, instead of a Poison distribution as random networks or small-world networks do. They propose a method composed of two mechanisms, growth and preferential attachment, to construct networks which exhibit the power-law degree distribution (the model is called B&A model hereafter). The network is built as follows: The network starts with a small number of isolated nodes, say m_0; at every time step, say t^{th} time step, a new node is added into the network; each new node has m ($m \leq m_0 + t$) connections, linking to m different nodes in the network, based on a preferential attachment mechanism: the more connections that a node has, the higher chance for the node to be chosen to connect with the new node. After t time steps, the procedure results in a network with $N = t + m_0$ nodes, and $m \times t$ connections. The final network exhibits scale-invariant characteristics in its degree distribution: the probability $P(k)$ that a node has k connections follows a power law, $P(k) = \frac{2m^2}{k^3}$. The exponent is constant, $\gamma_{BA} = 3$, regardless of the parameters m_0 and m.

The B&A model to construct a scale-free network has been modified in various ways in later studies to overcome its several deficiencies, such as the fixed exponent of the power law function, which are not consistent with those found in different networks in the real world. Among these modifications, a model proposed by Li and Chen (2003) is worth mentioning. In B&A's model, it is assumed that the new nodes can have access to *all* existing nodes in the network, and choose to connect to those which have more connections. However, such access to the global network is not realistic, especially in large-scale human networks. In reality, a person only has his own limited world-view, as he can only reach a limited part of the world. Friends and personal connections a person can make are mostly limited to his local area (though some sporadic long distance connections are possible owing to the modern communicative means such as telephones, emails, or webcams). Based on these considerations, Li and Chen (2003) propose a local-world evolving network model, which adds a small modification on the preferential attachment mechanism. At each time instant, instead of choosing nodes from the whole network, a new node first selects M nodes randomly from the existing network, and applies the preferential attachment to these M nodes. In fact the local-world network model presents a more general model. Its limiting case of $M = m_0 + t$ corresponds to B&A's model, and the other limiting case of

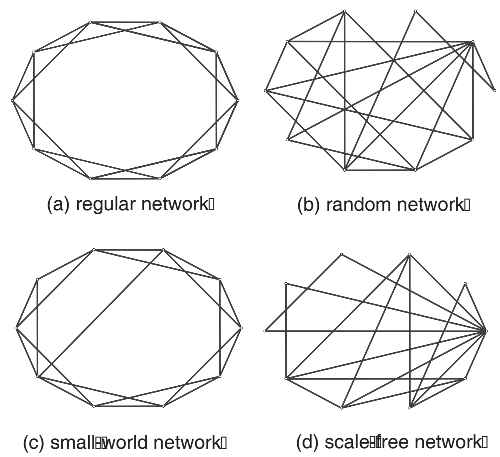


Figure 5.6: Four typical networks (a) regular network, (b) random network, (c) small-world network (based on Watts and Strogatz (1998), $p = 0.01$), (d) scale-free network (based on Barabási and Albert (1999), m0=m=2); all with the same size (N=10) and same average degree($<k>$=4).

$M = m$ corresponds to a random network model. When $m \leq M \leq m_0 + t$, the local-world model represents a transition of degree distribution from Poisson to power-law distribution, in other words, from random networks and to scale-free networks.

Figure 5.6 give examples of the four typical networks, all with the same size (N=10) and same average degree($<k>$=4). It is easy to see that there exist some highly connected nodes in a scale-free network, while the random network and small-work network are closer to regular network in which the nodes have homogeneous degree. To illustrate better the differences in degree distributions of these four typical networks, we compare four networks with a larger size (N=1,000, $<k>$=20), as shown in Figure 5.7. Due to the existence of the

Newman (2003) points out that in recent years a great deal of effort on complex networks has contributed to three main areas. One is to analyze the statistical properties which characterize the structure and behavior of networked systems, and to suggest appropriate ways to measure these properties. Besides the several basic measures introduced above, more and more complicated measures have been proposed to zoom in on the detailed aspects of structures, such as the betweenness centrality and assortativity which allow to detect community structures in networks (Newman and Park, 2003). The second area is to create models of networks that can help us to

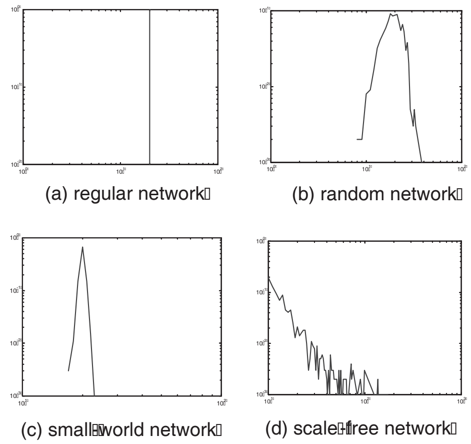


Figure 5.7: The degree distribution of four typical networks (N=1,000, <k>=20).

understand the meaning of these properties and explain how they come to being and how they interact with each other. In addition to these two areas, Newman specially highlights that there is a third area of network studies, which is to predict what the behavior of networks will be on the basis of certain structural properties, and study the effect of the network structures on the system behavior. He remarks that the last area is still in its infancy, compared to the first two areas. In this study, one of our interest is to examine the effect of different social structures on the diffusion pattern in language change. This work will serve as a case study in the third area.

5.2.5 The baseline model

As mentioned earlier, Shen (1997) proposes an analytical model using a logistic function (Equation 5.1) to model the diffusion of linguistic innovations, and shows that it can produce an expected S-curve dynamic, as reproduced in Figure 5.8(a). We transform this analytic model into a stochastic simulation model, in order to simulate the situation of an finite population (in the following simulations, the population size is set as N=500, unless specified otherwise). In order to make the two models match, the simulations model has the following assumptions: the change is uni-directional (only $U \rightarrow C$, no $C \rightarrow U$); agents all have the same probability, α, to change to adopt C; the agents are in a fully connected network so that any two agent could interact.

In the beginning, one agent has the innovation C while all others have U. Agents

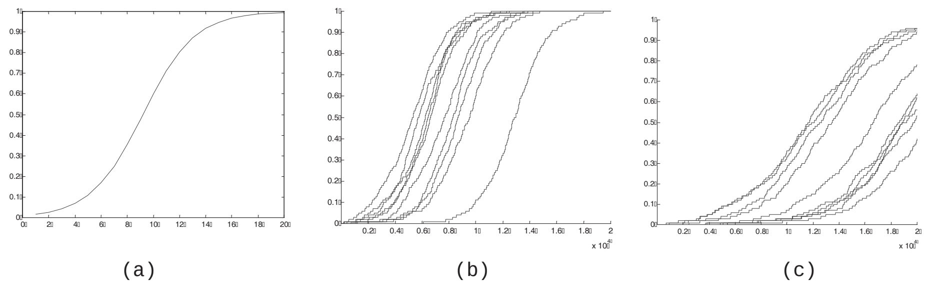


Figure 5.8: Diffusion in a homogeneous regular network with different degrees of density (N=500, $\alpha = 0.5$, 10 runs). (a) the analytic model from a logistic function; (b) D=100%; (c) D=50%.

start to interact; at each time instant, two agents are selected. If one agent has the innovation, the other agent who still uses U will have a probability ($\alpha = 0.5$) to change to C. After a number of interactions, the innovation C diffuses to the whole population, exhibiting an S-curve dynamics. As there are random parameters, such as whether the agent changes or not, which two agents are selected to interact, and so on, different runs give different results; but they all show S-curve patterns. Figure 5.8(b) shows the results of 10 runs. The average time for the diffusion to complete is about 11,000 steps. Each curve in a graph traces one diffusion process with the innovation randomly generated from one agent. The x-axis is the number of steps, and the y-axis is the percentage of agents using the changed form.

In reality, a fully connected network is very rare, even in very small networks. In the study of social networks, it is believed that there is an upper limit to the number of relations that each agent can sustain, constrained by time and cost of making and maintaining relations. The maximum value for density which is likely to be found in actual social network is suggested as 0.5 (Scott, 2000).

When the network connections become sparse, the diffusion slows down as the slope of the S-curve decreases. Figure 5.8(c) shows the situation when the density drops to 50%. Within the given number of 20,000 steps, only in 3 runs the diffusion is complete, and on average the innovation only reaches about 75% of the population. Despite the slow dynamics, the diffusion still proceed in S-curve shapes.

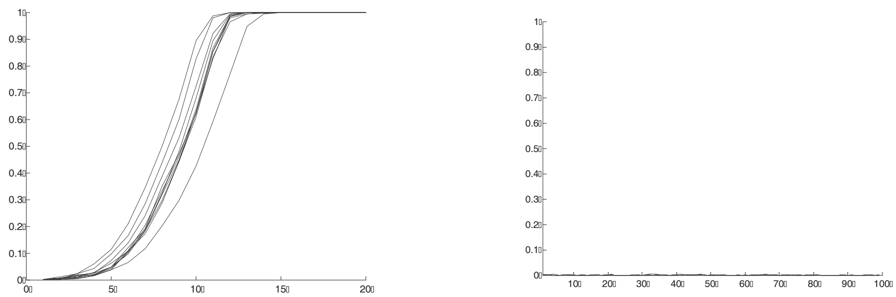
Real social networks hardly look like regular networks. Though people are more likely to connect to their family members, local neighbors, colleagues, and so on, they would have long-distance connections such as friends and relatives living in other cities. Real social networks are not like random networks either, as the majority of local connections is significant. Real social networks have been shown to exhibit characteristics of small-world networks, such as in collaboration network of film actors (Watts and Strogatz, 1998) or scale-free networks, such as in romantic relationship network (Liljeros et al., 2001). We will compare these different types of networks in the following experiments. Before examining the effects of networks, we recognize the current baseline model has some unrealistic assumptions. One important problem is that the agents are homogeneous in age and their learning capacity, and they are assumed as immortal. We will present some modifications of these assumptions in the model.

5.2.6 Model with age structure

In reality, adults and children may have different responses to linguistic innovations. It is often assumed that adults seldom change their linguistic behaviors since adolescence. The "apparent time" approach (Labov, 1994) to reconstruct the historical profile of ongoing changes in sociolinguistic studies is based on this assumption. In the following models, we add age structure and changing population to the previous static model, following Nettle (1999c) mentioned in Section 4.1.2.

In the model, each agent in the population passes through five life stages. Agents at stage 1 are infants who only learn from others, and do not influence others; agents at stage 2 are children who both learn and affect others; agents at stages 3-5 are adults who use what they have learned before adulthood to teach the infants and children. The population is initialized with an equal proportion of agents in these five stages. Therefore the ratio between adults and children is 2:3[11]. At the end of each life stage, each agent advances to the next stage, and agents at stage 5 are replaced with new

[11] This ratio may not conform to the demography in a modern society. For example, in Hong Kong, the ratio between children (<14 year old) and adults (>15 years old) is about 1:4. However, as the aim of the model is to illustrate the importance of considering non-homogeneous and dynamic populations, the actual ratio is not crucial, and will not affect the quantitative conclusion drawn from this model. Furthermore, this ratio with a smaller difference between adults and children may be more appropriate if we consider the pre-modern human society or hunter-gatherer society.



(a) adult change prob=50%   (b) adult change prob=0.1%

Figure 5.9: Diffusion in an age-structured small-world network ($N = 500$, $< k >= 10$, $p = 0.01$, each time step 4000 interactions). (a) adults change ($\alpha = 0.5$) (b) adults rarely change ($\alpha = 0.001$).

infants.

Figures 5.9(a) and (b) compare the diffusion in a small-world network under two different situations: (1) adults change with a certain high probability ($\alpha = 0.5$) when encountering the innovation; (2) adults rarely change ($\alpha = 0.001$). We can see from the simulation results that in a changing population, if adults can learn the innovation easily, the innovation can spread quickly; on the contrary, if the innovation is rarely adopted by the adults, the innovation cannot spread but dies out.

In the above model, the diffusion is from an "interaction" perspective, that is, at each time step, every agent has a given number of interactions with his connected neighbors and each interaction gives the agent a chance to change when the innovation is encountered. This implementation has a difficulty in determining the appropriate number of interactions for the agent at one time step. To examine the long term effect of learning on language change, the actual detailed sequence of interaction during learning period may not be important at all. What is important is what the children finally learn and continue to use afterwards.

Therefore, in the following, we simulate the diffusion process from a "learning" perspective. In the model, adults rarely change, and innovation mainly spreads through children's learning. The learning process is simplified as follows: at each time step, each learner takes input from all his connected teachers. If all teachers use the same form, then it is no doubt that the learner will only learn this form. If both forms, i.e., U and C, are present in the learning environment, the leaner compares the accumulative impact of the two variants and chooses the one which has a higher impact. Nettle's model in fact takes this learning perspective. In his model, a learner samples from the entire population (except the infants), but a teacher's impact on the learner decreases as his distance to the learner increases. As Nettle's model simulates a weighted regular network which is not realistic, we adopt a small-world network structure instead. The following simulation experiments will mostly be based on the model with this implementation, unless specified otherwise.

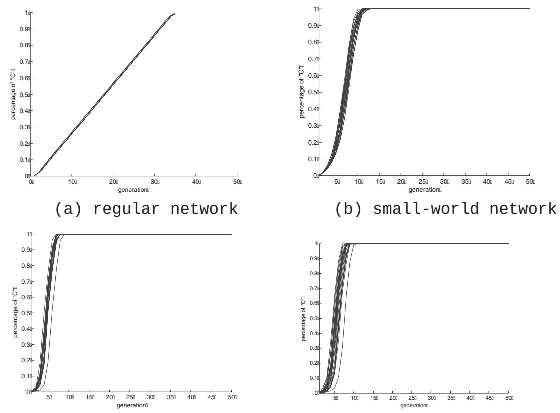
In later simulation, the experiments measure how likely the innovation can override the original norm to become the new norm adopted by the majority of the population. For each set of experimental parameters, the simulation is run 100 times. The number of runs with successful diffusion is taken as the index for comparison with other sets of parameters.

5.2.7 Different functional biases toward the innovation

While various linguistic innovations appear constantly, some are easier to learn than others, such as those with shorter form, salient in perception, or involving a simpler rule, and so on. These are considered as the functional aspects of the linguistic items. The various functional criteria proposed to account for language universals, as discussed in Section 1.1.2.1, can be used to explain language change as well. Linguistic innovations which have a higher functional advantage would be preferred in language acquisition, and therefore have a higher chance to spread across the population.

In the following simulation experiments, different functional biases, represented by an abstract index, are compared under different conditions. Figures 5.10, 5.11 and 5.12 show the diffusion in four types of networks under four different conditions.

When there is only one adult introducing the innovation (here innovation is only created by adults, and we do not consider the case of imperfect learning), and the innovation has a huge functional bias ($\beta = 20$), diffusions are all successful, as shown in Figure 5.10. But different networks show different dynamics. The innovation diffuses in a linear way in regular networks, while in the other types of networks, the



(a) regular network (b) small-world network

(c) random network (d) scale-free network

Figure 5.10: Diffusion dynamics in four types of networks in 10 runs (network size $N = 500$, average degree $< k >= 20$, functional bias $\beta = 20$, number of innovators $I = 1$).

diffusions show similar sharp S-curves. The rate of diffusion is about 3 times slower in a regular network than in other networks. The slow diffusion in the regular network can be explained by its lack of "short-cut" connections between distant nodes in the population, while the other three networks have these "short-cuts".

A functional bias with the value of 20 and innovation originated from only one adult are too extreme and unrealistic conditions. When we decrease the functional bias from 20 to 10, if there is still only one innovator, there is no diffusion at all for all networks. If there are more innovators, say 10, the innovation is successful in most of the cases for the same functional bias of 10, as shown in Figure 5.11. Regular networks and small-world networks show similar gradual diffusion, while random network and scale-free network still exhibit the rapid diffusion. Here we observe clearly the stochasticity of the diffusion process. Under the same condition, except the origin of the innovation from different agents, the innovation has different fate. Some can be successful in diffusion, while some die out. This is seen in some cases in regular network and small-world networks.

We further decrease the function bias from 10 to 2. The innovation can only spread when there is a large number of innovators. We show in Figure 5.12 the situation when $I = 100$, which is only likely in the case of a massive immigration flow. Regular and small-world networks again show gradual diffusions, while random and

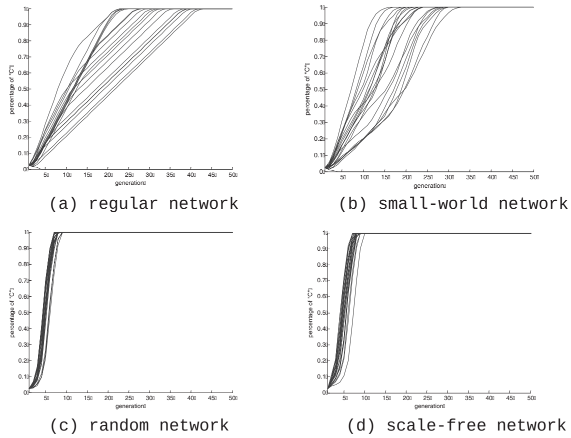


Figure 5.11: Diffusion dynamics in four types of networks in 10 runs. ($N = 500$, $< k >= 20$, $\beta = 10$, $I = 10$)

.

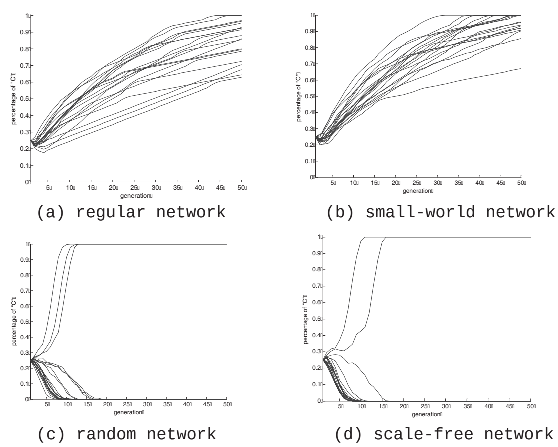


Figure 5.12: Diffusion dynamics in four types of networks in 10 runs. ($N = 500$, $< k >= 20$, $\beta = 2$, $I = 100$).

scale-free networks show rapid diffusion in sharp S-curves. But the latter two types of networks have fewer successful diffusions than the former two.

We further compare more systematically the four types of networks under different conditions of functional bias and percentage of innovators. We fix the number of innovators as 10, and check the probability of successful diffusion and the diffusion rate. The upper panel of Figure 5.13 gives the probability of successful diffusion over 100 runs for different functional biases. We can see that for a range of small functional biases ($\beta = 3$ 7), scale-free and random networks have much smaller probabilities of successful diffusion than regular and small-world networks, or we can say that the

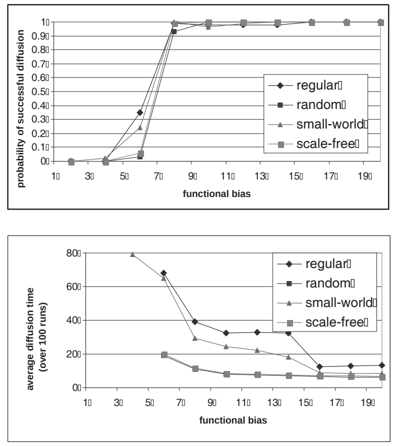


Figure 5.13: The probability of successful diffusion under different functional biases, and the average diffusion time in four types of networks over 100 runs. ($N = 500$, $< k >= 20$, $I = 1$).

first two types of networks require a higher threshold of functional bias for successful diffusion. But the first two types of networks take much less time to complete the diffusion than the latter two, as shown by average diffusion time over 100 runs in the lower panel of Figure 5.13.

Similar observations exist for the parameter of percentage of innovators. As shown in Figure 5.14, under the same condition of functional bias of 4, scale-free and random networks have higher thresholds for the input of innovations, but diffuse at much higher rates than regular and small-world networks.

When either functional bias or percentage of innovators is high, there are little differences among the four types of networks. However, in most cases of the real world, innovations do not have very high functional biases, sometime even no bias; also they often only occur in a small number of innovators except for massive contact situations. Under these conditions (small functional bias and small number of innovators), the four types of networks show different characteristics, as summarized in Table 5.5. The dynamics in small-world networks is similar to that in regular networks: high success probability, but slow diffusion rate. Scale-free networks are similar to random networks: fast diffusion rate, but lower success probability.

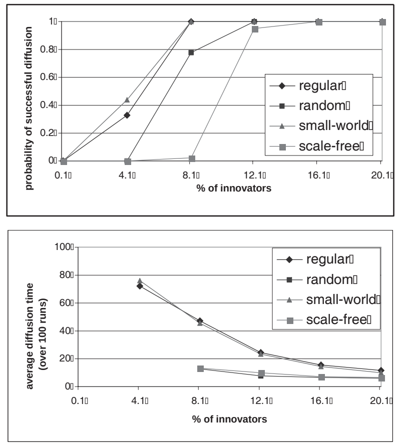


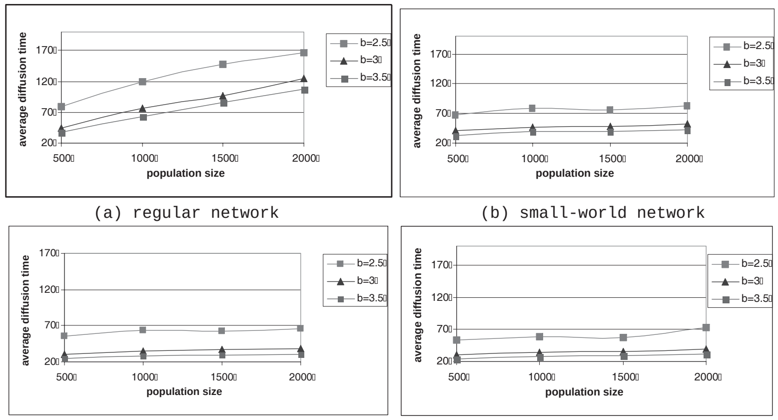
Figure 5.14: The probability of successful diffusion and the average diffusion time under different numbers of innovators in four types of networks over 100 runs. ($N = 500$, $< k >= 20$, $\beta = 4$).

Table 5.5: Comparison of the four types of networks in probability of successful diffusion and rate of diffusion.

	Probability of success	Rate of diffusion
Regular network	High	Gradual
Small-world network	High	Gradual
Random network	Low	Abrupt
Scale-free network	Low	Abrupt

5.2.8 Effect of population size

Nettle (1999a) shows by his model that a larger community requires longer time for changes to complete, and thus fewer changes will occur. In his model, as mentioned earlier, the social network is a kind of weighted regular network. Here we compare regular networks with other three types of networks as shown in Figure 5.15. Our model shows that in regular network, the diffusion increase almost linearly as the increase of population size, similar to Nettle's model. However, the other three types of networks do not show a similar result. Instead, there is little increase in the diffusion time, compared to the regular network. One possible reason is that there are many random connections linking different sectors of the populations in the three types of networks. More importantly, there exist hubs in random and scale-free networks, which can be seen from their degree distributions. We note that to extend the discussion of the effect of population size to linguistic diversity, more careful examinations of the relation



(a) regular network (b) small-world network

(c) random network (d) scale-free network

Figure 5.15: The relation between community size and the rate of change in four types of networks.

between language change and diversification are necessary.

5.2.9 Effect of two types of learners

In the above experiments, the agent in the model, faced with the presence of the two competing variants, only learns and uses one of them. This is inconsistent with our empirical findings reported in Chapter 3, where most of the speakers use two variants for some of the morphemes subject to the change. The co-existence of variants in individual speakers (i.e., VT-III) is common. Therefore the model should take into account this consideration. The dynamics of change in the situation that agents allow co-existence of variants should be different from the situation that does not.

In Chapter 3, we hypothesize two types of learners, i.e., "categorical" and "statistical" learners, in terms of their capacity to accommodate competing variants. Niyogi (2002) has a similar classification of learning behavior, and he uses analytical models to compare the two situations. In his models, a categorical learner only learns and uses the form which is encountered *often and early enough* (i.e., the variant which first has k occurrences) during the acquisition period, while a probabilistic learner acquires both forms and uses them in proportion to their frequency. The latter is called a "blending" learning behavior in his term. His models show that if children learn categorically, the mixed population will not be stable and will always converge to using only one form, except one special condition that children learn only from two parents and the probability of learning from one parent is exactly 0.5. On the other hand, for a population with probabilistic learners, the two variants will co-exist for all time. He claims that "categorical behavior on the part of learners results in an inherent tendency of linguistic population to change with time to a homogeneous stable mode with only one linguistic form surviving. Blending behavior on the part of the learner leads to both forms being preserved in the population at large. This is in contrast to models of inheritance in evolutionary biology where blending inheritance eliminates variation while particulate preserves it" (ibid, p274). He conjectures that "it is the categorical nature of language that forces change" (ibid, p267).

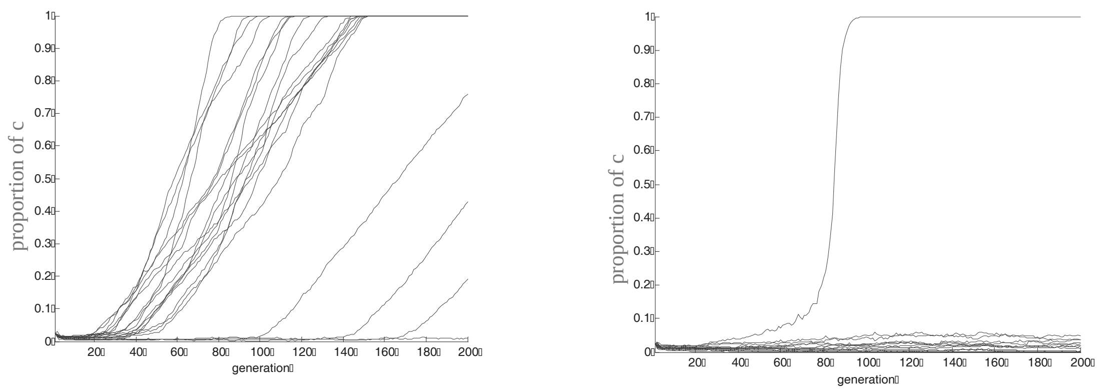
Niyogi attempts to apply these theoretical results to analyze an on-going sound change in the Wu dialects of China described in Shen (1997). Two vowels, one diphthong [ɸy] and one monophthong [ɸ], are merging, as attested in many emerging homophonous morphemes which used to be minimal pairs. Niyogi considers that Shen's data suggest that people show categorical behaviors, i.e., people use consistently one of the two competing forms, and the data is consistent with his conclusion that categorical learning is the force for language change. However, we examine the questionnaires that Shen used in his survey and recognize that the categorical appearance is an artifact due to the experimental design: the questionnaire did not provide the possibility for the subjects to show the co-existence of the two forms.

There are some other problems in Niyogi's models. First, they are based on some unrealistic conditions, such as infinite population size and homogeneity in the population, i.e., all agents learn in the same way. However, it is hardly the case in reality. As suggested in our empirical study, people may have different learning styles. Faced with the linguistic environment in which variants are present, some people may only acquire one form, while some others may learn two forms. Moreover, the two learning situations assumed in his model, i.e., learners either learn from the whole population or from their parents, are not realistic. Though Niyogi proves that statistical (or blending) learners would always preserve the two forms in the population, the outcome of the stochastic model without these unrealistic conditions may differ greatly. We will illustrate this point in the following experiments.

Second, though Niyogi's model proves that categorical learners always result in a population converging to one form while the other variant dies out, in fact the convergence is not always equal to a language change. His models have not considered the initial condition of a language change. In a general situation, as we have discussed for the "threshold problem", the change always originates as a linguistic innovation from a small number of speakers. If a form can only be learned when its frequency in the linguistic input is beyond a certain threshold, as Niyogi's model of categorical learners assumes, then it is very likely that a learner will learn the predominant form, and it is hardly possible for the innovative and infrequent form to be acquired. If the innovation is not acquired and used by subsequent generations, it will die out with the old generations and no change happens.

We use our model to test the effects of the two types of learners proposed. Unlike Niyogi's model, the learners in our model learn from all connected neighbors. At each life stage, learners evaluate the impact of the forms, if they encounter more than one during their learning period. The impact of the variant form is measured by the product of its functional bias and its frequency. A categorical learner adopts only the form which has higher impact, while a probabilistic learner may adopt both forms and uses them probabilistically proportional to their impact. At the beginning when the innovation C is still rare in the population, learners will only encounter U and they will only learn and use U. But at later stages of the change when the innovation has diffused to more speakers, learners are likely to be exposed to both U & C. If a child has encountered U three times and C twice from his teachers, and if the functional bias toward C is 2, then a categorical learner will use only C consistently in his adulthood, while a probabilistic learner will use both forms with different probabilities, $Prob(U) : Prob(C) = 3 : 4$.

With the existence of probabilistic learners, the innovation with a small functional bias can spread much more easily than in a population with only categorical learners. Figure 5.16 shows the diffusion of an innovation with a functional bias of value 2, starting from only one adult, in a small-world and a scale-free network. Under this condition, the innovation has no chance to diffuse at all in a population with all



(a) small-world network (b) scale-free network

Figure 5.16: The diffusion dynamics in a population with all probabilistic learners in two networks. ($N = 500$, $< k >= 20$, $\beta = 2$).

categorical learners. If the learners are all probabilistic learners, diffusion are possible. And consistent to earlier findings, small-world networks ensure more successful diffusion, but require longer time to complete the diffusion than scale-free networks.

The above results seem to contradict the conclusion from Niyogi's models. As mentioned earlier, his analytical models are based on some unrealistic conditions. Especially in a finite population, his prediction that probabilistic learners preserve variants and do not lead to change is only possible when the innovation starts with about the same number of agents as the original norm. Even so, the variants do not always persist. Figure 5.17 shows the results of 10 runs in this situation. For some runs, the proportions of the two forms fluctuate for a long time, but in most of the runs there is a clear tendency for one of the variants to dominate and the other to die out.

Wonnacott (2003) has similar discussions on the effect of different types of learners on language change, though she focuses on the condition when the S-curve trajectory is likely to occur. Our results are consistent with her findings, which is that the probabilistic learner, or "probabilistic learner" in her term, will result in changes with S-curve diffusion patterns. However, her model only considers the situation where the two forms are of equal proportion at the beginning, which is only one special situation in language change, such as when two different populations of similar size get mixed,

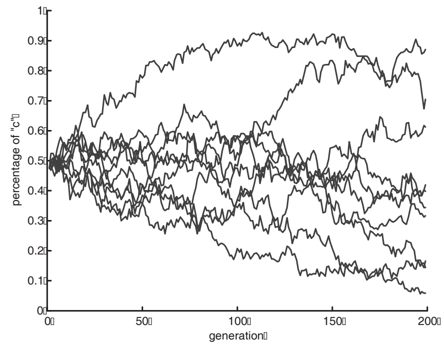


Figure 5.17: The diffusion dynamics of 10 runs under the condition that learners are all probabilistic learners, the innovation has no functional advantage and originates in 47% of the population.

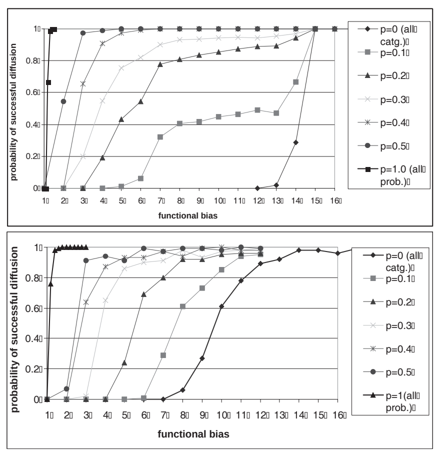


Figure 5.18: Probability of successful diffusion in populations with different proportions of probabilistic learners, under different functional biases. Upper panel: small-world network; lower panel: scale-free network. ($N = 500$, $< k >= 20$, $I = 1$)

.

or a local population receives a large influx of immigrants.

We compare the effect of probabilistic learners under conditions of different functional bias values. In a small-world network, if all learners are categorical, the threshold of functional bias for successful diffusion is 13, but with 10% of probabilistic learner, the threshold decreases to 5. If half of the population are probabilistic learners, then an innovation with functional bias of 2 will have more than 50% of chance to successfully spread. If the whole population are all probabilistic learners, a functional advantage of 1.3 will allow 99% successful diffusion. Similar phenomena can be observed in scale-free networks.

Combining with the empirical findings and the simulation results, we submit that it is because of the existence of probabilistic learners that language change is so frequent - many innovations can successfully spread as long as they have a small functional bias to replace the original norm.

Chapter 6

Conclusions, discussions and future work

6.1 Self-organization as a unifying framework

In this thesis, I advocate the framework of self-organization for the study of language evolution. Self-organization has been recognized as a new scientific paradigm for the study of complex adaptive systems in various disciplines. The essence of this framework is to provide a perspective different from reductionism, which has been considered as the doctrine in scientific investigation. To explain complex patterns in systems, the self-organization theory views these patterns as *emergent* phenomena at the global level. These patterns cannot be understood by only studying the properties of the constituting components, i.e. "agents", themselves. In addition, it is important to examine the effects of long-term local interactions between agents in the system. Usually, it turns out that the constraints governing the agents' actions and interactions are very simple, but the local interactions between agents can produce very complex patterns, such as termite mounds, honeybee combs, etc.

The self-organization framework may unify the two existing perspectives on language evolution, as the biological perspective will contribute to identifying the properties of *individual agents* (i.e. the biological basis for language learning and processing), and the cultural perspective will contribute to examining the conditions and effects of *interactions between agents*. These two lines of investigation correspond to the two main tasks in analyzing a self-organizing system. Under this framework, independent areas of language evolution, including language acquisition, language change, and language origin, can be integrated into one coherent framework, aiming to construct a more complete picture of how language has evolved and is evolving.

6.2 Investigations at three levels: system, population, and individual

There are three ways to view language evolution in a self-organization framework. One is to view an abstract language system itself as self-organizing; the other two views are applied to the two levels of existence of language: the communal language and the idiolect, each undergoing self-organization processes.

We exemplify the first view in a study of homophony. The existence and evolution of homophones are considered as the result of self-organization in the language system. Homophones emerge as the result of sound mergers, and distribute in a language in a non-random way: homophone words tend to differentiate from each other in grammatical classes and in frequencies. Moreover, the degree of homophony in a language is highly correlated with the degree of monosyllabicity: The more the monosyllabic words a language has, the more homophones it will have. The degree of monosyllabicity is also the result of the self-organization in the language system: in the most frequent words, the degree of monosyllabic words are similarly high in different languages, and the degrees all drop in less frequent words; but different languages level off at different degrees, which are correlated with the size of phonological inventory.

In the first line of analysis, the language is only considered as an abstract system, and how the self-organization is implemented is not specified. In comparison, the latter two levels of self-organization are grounded with actual existence, i.e. language in the individuals and the language in the community. A case study of an on-going sound change in Cantonese is taken as an example to illustrate the self-organization process at these two levels. Sound change does not progress as if the individual speakers are simultaneously applying a rule of phonological change in their language behaviors.

Instead, language change can be considered as innovation diffusion. It progresses as adults change through interaction with the innovating adults, as well as learners of new generations adopt the innovation. As adults' changes are much less frequent, the main driving force of language change is the learning of new generations. A consequence of this process is the proportion of speakers using different variants in the population changes, similar to how population genetics conceives biological evolution: it is the change in the distribution of individual genotypes over time, rather than the individual themselves changing directly.

Language acquisition is a self-organizing process. Being exposed to the heterogeneous linguistic environment as exemplified to an extreme in the situation with an on-going sound change, different learners exhibit different learning styles which result in different language systems. Especially, we observe that most speakers can learn simultaneously both variants of a change, but speakers differ in the number of co-existing variants. More importantly, the set of words subject to the change may have very different distributions in the two variant categories in different speakers, even in identical twins.

The self-organization at the population level is reflected in that the language community is stratified in different social parameters, as sociolinguistic studies have revealed. For example, in our study of sound change, it is found that speakers with different educational backgrounds exhibit different degrees of change. The stratification is the result of the interactions between speakers. Compared to language change, the effect of interactions in a population of agents is more clearly demonstrated in the study of language emergence, as shown in various computational models.

6.3 Heterogeneity as the norm

When adopting a self-organization framework for studying language evolution, we readily take heterogeneity as the nature of language as the starting point of our investigation. Heterogeneity should be considered the norm.

As we discussed above, language acquisition can be considered as a self-organization process. The same linguistic input may result in different idiolects, as there exist different learning styles, such as the categorical and statistical styles we proposed in Chapter 3. Even when we assume the same learning styles, it is hardly possible for different learners to have exactly the same linguistic environment. The minor differences in the initial condition may result in very different idiolects, due to the characteristics of nonlinearity in self-organizing systems.

As we recognize the existence of the large degree of heterogeneity across speakers, it becomes unreasonable to adopt the paradigm of an ideal speaker-listener pair to analyze a language system, because there does not exist such a pair. Also, there is no such thing as a well-structured language system appearing as "*tout se tient*" (Meillet, 1903-1904, p461). The language system at the population level should contain a large degree of heterogeneity. The linguists' attempt to construct a coherent grammatical system for a language may be a task unachievable, because no such system actually exists. As Lü (1979) illustrates in his review of the problems in the analyses of Chinese grammars, almost all levels of categorization, such as transitive-intransitive verbs, grammatical classes, word-or-phrase, phrase-or-sentence, and so on, have many ambiguous cases, and conflicts may arise when different criteria are employed. The determinination of "wordness" in Chinese is a notorious example; syntactical or lexical criteria may result in different analyses.

In fact such problems may be better considered as artifacts created by linguists. Most ordinary language users do not aim to construct a language system as systematic as linguists assume. The task of language acquisition is to *use* language, instead of "*grammar identification*" (Seidenberg, 1997). An idiolect is constructed incrementally in a piecemeal manner during the process of using language to communicate (Tomasello, 2003).

A language can be viewed as a changing "mosaic" (Wang, 1982a), either at the population or the individual level. If there are indeed homogeneous elements across speakers, they are the physiological and cognitive constraints at the lowest level for language processing and learning. These are the properties of the agents in the self-organization framework that we aim to ultimately reveal.

6.4 Computational modeling as a promising methodology

In addition to empirical studies, this thesis adopts a methodology of computational modeling, which is a burgeoning area in studying language evolution. Computational modeling complements traditional linguistic studies by providing a convenient and effective way to test various hypotheses for language evolution, and compare various conditions under control. The modeling methodology has become more and more a central part of linguistic investigation.

As reviewed in Chapter 4, computational models have been used in studying different scales of language evolution, i.e. acquisition, change and origin. There are various types of models with different levels of resolution: some assume language as a synthetic whole and examine the evolutionary dynamics under various selection conditions, being either biological or cultural; some model language with more detailed representation in some specific domains, such as phonology, lexicon or certain syntactic structures, but without actual language use in communication; some take language as mainly a complex set of meaning-form associations, and simulate its evolution by going through iterative communicative interaction and/or learning process.

Among the various types of models, agent-based models are the ones which fit the self-organization framework the best, which are often adopted in the models at the finest resolution of implementation of language. In these models, the two levels of existence of language are readily incorporated. The idiolects are represented with detailed linguistic components, confined by built-in biological constraints as well as social constraints in interactions. The effects of these various constraints can be studied systematically. In these models, the emergence of a communal language in a heterogeneous population is a natural product (in some models, the population finally converges to a homogeneous language; but the internal representation of language in the agents may not be the same, especially in the connectionist models.)

6.5 Future directions

Following the studies which have been carried out in the thesis, several interesting research topics are worthy of further extension and exploration. I will discuss them briefly in the following.

6.5.1 Cross-dialectal investigation of /n-/⟷/l-/ changes

The sound change between /n-/ and /l-/ examined for Cantonese is in fact an interesting case of a pan-dialectal change. It seems the merger between /n-/ and /l-/ happens in all major dialects: in addition to Yue (Cantonese being one of the varieties), it also occurs in Min, Gan, Kejia, Wu, and Xiang (Karlgren, 1937/1994; Yuan, 1960/2001; Ho, 1988; Zhan et al., 1990; Hou, 2002, etc.).

The parallel change in different dialects would serve as a good case study for differentiating vertical and horizontal transmission. We are interested in identifying the sources of the change in various dialects from one of the three possibilities: 1) the change is borrowed from other dialect(s), 2) the same change is inherited from Middle Chinese, or 3) the change developed independently in individual dialects. We hope that some traces of the change in early times can be identified from the analyses of the historical texts of various dialects.

Besides the research in the histories of dialects, it is also important to compare the contemporary situation of the dialects. While we hypothesize a reversal change, /l-/→/n-/, is taking place in parallel with the on-going change /n-/→/l-/ in Cantonese, it is interesting to see whether such a reversal change also happens in other dialects, due to the strong influence of Putonghua in recent decades.

In addition to the analyses of the situation of the change in dialects, we are interested in knowing why such changes are so common and wide-spread in Chinese dialects. We will look for explanations from phonetics concerning the physiological and psychological bases for the changes. Considering production, the two sounds, /n-/ and /l-/, are both plain consonants which differ mainly in manner of articulation, one being a nasal stop sound and the other an oral lateral sound[1]. "Production ease"

[1]According to Ladefoged and Maddieson (1996), nasal consonant is one "in which the velum is

may serve as one possible explanation for the change /n-/→/l-/, assuming /l-/ may require less effort than /n-/[2]. However, some counter-argument may be taken from the acquisition data, as it is found that /l-/ is usually acquired later than /n-/ (Wong and Stokes, 2001; Zhu and Dodd, 2000), suggesting /l-/ may be more difficult for production.

Besides production, the trigger for the change /n-/→/l-/ may come from perception. However, acoustically, /l/ is significantly different from /n/. While the first formants of /l/ and /n/ are similarly low, but for /l/ "the second formant have a center frequency, . . . and the third formant has typically a relatively strong amplitude and high frequency; and there may also be several closely spaced additional formants above the frequency of F3" (Ladefoged and Maddieson, 1996, p193). An illustration of the spectrograms for a pair of sounds /ni/ and /li/ are shown in Figure 6.1. In fact, the phonemic distinction between /n/ and /l/ is widely attested in a large number of languages in the world[3]. Also, Niyogi and Surendran (2003) show that the functional loads for the two phonemes /n-/ and /l-/ in Cantonese are quite high and the merger is supposed to be unlikely, given the resultant large expense of possible ambiguity. Therefore, the reason for the wide spread change between /n-/ and /l-/ is intriguing and worthy of further investigation[4].

It is not clear yet whether there exists an intermediate sound between /n-/ and /l-/ which Karlgren called "naso-oral" (1936/1994, p176). And it is interesting to know if there exists a categorical perception boundary along the continuum in the acoustic space. Perception experiment will be necessary to answer this question.

Moreover, there is a special case in language change, so-called "near-merger" (Labov,

lowered and there is a closure in the oral cavity somewhere in front of the velic opening. Hence, air from the lungs is directed out through the nasal passage alone."(ibid, p102). /l/ is a lateral sound, "in which the tongue is contracted in such a way as to narrow its profile from side to side so that a greater volume of air flows around one or both sides than over the center of the tongue." (ibid, p182).

[2]I would like to thank Prof. Robert Bauer who raised the discussions for the constraint of "least effort" as the reason for the change in question.

[3]From the statistics of the latest version of the UCLA Phonological Segment Inventory Database (UPSID) (Maddieson and Precoda, 1990), among 451 languages, 80.5% languages have /n/, and 68.7% have /l/, and 60.8% have both /n/ and /l/, while only 19.5% have /n/ but not /l/, and 8% have /l/ but not /n/, and 11.8% have neither /n/ nor /l/. We are thankful for Dr. Christophe Coupé for providing with the statistical data.

[4]In addition to the change between /n/ and /l/ in the initials in Chinese dialects, one similar change happens to the finals as observed by Prof. Robert Bauer: in Thai, /-l/ is merging with /-n/. That is, words which are spelled with /-l/ in Thai orthography are pronounced with /-n/, such as 'Mahidol' is pronounced as if it were spelled 'Mahidon'.

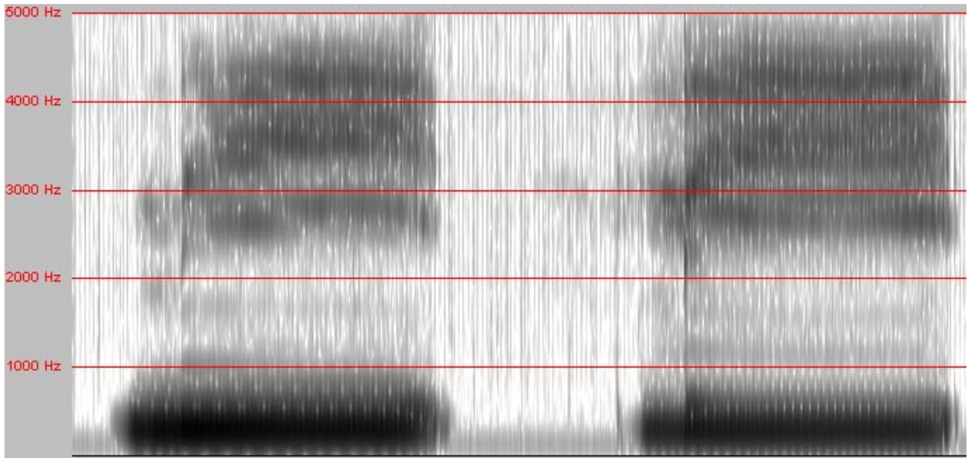


Figure 6.1: Spectrograms illustrating syllables /ni/ (left) and /li/ (right) of an adult female voice.

1994), which refers to the situation that a speaker consistently makes the distinction between two sounds, but they cannot hear the difference themselves. It seems that such cases exist in our Cantonese data. But more systematic perception experiments on such subjects will be necessary in order to confirm their presence. Also it is interesting to know if such cases exist in other dialects.

6.5.2 Psychological reality of the phonological rules

The two types of learners hypothesized here lead us to rethink the explanation for language change and language processing. While linguists are inclined to hypothesize various rules, be it transformational rules or variable rules, etc., it is not clear yet if all these rules have a psychological reality. Some studies have challenged the existence of the abstract rules, such as the tone sandhi rule in Min dialect (Hsieh, 1970). Also the actual form of the rule has been refined progressively, such as the vowel shift rule in English is found to be orthographically conditioned, rather than phonologically conditioned as originally believed (Jaeger, 1984). Even when the rules are shown to exist in some speakers, it is not clear if they exist uniformly in all speakers, as a large degree of heterogeneity exists in the population as we discussed above. Children differ greatly in language acquisition (Shore, 1995) and speakers have different linguistic competence (Fillmore et al., 1979); it is hardly safe to assume all speakers have the same internal representations of the language system.

From our study of the change between /n-/ and /l-/, we speculate that in some, if not all, speakers' mental lexicon, a morpheme with variation may have two phonetic representations, both /n-/ and /l-/. This view is similar to what Bybee (2002) proposes with her exemplar model. What linguists consider as variation between phonemes is in fact the individual alternation of usage of a limited set of morphemes. There is no such rule determining the relations between the two phonemes. Rather, the variation is lexically based, at least in some speakers. Maybe some highly analytical speakers are able to construct an explicit rule of phonological change between phonemes, similar to linguists' analyses. But certainly not all speakers would share a uniform variable rule applicable to the whole set of relevant words.

The existence of grammatical processing "rules" as opposed to "word" has been a main controversy in psycholinguistics and cognitive science, as exemplified by the debate on past-tense processing (Pinker, 1999). With the development of neuroscience techniques (such as Event-related Potential (ERP), functional Magnetic Resonance Imaging (fMRI), and MagnetoEncephaloGraphy (MEG)), this controversy can be better dealt with by probing into the brain. A recent proposal for the dual-mechanism of past tense formation suggests that two types of processing, i.e. the memorization of words in the mental lexicon, and the rule-governed combination of words by the mental grammar, may be rooted with the two types of memories: the mental lexicon depends on declarative memory and is rooted in the temporal lobe, whereas the mental grammar involves procedural memory and is rooted in the frontal cortex and basal ganglia (Ullman, 2001). Also some studies have shown that different brain regions are involved differently for syntactic and semantic processing (e.g., Newman et al., 2001). But this distinction has only been shown above the level higher than lexicon. It is not clear if the processing between words and rules at the level of lexicon and phonology can be identified. If there is a way to detect the activation of a phonological rule during the access of a lexical item, then we may be able to differentiate the two types of VT-III in different speakers: some pairs of variants co-exist as distinctive items in the mental lexicon, some variants are generated from the application of a rule /n-/→/l-/ on some lexical items. We believe that such experiments, which will enable us to zoom in on the actual language representation and processing, will be possible in the near future.

6.5.3 Further development of the model of language change

The model of language change we present in Chapter 5 provides an interesting framework to systematically examine the effect of various language internal and external factors. It invites extensions in many directions. So far we only simulate the diffusion of a single linguistic innovation, and the nature of the innovation is not specified, being a new lexical item, a sound change concerning a set of words, or a grammatical change such as a word order change. The most appropriate correspondence is a new pronunciation of one lexical item. However, a sound change usually affects a number of words. We cannot assume the innovation is a sound change and the agents adopt it or not as a whole, unless we believe the Neogrammarian view that all relevant words change in the same way simultaneously in speakers. Our empirical study of the Cantonese on-going sound change has shown that different words subject to the same change may have different states in one speaker, either unchanged (U), changed (C) or in variation (V). Therefore, the words have to be considered individually. In addition to the phonetic form, each word has a set of properties, including grammatical class, frequency, having homophones or near-homophones or not, being taboo or not, etc. Each of these properties may play a role in determining the word's propensity to change.

In addition to multiple words, we also need to take into account different learning styles in speakers. The two learning styles we proposed in Chapter 3, i.e. categorical and probabilistic learning, may be extended to three types of learners in the multiple-word situation: 1) words are totally independent and each is learned probabilistically based on the frequency of the linguistic input in the environment; 2) the probability to learn a changed form for a word is correlated with the number of words which have already had the changed form; 3) all words are bound together and always learned and changed categorically. We may vary the ratios of these three types of learners in the population to see how the change dynamics differ.

Empirically, it has been found that words starting later may have a faster change rate than early starters, exhibiting a snow-ball effect (Ogura and Wang, 1996; Shen, 1997). Wang et al. (2004) design an analytical model to show the condition for this snow-ball effect to take place. But the analytical model cannot zoom in on the actual diffusion process by taking into account more realistic considerations. Our simulation can be extended to address these issues.

We may also further extend the model to simulate the situation when two competing changes happen at the same time to see how a change is interrupted by the other and what kind of residues will be left (Wang, 1969a). The situation of the bidirectional change such as in our case of Guangzhou Cantonese, where both /n-/→/l-/ and /l-/→/n-/ are competing, will be an interesting case for modeling. The model may predict what is the future of this particular change under various conditions.

The model can be extended to study various situations of language contact. The effect of different population size, contact intensity and contact duration can be quantitatively examined. There are abundant data from historical linguistics on various contact situations. For example, the history of English provide two distinctive situations of language contact: 1) a large number of immigrants with low social impact, as in the settlement of a large number of Scandinavians in the north of England for about 250 years; 2) a small group of immigrants with high prestige, as in the case of French influence on English in about 300 years after the Norman Conquest in 1066 (Ogura and Wang, 2004). The model will be able to manipulate various conditions, such as the ratios of population size between two groups in contact, the social status, etc., to compare these two situations.

Empirical sociolinguistic studies have found the personal networks of individual speakers correlate with their linguistic behavior in response to a change in the community. Two different hypotheses regarding the leader of the change have been proposed. Some suggest that leaders of linguistic change are centrally located in social networks (Labov, 2001), while some remark that marginal members with weak ties within the community are leaders (Milroy, 1980/1987). Simulation models may help to test these different hypotheses. We may also incorporate new findings from recent rapid advancement in network analyses (e.g., Newman and Girvan, 2003) or network modelings (e.g., Ravasz and Barabási, 2003).

6.5.4 More interactions between modeling and empirical studies

Currently, modeling studies of language evolution mostly make use of the existing knowledge from empirical studies and put them directly in the models as built-in assumptions. However, there may exist some reciprocal benefits. As demonstrated by the honeybee comb case in Chapter 1, when the models do not produce the observed pattern in the real world, the failure will lead to new questions for empirical investigations. Moreover, when modelers attempt to make their assumptions as explicit as implementable in a computational model, they need to have valid justifications for these assumptions. Sometimes, some assumptions may appear as trivial and have been taken for granted. But when we force ourselves for justification, we may realize that they are not so trivial.

The use of conjunction in linguistic expression may be given here as a very simple example. It seems very straightforward for us to use a lexical item 'and' in English or '和' in Chinese to join two constituents as one compound entity. In fact this way to express the concept of "coordination" in English and Chinese is only one of the many forms found in languages. There are many other alternatives of expressing coordination and the types of coordinations of constituents differ a lot among languages (Mithun, 1988)[5]. These different ways of expressing coordination may come into languages in different trajectories. Now let's try to imagine how the forms for conjunction in English or Chinese came about. We may assume that in our ancestors' mind that they knew they could create a word to join two parallel things together? Then, what would be the actual process? Suppose after the stage of using simple sentences to express one-degree predicate structures. There may appear one person who accidentally realized that the two subsequent sentences share some similar parts, and then he decided to combine the two sentences in a creative way. There are at least two possible scenarios when we try to zoom in on the process. The first one

[5]For example, in English the following sentence is a good construction: '[Jenny makes] and [Randy sells] the prints.' (Croft, 2001, p189) But such kind of coordination is not possible in Chinese.

follows an analysis of conjoining and deletion given in Wang (1967).

Scenario 1:

1. S1: Tiger come. Lion come.
2. S2: Tiger come *and* lion come. (by conjoining two sentences together with one lexical item)
3. S3: Tiger and lion come. (by deleting the repetitive parts)

We may have an alternative scenario.

Scenario 2:

1. S1: Tiger come. Lion come.
2. S2: Tiger lion come. (by putting two constituents together as a kind of embedding which is assumed as an innate ability in humans)
3. S3: Tiger *and* lion come. (by creating a lexical item 'and' to make the coordination explicit, as sometime it is necessary to disambiguate between "tiger lion" as a type of lion and "tiger and lion" as two animals.)

The appearance of a lexical item 'and' in Step 2 in Scenario 1 and Step 3 in Scenario 2 is not trivial. And the two scenarios require different assumptions. The first scenario assume the creation of a lexical item to signify a coordination relation from scratch can happen readily similar to the creation of any lexical item, while the second scenario assumes there is a need for disambiguation. While we have no evidence so far to show which scenario is more plausible, the second scenario, which assumes function and communication need as the drive, appears to be a little bit more appealing.

In conjecturing these scenarios, we may come up with questions concerning language acquisition. Does the child understand the concept of coordination very early? How old must children be before they understand "and" and use "and"? Are there any variation stages before they can use the conjunction 'and' consistently? How about the acquisition of other languages with very different types of structures and forms for coordination? And what is the process of the development of forms for coordination in creole languages? Moreover, do other animals have such knowledge of the concept 'and' and how do they express it? ...Many questions are triggered by our attempt to formulate a scenario for modeling the emergence of coordination. These questions may stimulate some interesting topics for empirical studies in language acquisition and in other animals' communicative systems.

Our attempt to hypothesize the emergence of coordination structure exemplifies one conventional practice of linguistic investigations - verbal hypothesizing, especially for language origin. Nowadays we have more and more powerful and accessible empirical experimental tools and computational models. It is hoped that our enquiry to how language emerged and evolved would become more precise and quantitative. In fact many students in linguistics have been working in this direction in the various areas as we have discussed in this thesis. Our attempt here is to weave these diverse areas into one coherent picture under the umbrella of self-organization, a paradigm which has been advocated by scientists in other disciplines. As Marcel Proust said, "The real voyage of discovery consists not in seeking new landscapes but in having new eyes". We hope that the theory of self-organization brings new eyes for linguistics, and furthermore, the insights from these new eyes will lead us to seek for more new landscapes.

Bibliography

Aitchison, J. (1996). *The Seeds of Speech: Language Origin and Evolution*. Cambridge University Press, Cambridge.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 71(1):47–97.

Altmann, G. (1983). A law of change in language. *Quantitative Linguistics*, 18:104–115.

Andersen, H. (1978). Perceptual and conceptual factors in abductive change. In Fisiak, J., editor, *Recent Development in Historical Phonology*, pages 1–23. Mouton, The Hague.

Antilla, R. (1989). *Historical and Comparative Linguistics*. John Benjamins, Amsterdam/Philadelphia, 2nd edition.

Arbib, M. (in press). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*.

Arita, T. and Koyama, Y. (1998). Evolution of linguistic diversity in a simple communication system. In Adami, C., Belew, R., Kitano, H., and Taylor, C., editors, *Artificial Life VI*, pages 9–17. MIT Press, Cambridge, MA.

Aronoff, M. (1976). *Word Formation in Generative Grammar (Linguistic Inquiry Monograph, 1)*. MIT Press, Cambridge, MA.

Au, C.-P. and Coupé, C. (2003). A computational model on the emergence of vowel and diphthong perception. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*. Barcelona.

Baayen, H., Piepenbrock, R., and van Rijn, H. (1995). *The CELEX Lexical Database, Release 2 (CD-ROM)*. Linguistic Data Consortium, University of Pennsylvania.

Bailey, C.-J. (1973). *Variation and Linguistic Theory*. Center for Applied Linguistics, Washington, D.C.

Baldwin, J. (1896). A new factor in evolution. *The American Naturalist*, 30:441–451,536–553.

Barabási, A.-L. (2002). *Linked: The New Science of Networks*. Perseus Publishing, Cambridge, Mass.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Barrett, M. (1978). Lexical development and overextension in child language. *Journal of Child Langauge*, 5:205–219.

Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In Brooks, R. and Maes, P., editors, *Artificial Life IV*, pages 160–171. MIT Press, Cambridge, MA.

Batali, J. (1998). Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M., and Knight, C., editors, *Approaches to the Evolution of Language*, pages 405–426. Cambridge University Press, Oxford.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Briscoe, E. J., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Chapter 5. Cambridge University Press.

Bauer, R. S. (1983). Cantonese sound change across subgroups of the Hong Kong speech community. *Journal of Chinese Linguistics*, 11(2):301–354.

Baxter, W. H. (1992). *A Handbook of Old Chinese Phonology*. Mouton De Gruyter.

Belew, R. K., Mitchell, M., and Ackley, D. H. (1996). Computation and the natural sciences. In *Adaptive Individuals in Evolving Populations: Models and Algorithms (eds.)*, Chapter 24. Addison-Wesley, Reading, Mass.

Bhattacharjee, Y. (2003). From heofonum to heavens. *Science*, 303:1326–1328.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Science*, 7:173–221.

Bickerton, D. (1990). *Language and Species*. The University of Chicago Press.

Bickerton, D. (1991). Language origins and evolutionary plausibility. *Language and Communication*, 11(1):37–39.

Blackmore, S. (1999). *The Meme Machine*. Oxford University Press, New York.

Bloch, B. (1948). A set of postulate for phonemic analysis. *Language*, 24:3–46.

Bloomfield, L. (1933). *Language*. The University of Chicago Press.

Blum-Kulka, S. and Snow, C. E. (2002). *Talking to Adults: The Contribution of Multiparty Discourse to Language Acquisition (ed.)*. Lawrence Erlbaum Associates, Mahwah, N.J.

Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, New York; Oxford.

Bourgerie, D. S. (1990). *A Quantitative Study of Sociolinguistic Variation in Cantonese*. PhD thesis, The Ohio State University.

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition*, 61(1-2):1–38.

Brent, M. R. and Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1-2):93–125.

Briscoe, E. J. (2000a). An evolutionary approach to (logistic-like) language change.

Briscoe, E. J. (2000b). Grammatical acquisition: Inductive bias and coevolution of language and the language acquisition device. *Language*, 76(2):245–296.

Briscoe, E. J. (2002). *Linguistic Evolution through Language Acquisition: Formal and Computational Models (ed.)*. Cambridge University Press, Cambridge.

Buchanan, M. (2002). *Nexus: Small Worlds and the Groundbreaking Science of Networks*. W. W. Norton, New York; London.

Burling, R. (2000). Comprehension, production and conventionalization in the origins of language. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, Cambridge.

Bybee, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Benjamins, Philadelphia.

Bybee, J. L. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14:261–290.

Calvin, W. H. (1996). *The Cerebral Code: Thinking a Thought in the Mosaics of the Mind*. MIT Press, Cambridge, MA.

Camazine, S., Deneubourg, J.-L., Franks, N., Sneyd, J., Theraulaz, G., and Bonabeau, E. (2001). *Self-Organization in Biological Systems*. Princeton University Press, Princeton, NJ.

Cangelosi, A., Greco, A., and Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In *Simulating the Evolution of Language*. Springer, London.

Cangelosi, A. and Harnad, S. (2000). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4(1):117–142.

Cangelosi, A. and Parisi, D. (1998). The emergence of a "language" in an evolving population of neural networks. *Connection Science*, 10(2):83–97.

Cangelosi, A. and Parisi, D. (2001). *Simulating the Evolution of Language (ed.)*. Springer-Verlag, London.

Capra, F. (1996). *The Web of Life: A New Scientific Understanding of Living Systems*. Anchor Books, New York.

Caramazza, A. and Berndt, R. (1978). Semantic and syntactic processes in aphasia: A review of the literature. *Psychological Bulletin*, 85:898–918.

Cavalli-Sforza, L. L. and Feldman, M. W. (1981). *Cultural Transmission and Evolution*. Princeton University Press, Princeton, NJ.

Chao, Y.-R. (1934). The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica*, 4(4):363–397. Reprinted in Readings in Linguistics, ed. Martin Joos, p38-54.

Chao, Y.-R. (1947). *Cantonese Primer*. Greenwood Press, New York.

Chao, Y.-R. (1959). Ambiguity in Chinese. *Studia Serica Bernhard Karlgren Dedicata*, pages 1–13.

Chen, B.-Y. (1989). Xitong yanbian de mudixing (系统演变的目的性The teleology of system evolution). *Zhexue Yanjiu (哲学研究)*, 9.

Chen, B.-Y. (1999a). *Ershi Shiji Zhongguo Yuyanxue Fangfa Lun (二十世纪中国语言学方法论Chinese Linguistic Methodology in the 20th Century)*. Shandong Jiaoyu Chubanshe (山东教育出版社), Jinan, Shangdong.

Chen, M. Y.-C. and Wang, W. S.-Y. (1975). Sound change: actuation and implementation. *Language*, 51(1):255–281.

Chen, P. (1999b). *Modern Chinese: History and Sociolinguistics*. Cambridge University Press, Cambridge.

Cheng, C.-C. (1996). Quantifying dialect mutual intelligibility. In Huang, C.-T. J. and Li, Y.-H. A., editors, *New Horizons in Chinese Linguistics*, pages 269–292.

Cheng, C.-C. (1998). Quantification for understanding language cognition. *Quantitative and Computational Studies on the Chinese Language*, pages 15–30.

Cheng, C.-C. and Wang, W. S.-Y. (1972). Tone change in Chaozhou Chinese - a study in lexical diffusion. In Kachru, B. and et al., editors, *Papers in Linguistics in Honor of Henry and Renee Kahane*, pages 99–113. University of Illinois Press. Reprinted in The Lexicon in Phonological Change, William S-Y Wang, ed. Mouton, 1977.

Cheng, X.-Q. (1992). *Xian Qin Hanyu yanjiu (先秦汉语研究Studies of the Pre-Qin Chinese)*. Shandong jiaoyu chubanshe (山东教育出版社).

Cheung, S. H.-N. (1972). *Xianggang Yueyu Yufa de Yanjiu (香港粤语语法的研究Cantonese as Spoken in Hong Kong)*. The Chinese University of Hong Kong, Hong Kong.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (1972). *Language and Mind*. Harcourt Brace Jovanovich, New York.

Chomsky, N. (1976). On the nature of language. In Harnad, S. R., Steklis, H. D., and Lancaster, J., editors, *Origins and Evolution of Language and Speech*, pages 46–57. The New York Academy of Sciences, New York.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. Praeger, New York.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge, MA.

Christiansen, M. (1994). *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistics Structures*. PhD thesis, University of Edinburgh.

Christiansen, M., Allen, J., and Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, 13:221–268.

Christiansen, M. and Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23:157–205.

Christiansen, M. H., Dale, R., Ellefson, M., and Conway, C. (2002). The role of sequential learning in language evolution: Computational and experimental studies. In Cangelosi, A. and Parisi, D., editors, *Simulating the evolution of language*, pages 165–187. Springer-Verlag, London.

Christiansen, M. H. and Devlin, J. T. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, pages 113–118. Lawrence Erlbaum Associates, Mahwah, NJ.

Christiansen, M. H. and Kirby, S. (2003). *Language Evolution: The States of the Art (eds)*. Oxford University Press.

Cihui (1995). *Hanyu Fangyan Cihui (汉语方言词汇Chinese Dialect Word List)*. Beijing Daxue Zhongguo Yuyan Wenxue Xi Yuyanxue Jiaoyanshi, Yuwen Chuban She (北京大学中国语言文学系语言学教研室,语文出版社), 2nd edition.

Clark, E. V. (1987). The Principle of Contrast: A constraint on language acquisition. In MacWhinney, B., editor, *Mechanisms of Language Acquisition*, pages 1–33. Erlbaum, Hillsdale, NJ.

Clark, E. V. (1993). *The Lexicon in Acquisition*. Cambridge University Press, Cambridge, England.

Coates, W. A. (1969). Near-homophony as a factor in language change. *Language*, 44(3):467–479.

Comrie, B. (1981). *Language Universals and Linguistic Typology: Syntax and Morphology*. Blackwell, Oxford.

Comrie, B. (2000). Language contact, lexical borrowing, and semantic fields. *Studies in Slavic and General Linguistics*, 28:73–86.

Coupé, C. (2003). *From the Origin of Language to the Origin of Languages: Modelling of the Emergence and Evolution of Linguistic Systems*. PhD thesis, Universite de Lyon.

Croft, W. (1990). *Typology and Universals*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach*. Longman, Harlow, England; New York.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press, Oxford; New York.

Crothers, J. (1978). Typology and universals of vowel systems. In Greenberg, J. H., editor, *Universals of Human Language (Phonology)*, volume 2, pages 93–152. Stanford University Press, Stanford, CA.

Curtiss, S. (1977). *Genie: A Psycholinguistic Study of a Modern-day "Wild Child"*. Academic Press, New York.

Cutler, A., Hawkins, J., and Gilligan, G. (1985). The suffixing preference: A processing explanation. *Linguistics*, 23:723–758.

Cuyckens, H. and Zawada, B. E. (2001). Introduction. In Cuyckens, H. and Zawada, B., editors, *Polysemy in Cognitive Linguistics : Selected Papers from the International Cognitive Linguistics Conference, Amsterdam, 1997*, pages ix–xxiii. John Benjamins, Amsterdam; Philadelphia.

Dai, X. J. (1990). Historical morphonologization of syntactic words: Evidence from Chinese derived verbs. *Diachronica*, 7(1):9–46.

Darwin, C. (1871). *Descent of Man.* Prometheus Books, Amherst, N.Y.

Dauzat, A. (1927). *Les Patois.* Paris.

Dawkins, R. (1976). *The Selfish Gene.* Oxford University Press.

de Boer, B. (1997). Generating vowel systems in a population of agents. In Husbands, P. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 503–510. MIT Press, Cambridge, MA.

de Boer, B. (2000). Emergence of sound systems through self-organisation. In Knight, C., Studdert-Kennedy, M., and Hurford, J. R., editors, *The Evolutionary Emergence of Language*, pages 177–198. Cambridge University Press, Cambridge; New York.

de Boer, B. (2001). *The Origins of Vowel Systems.* Oxford University Press.

de Saussure, F. (1983). *Course in General Linguistics (Cours de linguistique generale).* Open Court, LaSalle, IL.

de Waal, F. (1998). *Chimpanzee Politics: Power and Sex among Apes.* Johns Hopkins University Press, Baltimore.

Deacon, T. (1997). *The Symbolic Species.* W. Norton and Co., New York.

Denison, D. (2003). Log(ist)ic and simplistic s-curves. In Hickey, R., editor, *Motives for Language Change*, pages 54–70. Cambridge University Press, Cambridge; New York.

Derwing, B. L. and Baker, W. J. (1977). The psychological basis for morphological rules. In Macnamara, J., editor, *Language Learning and Thought*, pages 85–110. Academia Press, New York.

Dixon, R. M. W. (1997). *The Rise and Fall of Languages.* Cambridge University Press, Cambridge; New York.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, 68:81–138.

Du Bois, J. (1987). The discourse basis of ergativity. *Language*, 63:805–855.

Duanmu, S. (2000). *The Phonology of Standard Chinese.* Oxford University Press, Oxford; New York.

Dugatkin, L. A. (2000). *The Imitation Factor: Evolution beyond the Gene.* Free Press, New York.

Dunbar, R. (1993). Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16:681–735.

Eckert, P. (2000). *Linguistic Variation as Social Practice : The Linguistic Construction of Identity in Belten High*. Blackwell Publishers, Oxford, Malden, Mass.

Ellefson, M. R. and Christiansen, M. H. (2000). Subjancy constraints without universal grammar: Evidence from artificial language learning and connectionist modeling. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pages 645–650. Lawrence Erlbaum Associates, Mahwah, NJ.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14:179–211.

Elman, J., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1998). *Rethinking Innateness - A Connectionist Perspective on Development*. Bradfork Book/MIT Press.

Empson, H. (1992). *Mapping Hong Kong: A Historical Atlas*. Hong Kong Zhengfu yinwu ju 香港政府印务局, Hong Kong.

Erdös, P. and Rènyi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297.

Fan, G. and et al. (1997). *Yueyu Pinyin Zibiao (粤语拼音字表A syllabary of Cantonese)*. Hong Kong Linguistic Society, Hong Kong.

Feng, S. (1995). *Prosodic Structure and Prosodically Constrained Syntax in Chinese*. University Microfilms, Inc., Ann Arbor, MI.

Feng, S. (1998). Prosodic structure and compound words in Classical Chinese. In Packard, J., editor, *New approaches to Chinese Word Formation*, pages 197–260. Mouton de Gruyter, Berlin.

Ferguson, C. A. (1979). Phonology as an individual access system: Some data from language acquisition. In Fillmore, C. J., Kempler, D., and Wang, W. S.-Y., editors, *Individual Differences in Language Ability and Language Behavior*, pages 189–201. Academic Press.

Ferguson, C. A. and Farwell, C. B. (1975). Words and sounds in early language acquisition: English initial consonants in the first fifty words. *Language*, 51:419–439.

Ferrand, L. and Grainger, J. (2003). Homophone interference effects in visual word recognition. *The Quarterly Journal of Experimental Psychology: Section A*, 56(3):403 – 419.

Ferrer i Cancho, R. and Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Science*, 100(3):788–791.

Fillmore, C. J., Kempler, D., and Wang, W. S.-Y. (1979). *Individual Differences in Language Ability and Language Behavior (eds.)*. Academic Press.

Fischer, J. L. (1958). Social influences on the choice of a linguistic variant. *Word*, 14:47–56.

Fitch, W. T. (2002). Comparative vocal production and the evolution of speech: Reinterpreting the descent of the larynx. In Wray, A., editor, *The Transition to Language*, Chapter 2. Oxford University Press, Oxford.

Fouts, R. (1991). Dirty bathwater, innateness neonates and the dating game. *Language and Communication*, 11(1/2):41–43.

Frommer, I. and Pundoor, G. (2003). Small-words: A review of recent books. *Networks*, 41(3):174–180.

Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science*, 167:86–87.

Gallup, G. G. J., Anderson, J. R., and Shillito, D. J. (2002). The mirror test. In Bekoff, M., Allen, C., and Burhgardt, G. M., editors, *The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition*. MIT Press, Cambridge, MA.

Ghazanfar, A. A. and Logothetis, N. K. (2003). Neuroperception: Facial expressions linked to monkey calls. *Nature*, 423:937 – 938.

Gillièron, J. and Roques, M. (1912). *Ètudes de Gèographie Linguistique d'apres l'Atlas Linguistique de la France*. Paris.

Goldstein, M. H., King, A. P., and West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Science*, 100(13):8030–8035.

Gong, T., Ke, J.-Y., Minett, J. W., and Wang, W. S.-Y. (2004). Language emergence and simple social structure. In *Proceedings of The 5th Conference on Language Evolution*. Leipzig, Germany.

Gould, J. L. and Marler, P. (1987). Learning by instinct. *Scientific American*, pages 74–85.

Gould, S. J. and Eldredge, N. (1977). Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology*, 3:115–151.

Gould, S. J. and Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London*, 205:281–188.

Gould, S. J. and Vrba, E. S. (1982). Exaptation – a missing term in the science of form. *Paleobiology*, 8:4–15.

Greenberg, J. H. (1963). *Universals of Language (ed.)*. MIT Press, Cambridge, MA.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford University Press, Stanford.

Guo, S. (1938). Zhongguo yuci zhi tanxing zuoyong (中国语词之弹性作用the elastic function of Chinese word length). *Yanjing Xuebao (燕京学报)*, 24.

Guo, X.-L. (1980). Hanyu disan rencheng daici de qiyuan he fazhan (汉语第三人称代词的起源和发展the origin and development of the third person pronoun in Chinese). *Yuyanxue Luncong (语言学论丛)*, 6. Reprinted in Hanyushi Lunji, Shangwu Yishuguan, 1997.

Guy, G. R. (1980). Variation in the group and the individual. In Labov, W., editor, *Locating Language in Time and Space*, pages 1–36. Academic Press, New York.

Haegeman, L. (1991). *Introduction to Government and Binding Theory*. B. Blackwell, Oxford, 1st edition.

Haiman, J. (1983). Iconic and economic motivation. *Language*, 59:781–819.

Haken, H. (1984). *Laser Theory*. Springer-Verlag, Berlin.

Hankamer, J. (1973). Unacceptable ambiguity. *Linguistic Inquiry*, 4:17–68.

Hare, M. and Elman, J. (1995). Learning and morphological change. *Cognition*, 56:61–98.

Hashimoto, O.-K. Y. (1972). *Phonology of Cantonese*. Cambridge University Press, London.

Hauser, M. and Fitch, W. (2003). What are the uniquely human components of the language faculty? In Christiansen, M. H. and Kirby, S., editors, *Language Evolution: The States of the Art*. Oxford University Press, Oxford.

Hauser, M., Newport, E. L., and Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: statistical learning in cotton-top tamarins. *Cognition*, 78:B53–B64. mokey can keep track of relative serial order among pattern elements, and also the frequency or consistency with which various combinations and orders appear.

Hauser, M., Weiss, D., and Marcus, G. (2002a). Rule learning by cotton-top tamarins. *Cognition*, 86:B15.

Hauser, M. D. (1996). *The Evolution of Communication*. MIT press, Cambridge, MA.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002b). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298:1569–1579.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.

Hawkins, J. A. and Gell-Mann, M. (1992). *The Evolution of Human Languages (eds.)*. Addison-Wesley, Reading, MA.

Heyes, C. and Huber, L. (2000). *The Evolution of Cognition (ed.)*. MIT Press, Cambridge, MA.

Higgins, J. (1995). Quantifying english homophones and minimal pairs. In Lewis, J. W., editor, *Studies in General and English Phonetics: Essays in Honor of Professor J.D. O'Connor*, pages 326–334. Routledge, London.

Hinton, G. and Nowlan, S. (1987). How learning guides evolution. *Complex Systems*, 1:495–502.

Ho, D.-A. (1988). *Rules and Directions: The Changing Structure of Phonology*. Institute of History and Philology, Academia Sinica, Taiwan. Special Publications no. 90, in Chinese.

Hockett, C. F. (1961). Linguistic elements and their relations. *Language*, 37:29十53.

Hockett, C. F. (1968). *The State of the Art*. Mouton, The Hague, Paris.

Holland, J. H. (1998). *Emergence: From Chaos to Order*. Addison-Wesley, Reading, MA.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W., editor, *Current Progress in Historical Linguistics*, pages 96–105. North Holland, Amsterdam.

Hooper, J. B. (1980). Child morphology and morphophonemic change. In Fisiak, J., editor, *Historical Morphology*, pages 157–187. Mouton, Berlin.

Hopper, P. (1988). Emergent grammar and the a priori grammar postulate. In Tannen, D., editor, *Linguistics in Context: Connecting Observation and Understanding. Advances in Discourse Processes, Volume 29*, pages 117–134. Ablex Publishing, Norwood, NJ.

Hou, J.-Y. (2002). *Xiandai Hanyu Fangyan Gailun (*现代汉语方言概论*A Conspectus on Modern Chinese Dialects)*. Shanghai Jiaoyu Chubanshe & Shanghai Shiji Chuban Jituan (上海教育出版社&上海世纪出版集团), Shanghai.

Houde, J. F. and Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, 279:1213–1216.

Hsieh, H.-I. (1970). The psychological reality of tone sandhi rules in Taiwanese. In *Papers from the Sixth Meeting of the Chicago Linguistic Society*, pages 489–503.

Hsieh, H.-I. (1972). Lexical diffusion: Evidence from child language acquisition. *Glossa*, 6:89–104.

Hurford, J. R. (1987). *Language and Number: The Emergence of a Cognitive System.* Basil Blackwell, Oxford.

Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language evolution device. *Lingua*, 77:187–222.

Hurford, J. R. (1990). Nativist and functional explanations in language acquisition. In Roca, I., editor, *Logical Issues in Language Acquisition*, pages 85–136. Dordrecht, Foris, Holland.

Hurford, J. R. (1999a). The evolution of language and languages. In Dunbar, R., Knight, C., and Power, C., editors, *The Evolution of Culture*, pages 173–193. Edinburgh University Press, Edinburgh.

Hurford, J. R. (1999b). Functional innateness: Explaining the critical period for language acquisition. In Darnell, M., Moravscik, E., Newmeyer, F., Noonan, M., and Wheatley, K., editors, *Functionalism and Formalism in Linguistics, Volume II: Case Studies*, pages 341–363. John Benjamins, Amsterdam.

Hurford, J. R. (2003). Why synonymy is rare: Fitness is in the speaker. In Banzhaf, W., Christaller, T., Dittrich, P., Kim, J. T., and Ziegler, J., editors, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL), Lecture Notes in Artificial Intelligence, Vol. 2801*, pages 442–451. Springer-Verlag, Berlin, Heidelberg.

Hurford, J. R., Studdert-Kennedy, M., and C., K., editors (1998a). *Fitness and the selective adaptation of language*, Cambridge. Cambridge University Press.

Hurford, J. R., Studdert-Kennedy, M., and Knight, C. (1998b). *Approaches to the Evolution of Language (ed.).* Cambridge University Press, Oxford.

Hutchins, E. and Hazlehurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In Gilbert, G. N. and Conte, R., editors, *Artificial Societies: The Computer Simulation of Social Life*. UCL Press, London.

Hutchins, E. and Hazlehurst, B. (2001). Auto-organization and emergence of shared language structure. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, Chapter 13, pages 279–306. Springer-Verlag, London.

Jackendoff, R. (1999). Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences*, 3(7):272–279.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution.* Oxford University Press, Oxford; New York.

Jacob, F. (1977). Evolution and tinkering. *Science*, 196(4295):1161–1166.

Jaeger, J. (1984). Assessing the psychological status of the vowel shift rule. *Journal of Psychological Research*, 13:13–36.

Jakobson, R. (1960). Linguistics and poetics. In Sebeok, T. A., editor, *Style in Language*. MIT Press, Cambridge, MA. reprinted in Selected writings III, 18-51, Mouton, The Hague, 1971.

Janik, V. (1999). Origins and implications of vocal learning in bottlenose dolphins. In Box, H. and Gibson, K., editors, *Mammalian Social Learning: Comparative and Ecological Perspectives*, pages 308–326. Cambridge University Press, Cambridge.

Jantsch, E. (1981). *The Evolutionary Vision: Toward a Unifying Paradigm of Physical, Biological and Sociocultural Evolution*. MIT Press, Cambridge, MA.

Jerison, H. (1977). The theory of encephalization. In Diamond, S. and Blizard, P., editors, *Evolution and Lateralization of the Brain. Annals of The New York Academy of Sciences, vol. 299*, pages 601–606. New York Academy of Sciences.

Jespersen, O. (1922). *Language: Its Nature, Development and Origin*. G. Allen & Unwin, London, 1st edition.

Jespersen, O. (1933). Monosyllabism in english. In *Selected Writings of Otto Jespersen*, pages 617–641. Senjo Publishing, Tokyo. Binnial Lecture on English Philology, in the British Academy, Nov. 6, 1928.

Jiang, J.-C. (1991). Lun jindai hanyu de shangxian (论近代汉语的上限). *Gu Hanyu Yanjiu (古汉语研究)*, 2.

Jowett, B. (1953). *The Dialogues of Plato*. Oxford University Press, Oxford, 4th edition.

Jurafsky, D., Bell, A., and Girand, C. (2002). The role of the lemma in form variation. In Gussenhoven, C. and Warner, N., editors, *Papers in Laboratory Phonology VII*, pages 1–34. Mouton de Gruyter, Berlin/New York.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. MIT Press, Cambridge, MA.

Kant, I. (1952). *The Critique of Judgement*. Clarendon Press, Oxford.

Karlgren, B. (1937). *Études sur la Phonologie Chinoise (Zhongguo yinyunxue yanjiu 中国音韵学研究A study on Chinese phonology)*. Shangwu Yinshu Guan (商务印书馆), Shanghai. translated by Zhao Yuanren, Luo Changpei and Li Fanggui, reprintee in Beijing, 1994.

Karlgren, B. (1949). *The Chinese Language: An Essay on its Nature and History*. The Ronald Press Company, New York.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, New York.

Kauffman, S. A. (1995). *At Home in the Universe*. Oxford University Press, New York.

Ke, J.-Y. (2003). Complex dynamic networks and human languages. Seminar in Centre for Chaos Control and Synchronization, Department of Electronic Engineering, City University of Hong Kong, April 17, 2003.

Ke, J.-Y., Au, C. P., Minett, J. W., and Wang, W. S.-Y. (2002). Self-organization and selection in the emergence of vocabulary. *Complexity*, 7(3):41–54.

Ke, J.-Y., Gong, T., and Wang, W. S.-Y. (2004). Language change and social networks. In *The 5th Conference on Language Evolution*.

Ke, J.-Y., Ogura, M., and Wang, W. S.-Y. (2003). Optimization models of sound systems using Genetic Algorithms. *Computational Linguistics*, 29(1):1–18.

Keller, R. (1994). *On Language Change: The Invisible Hand in Language*. Routledge, London.

Kelly, M. H. and Ragade, A. (2000). Grammatical relationships between homonyms: Effects on language comprehension and the structure of the English vocabulary.

Kennedy, G. A. (1951). The monosyllabic myth. *Journal of American Oriental Society*, 71:161–166.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Kiparsky, P. (1968). Linguistic universals and linguistic change. In Bach, E. and Harms, R. T., editors, *Universals in Linguistic Theory*, pages 171–202. Holt, Rinehart and Winston, New York.

Kirby, S. (1999a). *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford University Press, New York.

Kirby, S. (1999b). Syntax out of learning: The cultural evolution of structured communication in a population of induction algorithms. In Floreano, D., Nicoud, J.-D., and Mondada, F., editors, *Advances in Artificial Life - Lecture Notes in Computer Science 1674*. Springer.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, Oxford.

Kirby, S. (2001). Spontaneous evolution of linguistic structure - an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, 5(2):102–110.

Kirby, S. (2002a). Learning, bottlenecks and the evolution of recursive syntax. In Briscoe, E. J., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Chapter 6. Cambridge University Press.

Kirby, S. (2002b). Natural language from artificial life. *Artificial Life*, 8(2):185–215.

Klein, R. G. (1999). *The Human Career, Human Biological and Cultural Origins*. The University of Chicago Press.

Kluender, K., Diehl, R., and Killeen, P. (1987). Japanese quail can learn phonetic categories. *Science*, 237:1195–1197.

Knight, C., Studdert-Kennedy, M., and Hurford, J. R. (2000). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form (eds.)*. Cambridge University Press, Cambridge; New York.

Köhler, W. (1947). *Gestalt Psychology: An Introduction to New Concepts in Modern Psychology*. Liveright Pub, New York, revision edition.

Kohonen, T. (1995). *Self-Organization Maps*. Springer-Verlag, Berlin; Heidelberg.

Komarova, N. and Niyogi, P. (2004). Optimizing the mutual intelligibility of linguistic agents in a shared world. *Journal of Artificial Intelligence*, 154(1-2):1–42.

Komarova, N. L. and Nowak, M. A. (2001). Evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology*, 63(3):451–485.

Kosaka, J. (1997). *Hakuwa goi no kenkyu (A Study on the Lexicon of Baihua)*. Zhonghua shuju (中华书局), Beijing. Baihua yuhui yanjiu (白话词汇研究), translated by Jiang Lansheng (江蓝生) and Bai Weiguo (白维国).

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1(3):199–244.

Kroch, A. (1994). Morphosyntactic variation. In Beals, K. e. a., editor, *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*. Chicago Linguistics Society, Chicago.

Kuhl, P. and Miller, J. (1975). Speech perception by the chichilla: voiced-voiceless distinction in alveolar plosive consonants. *Science*, 190:69–72.

Kuhl, P. K. and Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics*, 32(6):542–550.

Labov, W. (1966). *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington, D.C.

Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language*, 45(4):715–762.

Labov, W. (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.

Labov, W. (1989). The exact description of the speech community: Short a in philadelphia. In Fasold, R. and Schiffrin, D., editors, *Language Change and Variation*, pages 1–57. Georgetown University Press, Washington, D.C.

Labov, W. (1994). *Principles of Linguistic Change (I)*. Blackwell, Oxford, U.K.

Labov, W. (2001). *Principles of Linguistic Change (II)*. Blackwell, Oxford, U.K.

Labov, W., Karen, M., and Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, 3:33–74.

Ladefoged, P. and Maddieson, I. (1996). *The Sounds of the World's Languages*. Blackwell, Oxford.

Lass, R. (1980). *On Explaining Language Change*. Cambridge University Press, Cambridge.

Lee, W. H. and Murashima, K. (2002). Phonetic variation in loanwords: n-/r- variation in Japanese of Hong Kong Cantonese speakers. *Journal of the Phonetic Society of Japan*, 6(2):98–104.

Leonard, L., Newhoff, M., and Mesalam, L. (1980). Individual differences in early child phonology. *Applied Psycholinguistics*, 1:7–30.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Li, C. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press, Berkeley and Los Angeles.

Li, N. (1990). Dongci fenlei yanjiu shuolue (动词分类研究说略a note on the catgorization of verbs). *Zhongguo Yuwen (*中国语文*)*, 4:248–257.

Li, P. (1999). Generalization, representation, and recovery in a self-organizing feature-map model of language acquisition. In Hahn, M. and Stoness, S., editors, *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, pages 308–313. Lawrence Erlbaum, Mahwah, NJ.

Li, P. (2003). Language acquisition in a self-organising neural network model. In Quinlan, P., editor, *Connectionist Models of Development: Developmental Processes in Real and Artificial Neural Networks*, pages 115–149. Psychology Press, Hove; Briton.

Li, W. Z. (2001). *Songdai Yuyan Yanjiu (宋代语言研究A Study on the Languages in the Song Dynasty).* Xianzhuang Shuju 线装书局, Beijing.

Li, X. and Chen, G. R. (2003). A local-world evolving network model. *Physica A*, 328(1-2):274–286.

Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of language. *Language and Speech*, 6:172–187.

Lieberman, P. (1984). *The Biology and Evolution of Language.* Harvard University Press, Cambridge, MA.

Lightfoot, D. (1991). Subjacency and sex. *Language and Communication*, 11(1):67–69.

Lightfoot, D. (1999). *The Development of Language: Acquisition, Change and Evolution.* Blackwell: Oxford.

Lightfoot, D. (2000). The spandrels of the linguistic genotype. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form.* Cambridge University Press, Cambridge.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48:839–862.

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Aberg, Y. (2001). The web of human sexual contacts. *Nature*, 411:907–908.

Lindblom, B. (1983). Economy of speech gestures. *The Production of Speech*, pages 217–247.

Lindblom, B. (1986). Phonetic universals in vowel systems. *Experimental Phonology*, pages 13–44.

Lindblom, B. (1998). Systemic constraints and adaptive change in the formation of sound structure. In Hurford, J. R., Studdert-Kennedy, M., and Knight, C., editors, *Approaches to the Evolution of Language - Social and Cognitive Bases*, pages 242–264. Cambridge University Press, Cambridge.

Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In Butterworth, B., Bernard, C., and Dahl, O., editors, *Explanations for Language Universals*, pages 181–203. Walter de Gruyter & Co.

Livingstone, D. (2000). A modified-neutral theory for the evolution of linguistic diversity. In *The 3rd International Conference on the Evolution of Language*. Springer-Verlag, Paris.

Livingstone, D. (2001). The evolution of dialect diversity. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, Chapter 5, pages 99–118. Springer-Verlag, London.

Livingstone, D. and Fyfe, C. (1999). Modelling the evolution of linguistic diversity. In D. Floreano, J. N. and Mondada, F., editors, *European Conference of Artificial Life 1999*, pages 704–708. Springer-Verlag, Berlin.

Lu, B. and Duanmu, S. (1991). A case study of the relation between rhythm and syntax in Chinese. In *The Third North America Conference on Chinese Linguistics*. Ithaca.

Lü, S. X. (1963). Xiandai hanyu shuangyin jie wenti chutan (现代汉语双音节问题初探a primary exploration to the question of disyllabification in Chinese). *Zhongguo Yuwen (*中国语文*)*, 1:11–23.

Lü, S. X. (1979). *Hanyu Yufa Fenxi Wenti (*汉语语法分析问题*The problems in the analyses of Chinese Grammar)*. Shangwu Yinshuguan (商务印书馆), Beijing.

Ma, Z. (1981). Xianqin fuyin ci chutan (先秦复音词初探a primary exploration to disyllabic words in pre-qin). *Beijing Daxue Xuebao (*北京大学学报*)*, 1.

MacWhinney, B. (1995). *The CHILDES Project: Tools for Analyzing Talk*. L. Erlbaum Associates, Hillsdale, N.J., 2nd edition.

Maddieson, I. and Precoda, K. (1990). Updating upsid. *UCLA Working Papers in Phonetics*, 74:104–111.

Malkiel, Y. (1967). Every word has a history of its own. *Glossa*, 1:137–149.

Malkiel, Y. (1979). Problems in the diachronic differentiation of near-homophones. *Language*, 55:1–36.

Marcus, G. F., Vijayan, S., Rao, S. B., and Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, 283(5398):77–80.

Markman, E. M. (1989). *Categorization and Naming in Children: Problems of Induction*. MIT Press, Cambridge, MA.

Martinet, A. (1952). Function, structure, and sound change. *Word*, 1:1–32.

Masini, F. (1993). *The Formation of Modern Chinese Lexicon and Its Evolution toward a National Language: The Period from 1840 to 1898*. Journal of Chinese Linguistics, California.

Mathesius, V. (1911). O potenciálnosti jevȯ jazykových. In Vachek, J., editor, *A Prague School Reader in Linguistics (1964)*, pages 1–32. Indiana University Press, Bloomington.

Matthei, E. H. (1989). Crossing boundaries: More evidence for phonological constraints on early multi-word utterances. *Journal of Child Language*, 16:41–54.

McClelland, J. L. and Plaut, D. C. (1999). Does generalization in infant learning implicate abstract algebra-like rules. *Trends in Cognitive Sciences*, 3:166–168.

McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Journal of the Acoustical Society of America*, 264:746–748.

McMahon, A. (1994). *Understanding Language Change*. Cambridge University Press, Cambridge.

McShane, J. (1980). *Learning to Talk*. Cambridge University Press, Oxford.

Meillet, A. (1903). Review of michel bréal, essai de sémantique: science de significations, 3e édition revue, corrigée et augmentée. *L'Année Sociologique*, 8:640–641.

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Development Psychology*, 31:838–850.

Meltzoff, A. N. (1996). The human infant as imitative generalist: A 20-year progress report on infant imitation with implications for comparative psychology. In Heyes, C. M. and Galef, B. G., editors, *Social Learning in Animals: The Roots of Culture*, pages 347–370. Academia Press.

Menn, L. and Matthei, E. H. (1992). The "two lexicon" account of child phonology: Looking back and looking ahead. In Ferguson, C. A., Menn, L., and Stoel-Gammon, C., editors, *Phonological Development: Models, Research, Implications*. York Press, Timonium, MD.

Miller, G. A. and Gildea, P. (1987). How children learn words. *Scientific American*, September.

Milroy, L. (1987). *Language and Social Networks*. Blackwell, Oxford, New Yrok, second edition.

Minett, J. W. and Wang, W. S.-Y. (2003). Language death and maintenance. Manuscript, Language Engineering Laboratory, City University of Hong Kong.

Mithun, M. (1988). The grammaticization of coordination. In Haiman and Thompson, editors, *Clause Combining in Grammar and Discourse*, pages 331–359. John Benjamins, Amsterdam.

Moore, B. R. (1996). The evolution of imitative learning. In Heyes, C. M. and Galef, B. G., editors, *Social Learning in Animals: The Roots of Culture*, pages 246–265. Academia Press.

Morgan, M. S. (1999). Learning from models. In Morgan, M. S. and Morrison, M., editors, *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge University Press, Cambridge; New York.

Mufwene, S. S. (2001). *The Ecology of Language Evolution*. Cambridge University Press, Cambridge; New York.

Mufwene, S. S. (2002). Competition and selection in language evolution. *Selection*, 3(1):45–56.

Munroe, S. and Cangelosi, A. (2003). Learning and the evolution of language: The role of cultural variation and learning costs in the baldwin effect. *Artificial Life*, 8(4):311–339.

Nelson, K. (1973). *Structure and Strategy in Learning to Talk*. Society for Research in Child Development. Monographs 38, 1-2, Serial No. 149. Reprinted in part in Mussen, Conger and Kagan, Readings in child development and personality, (3rd ed.). Harper & Row, 1975.

Nettle, D. (1995). Segmental inventory size, word length, and communication efficiency. *Linguistics*, 33:359–367.

Nettle, D. (1998). Coevolution of phonology and the lexicon in twelve languages of West Africa. *Journal of Quantitative Linguistics*, 5(3):240–245.

Nettle, D. (1999a). Is the rate of linguistic change constant? *Lingua*, 108:119–136.

Nettle, D. (1999b). *Linguistic Diversity*. Oxford University Press, Oxford.

Nettle, D. (1999c). Using Social Impact Theory to simulate language change. *Lingua*, 108:95–117.

Newman, A., Pancheva, R., Ozawa, K., Neville, H., and Ullman, M. (2001). An event-related fmri study of syntactic and semantic violations. *Journal of Psycholinguistic Research*, 30(3):339–364.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45:167–256.

Newman, M. E. J. and Girvan, M. (2003). Mixing patterns and community structure in networks. In Pastor-Satorras, R., Rubi, J., and Diaz-Guilera, A., editors, *Statistical Mechanics of Complex Networks*. Springer-Verlag, Berlin.

Newman, M. E. J. and Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(036122).

Newmeyer, F. J. (1991). Functional explanation in linguistics and the origins of language. *Language and Communication*, 11(1):3–28.

Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, 79:682–707.

Newport, E. L. and Aslin, R. N. (2004). Learning at a distance: I. statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48:127–162.

Newport, E. L., Hauser, M. D., Spaepen, G., and Aslin, R. N. (2004). Learning at a distance: Ii. statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*. in press.

Nichols, J. (1992). *Linguistic Diversity in Space and Time*. University of Chicago Press, Chicago and London.

Nishimura, T., Mikami, A., Suzuki, J., and Matsuzawa, T. (2003). Descent of the larynx in chimpanzee infants. *Proceedings of the National Academy of Science*, 100(12):6930–6933.

Niyogi, P. (2002). Theories of cultural evolution and their application to language evolution. In Briscoe, E. J., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Chapter 7. Cambridge University Press.

Niyogi, P. and Berwick, R. C. (1997). Evolutionary consequences of language learning. *Journal of Linguistics and Philosophy*, 20:697–719.

Niyogi, P. and Surendran, D. (2003). Measuring the functional load of phonological contrasts. Technical report, Chicago University.

Nowak, A., Szamrej, J., and Latané, B. (1990). From private attitude to public opinion: A dynamical theory of social impact. *Psychological Review*, 97:362–376.

Nowak, M., Plotkin, J. B., and Krakauer, D. C. (1999). The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162.

Nowak, M. A., Komarova, N. L., and Niyogi, P. (2001). The evolution of universal grammar. *Science*, 291:114–118.

Ogura, M. (1993). The development of periphrastic do in English, a case of lexical diffusion in syntax. *Diachronica*, X:51–85.

Ogura, M. and Wang, W. S.-Y. (1996). Snowball effect in lexical diffusion: The development of -s in the third person singular present indicative in English. In Britton, D., editor, *English Historica Linguistics 1994: Papers from the 8th International Congress on English Historical Linguistics*, pages 119–141. John Benjamins, Amsterdam.

Ogura, M. and Wang, W. S.-Y. (2004). Dynamic dialectology and complex adaptive system. In Dossena, M. and Lass, R., editors, *Methods and Data in English Historical Dialectology*. Peter Lang, Bern.

Ohala, J. (1993). The phonetics of sound change. In Jones, C., editor, *Historical Linguistics: Problems and Perspectives*, pages 237–278. Longman, London.

Oliphant, M. (1997). *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language*. PhD thesis, Department of Cognitive Science, University of California, San Diego.

Oliphant, M. and Batali, J. (1997). Learning and the emergence of coordinated communication. *The Newsletter of the Center for Research in Language*, 11(1).

Olson, D. R. (1977). The formalization of linguistic rules. In Macnamara, J., editor, *Language Learning and Thought*, pages 111–116. Academia Press, New York.

Oudeyer, P.-Y. (2002). Phonemic coding might result from sensory-motor coupling dynamics. In Hallam, B., Floreano, D., Hallam, J., Hayes, G., and Meyer, J. A., editors, *Proceedings of the 7th International Conference on the Simulation of Adaptive Behavior*, pages 406–416. MIT Press, Edinburgh.

Oudeyer, P.-Y. (2003). *L'auto-organisation de la Parole*. PhD thesis, University Paris VI.

Pan, P. G. (1981). Prestige forms and phonological variation in Hong Kong Cantonese speech. Master thesis, University of Hong Kong.

Parisi, D. (1997). An artificial life approach to language. *Brain and Language*, 59(1):121–146.

Paul, H. (1880). *Prinzipien der Sprachgeschichte*. Niemeyer, Halle.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60:320–342.

Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: from "learning" to parameter setting in biology and the study of language. *Cognition*, 31:1–44.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. L. and Hopper, P., editors, *Frequency and the Emergence of Linguistic Structure (eds.)*. Benjamins, Amsterdam.

Pine, J. M. (1994). The language of primary caregivers. In Galloway, C. and Richards, B. J., editors, *Input and Interaction in Language Acquisition*, pages 15–37. Cambridge University Press, Cambridge.

Pinker, S. (1994). *The Language Instinct*. William Morrow, New York.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language.* Basic Books, New York.

Pinker, S. and Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, 13:707–784.

Pinker, S. and Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28:73–193.

Pinker, S. and Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11):456–463.

Plunkett, K. and Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perception: imiplications for child language acquistion. *Cognition*, 38:43–102.

Popper, K. R. (1980). *The Logic of Scientific Discovery.* Hutchinson, London.

Premack, D. and Premack, A. (2003). *Original Intelligence: Unlocking the Mystery of Who We Are.* McGraw-Hill, New York ; London.

Premack, D. and Woodruff, G. (1978). Does the chimpanzee have a "theory of mind?". *Behavioral and Brain Sciences*, 4:515–526.

Pribram, K. H. (1981). *Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology.* Brandon House, New York.

Pribram, K. H. and King, J. (1996). *Learning as Self-organization.* L. Erlbaum Associates, Mahwah, N.J.

Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences.* W. H. Freeman, New York.

Pulvermüller, F. (1998). On the matter of rules: past tense formation and its significance for cognitive neuroscience. *Network: Computational Neural Systems*, 9:R1–R52.

Quine, W. V. O. (1960). *Word and Object.* Harvard University Press, Cambridge, Massachusetts.

Ramachandran, V. S. and Hubbard, E. M. (2001). Synaesthesia - a window into perception, thought and language. *Journal of Consciousness Studies*, 8:3–34.

Ravasz, E. and Barabási, A.-L. (2003). Hierarchical organization in complex networks. *Physicial Review E*, 67.

Redford, M. A., Chen, C. C., and Miikkulainen, R. (2001). Constrained emergence of universals and variation in syllable systems. *Language and Speech*, 44:27–56.

Rendell, L. and Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences*, 24(2):309–382.

Renkou, G. (1995). *Guangzhou Renkou Zhi (广州市人口志Guangzhou Demography).* Guangzhoushi Tongjiju (广州市统计局Guangzhou Statistics Bureau), Guangzhou.

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, 3:131–141.

Roberts, J. (2002). Child language variation. In Chambers, J., Trudgill, P., and Schilling-Estes, N., editors, *The Handbook of Language Variation and Change*, pages 333–348. Blackwell Publishers Ltd, Oxford, UK.

Romaine, S. (1992). The evolution of linguistic complexity in pidgin and creole languages. In Hawkins, J. A. and Gell-Mann, M., editors, *The Evolution of Human Languages*, SFI Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA.

Ross, J. R. (1979). Where's English. In Fillmore, C. J., Kempler, D., and Wang, W. S.-Y., editors, *Individual Differences in Language Ability and Language Behavior*, pages 127–166. Academic Press.

Rumelhart, D. E. and McClelland, J. L. (1986). On the learning of the past tenses of English verbs. In Rumelhart, D. E., McClelland, J. L., and the PDP Group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* MIT Press, Cambridge, MA.

Rumelhart, D. E., McClelland, J. L., and the PDP Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 2: Psychological and Biological Models.* MIT Press, Cambridge, MA.

Saeed, J. I. (1997). *Semantics.* Blackwell Pub., Malden, MA.

Sagart, L. (1999). *The Roots of Old Chinese.* John Benjamins Pub. Co., Amsterdam; Philadelphia.

Sankoff, D. and Labov, W. (1979). On the uses of variable rules. *Language in Society*, 8(2):189–222.

Sankoff, D. and Lessard, R. (1975). Vocabulary richness: A sociolinguistic analysis. *Science*, 190:689–690.

Sapir, E. (1921). *Language: An Introduction to the Study of Speech.* Harcourt, Brace, New York.

Savage-Rumbaugh, E., Rumbaugh, D., and Boysen, S. (1980). Do apes use language. *American Scientist*, 1(68):49–61.

Savage-Rumbaugh, E. S. and Lewin, R. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. Doubleday, New York.

Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. *Minds and Machines*, 9:9.

Schoenemann, P. T. (2002). Initial word and learning in children: A neural net model that does not require language-specific innate constrains.

Scott, J. (2000). *Social Network Analysis: A Handbook*. Sage Publications, London.

Seidenberg, M. and Elman, J. (1999). Do infants learn grammar with algebra or statistics? (letter to science commenting on marcus et al. 1999). *Science*, 284:433.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275:1599–1603.

Shen, Z. (1997). *Exploring the dynamics aspect of sound change*. Journal of Chinese Linguistics.

Shore, C. M. (1995). *Individual Differences in Language Development*. Sage Pub., Thousand Oaks, CA.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61(1-2):39–91.

Slobin, D. (1977). Language change in childhood and in history. In Macnamara, J., editor, *Language Learning and Thought*, pages 185–214. Academia Press, New York.

Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, 9(2):175–190.

Smith, A. D. M. (2005). Mutual exclusivity: communicative success despite conceptual divergence. In Tallerman, M., editor, *Evolutionary Pre-requisites for Language*. Oxford University Press.

Snow, C. E. and Ferguson, C. A. (1977). *Talking to Children: Language Input and Acquisition (ed.)*. Cambridge University Press, Cambridge.

Sokolov, J. L. and Snow, C. E. (1994). *Handbook of Research in Language Development Using CHILDES (ed.)*. L. Erlbaum Associates, Hillsdale, N.J.

Sproat, R. (2002). Corpus-based methods in Chinese morphology. pdf online. http://www.research.att.com/r̃ws/newindex/coling.pdf.

Steels, L. (1996a). A self-organizing spatial vocabulary. *Artificial Life Journal*, 2(3):319–332.

Steels, L. (1996b). Self-organizing vocabularies. MIT Press, Cambridge, MA.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, 1:1–34.

Steels, L. (1998). Synthesizing the origins of language and meaning using coevolution, self-organization and level formation. In Hurford, J. R., Studdert-Kennedy, M., and Knight, C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, pages 384–404. Cambridge University Press, Cambridge.

Steels, L. (1999). The spontaneous self-organization of an adaptive language. In Furukawa, K., Michie, D., and Muggleton, S., editors, *Machine Intelligence (15): Intelligent Agents*, Chapter 11. Oxford University Press, Oxford.

Steels, L. (2001). Grounding symbols through evolutionary language games. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, pages 211–226. Springer-Verlag, London.

Steels, L. and Kaplan, F. (2002). Bootstrapping grounded word semantics. In Briscoe, E. J., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Chapter 3. Cambridge University Press.

Steels, L. and Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Harvey, I. and Husbands, P., editors, *European Conference of Artificial Life 1997*. MIT Press, Cambridge, MA.

Stimson, H. (1966). A tabu word in the Peking dialect. *Language*, 42:285–294.

Straight, H. S. (1976). Comprehension versus production in linguistic theory. *Foundations of Language*, 14:525–540.

Stringer, C. and McKie, R. (1997). *African Exodus: The Origins of Modern Humanity*. Henry Holt, New York.

Taylor, A. (1990). Variation in past tense formation in the history of English. In Izvorski, R. and et al., editors, *University of Pennsylvania Working Paper in Linguistics*, volume 1, pages 143–159.

Thomason, S. (2000). On the unpredictability of contact effects. *Estudios de Sociolinguistica*, 1(1):173–182.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press, Cambridge, Mass.

Tomasello, M. (2000). Two hypotheses about primate cognition. In Heyes, C. and Huber, L., editors, *The Evolution of Cognition*, pages 165–185. MIT Press.

Tomasello, M. (2002). Some facts about primate (including human) social learning and communication. In Cangelosi, A. and Parisi, D., editors, *Simulating the Evolution of Language*, pages 327–340. Springer-Verlag, London.

Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Harvard University Press, Cambridge, Mass.

Tomasello, M. and Bates, E. (2001). *Language Development: The Essential Readings*. Blackwell Publishers, Malden, MA.

Tomasello, M. and Call, J. (1997). *Primate Cognition*. Oxford University Press, New York.

Tomasello, M., Kruger, A., and Ratner, H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16:495–552.

Tonkes, B. (2001). *On the Origins of Linguistic Structure: Computational Models of the Evolution of Language*. PhD thesis, School of Information Technology and Electrical Engineering, University of Queensland, Australia.

Trask, R. L. (1996). *Historical Linguistics*. Arnold, London.

Tremblay, S., Shiller, D. M., and Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, 423:866–869.

Tsou, B. and You, R. (2003). Dangdai hanyu xinci de duoyuanhua quxiang he diqu jingzheng (当代汉语新词的多源化趋向和地区竞争The diversity trend and area competition of new words in Contemporary Mandarin). *Yuyan Jiaoxue yu Yanjiu (语言教学与研究)*, 2(100):12–21.

Tsou, B. K. (1976). Homophony and internal change in Chinese. *Computational Analysis of Asian & African Languages*, 3:67–86.

Turkel, W. J. (2002). The learning guided evolution of natural language. In Briscoe, E. J., editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, Chapter 8. Cambridge University Press.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2:717–726.

Van Orden, G. C. (1987). A rows is a rose: Spelling, sound, and reading. *Memory and Cognition*, 15:181–198.

Vogt, H. (1954). Contact of languages. *Word*, 10:365–374.

Waddington, C. (1942). Canalization of development and the inheritance of acquired characters. *Nature*, 150:563–565.

Wagner, K., Reggia, J. A., Uriagereka, J., and Wilkinson, G. S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37–69.

Wang, L. (1958). *Hanyu Shi Gao (汉语史稿Studies on the History of Chinese*. Kexue Chubanshe (科学出版社), Beijing.

Wang, W. H. (1966). *Donghan-Sui Changyongci Yanbian Yanjiu (东汉-隋常用词演变研究A Study on Frequent Words in Donghan-Sui Dynasty*. Nanjing University Press, Nanjing.

Wang, W. S.-Y. (1967). Conjoining and deletion in mandarin syntax. *Monumenta Serica*, 26:224–236.

Wang, W. S.-Y. (1969a). Competing changes as a cause of residue. *Language*, 45(1):9–25.

Wang, W. S.-Y. (1969b). Project DOC: Its methodological basis. *Journal of the American Oriental Society*, 90:57–66.

Wang, W. S.-Y. (1971). The basis of speech. In Reed, C. E., editor, *The Learning of Language*, pages 267–306. Appleton-Century-Crofts, New York.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 280:61–72.

Wang, W. S.-Y. (1977). *The Lexicon in Phonological Change (ed.)*. Mouton, The Hague.

Wang, W. S.-Y. (1978). The three scales of diachrony. In Kachru, B., editor, *Linguistics in the Seventies*, pages 63–75. Department of Linguistics, University of Illinois, Paris.

Wang, W. S.-Y. (1979). Language change - a lexical perspective. *Annual Review of Anthropology*, (9):353–371.

Wang, W. S.-Y. (1982a). *Explorations in language evolution*. Osmanic University, Hyderabad.

Wang, W. S.-Y. (1982b). Variation and selection in language change. *Bulletin of the Institute of History and Philology*, 53:495–519.

Wang, W. S.-Y. (1989a). Language in China: A chapter in the history of linguistics. *Journal of Chinese Linguistics*, 17(2):183–222.

Wang, W. S.-Y. (1989b). Language prefabs and habitual thought. In *Forum Lectures at the 1989 TESOL Summer Institute, San Francisco State University*. reprinted in Explorations in Language, 1991, 397-412.

Wang, W. S.-Y. (1991). *Explorations in Language*. Pyramid Press, Taipei.

Wang, W. S.-Y. (1999). Language emergence and transmission. In Peyraube, A. and Sun, C., editors, *In Honor of Mei Tsulin: Studies on Chinese historical syntax and morphology: linguistic essays*, pages 247–257. Centre de Recherches sur l'Asie Orientale, Paris.

Wang, W. S.-Y. and Cheng, C.-C. (1977). Implementation of phonological change: The Shuang-Feng Chinese case. In Wang, W. S.-Y., editor, *The Lexicon in Phonological Change*, pages 148–157. Mouton.

Wang, W. S.-Y., Ke, J.-Y., and Minett, J. W. (2004). Computer modeling of language evolution. In Huang, C.-R. and Lenders, W., editors, *Computational Linguistics and Beyond: Perspectives at the Beginning of the 21st Century. Frontiers in Linguistics (I)*. Language and Linguistics, Academia Sinica, Taipei.

Watts, D. J. (1999). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, NJ.

Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age.* W.W. Norton, New York.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442.

Wei, Z. S. (1953). *Guangzhouhua Tongyin Zibiao (广州话同音字表A Collection of Homophones in Guangzhou Cantonese) (ed.).* Wenfang Chubanshe (文放出版社), Hong Kong.

Weinreich, U., Labov, W., and Herzog, M. I. (1968). Empirical foundations for a theory of language change. In Lehmann, W. P. and Malkiel, Y., editors, *Directions for Historical Linguistics*, pages 95–195. University of Texas Press, Austin.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., and Howell, F. C. (2003). Pleistocene homo sapiens from middle awash, ethiopia. *Nature*, 423:742–747.

Wildgen, W. (1990). Basic principles of self-organization in language. In Haken, H. and Stadler, M., editors, *Synergetics of Cognition: Proceedings of the International Symposium at Schloss Elmau, Bavaria, June 4-8, 1989*, pages 415–426. Springer-Verlag, Berlin; Hong Kong.

Wilson, E. O. (1972). Animal communication. *Scientific American*, 227:52–60.

Wong, W. W.-Y. and Stokes, S. F. (2001). Cantonese consonantal development: A nonlinear account. *Journal of Child Language*, 28(1):195–391.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18:47–67.

Wray, A. (2000). Holistic utterances in protolanguage: The link from primates to humans. In Knight, C., Hurford, J. R., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge University Press, Cambridge.

Wray, A. (2002). *The Transition to Language*. Oxford University Press.

Wu, Z.-W. (2001). *Xianqin Hanyu Fuyinci Yanjiu (*先秦汉语复音词研究*A Study on the Disyllabic Words in Pre-Qin*. Bashu Shushe (巴蜀出版社), Chengdu.

Xiang, X. (1993). *Jianming Hanyu Shi (*简明汉语史*A Brief History of Chinese)*. Gaodeng Jiaoyu Chubanshe (高等教育出版社), Beijing: China.

Yang, C. (2003). *Knowledge and Learning in Natural Language*. Oxford University Press, USA.

Yeung, S.-W. H. (1981). Some aspects of phonological variation in the Cantonese spoken in Hong Kong. Master thesis, University of Hong Kong.

Yuan, J.-H. (1960). *Hanyu Fangyan Gaiyao (*汉语方言概要*A Synopsis of Chinese Dialects)*. Wenzi Gaige Chubanshe (文字改革出版社), Beijing. 2nd, 2001, Yuwen Chubanshe.

Zee, E. (1999). Change and variation in hong kong cantonese. *Journal of Chinese Linguistics*, 27(1):120–167.

Zhan, B.-H. (2002). *Guangzhouhua Zhengyin Zidian (*广州话正音字典*Dictionary on Correct Pronunciations of Guangzhou Cantonese)*. Guangdong Renmin Chubanshe (广东人民出版社), Guangzhou, 1st edition.

Zhan, B.-H., Cheung, Y.-S., and et al. (1990). *Zhujiang Sanjiaozhou Fangyan Zongshu (*珠江三角洲方言综述*A Synthetic Review on Dialects in Pearl Delta Area)*. Guangdong Renmin Chubanshe (广东人民出版社), Guangzhou.

Zhang, L. (1997). *Hanyu Cihui de Liubian (*汉语词汇的流变*The Changes of Chinese Lexicon)*. Daxiang Chubanshe (大象出版社), Zhengzhou, China.

Zhu, H. and Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1):3–42.

Zhuangdong (1985). *Zhuangdong Yuzu Yuyan Cihui Ji (*壮侗语族语言词汇集*A Collection of Lexicons of Languages in the Zhuang-dong Language Family)*. Zhongyang Minzu Xueyuan Chubanshe, Beijing.

Zihui (1989). *Hanyu Fangyin Zihui (*汉语方音字汇*Chinese Dialect Character Pronunciation List)*. Beijing Daxue Zhongguo Yuyan Wenxue Xi Yuyanxue Jiaoyanshi, Wenzi Gaige Chubanshe (北京大学中国语言文学系语言学教研室,文字改革出版社), 2nd edition. 1st ed. 1962.

Zipf, G. K. (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology*. MIT Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley.

Appendices

1. Homophony

1.1. A collection of artistic usage of homophones

The following 20 cases are examples of homophones used for artistic purpose in campaign slogans or advertisements. They were collected in local areas of Hong Kong, except case 18. The target Chinese characters are homophonuous when they are pronounced in Cantonese.

1. 乐作剧 (恶作剧) [the name of a candidate for the election campaign for the drama society in CityU, 2001]: using music to make fun - prank (music (乐) - evil (恶))

2. 城大戏 (成大器) [the name of a candidate for the election campaign for the drama society in CityU, 2001]: city university's drama - become a great talent (city (城) - become (成) ; drama (戏) -utility (器))

3. 金资玉叶 (金枝玉叶) [the name of a candidate for the election campaign for the student union committee of the Department of Economics in CityU, 2001]: golden capital and jade leaves - imperial kingsmen/ladies from noble families (capital (资)-branch (枝))

4. 名音里 (明阴你) [the name of a candidate for the election campaign for the music society, 2003] (famous music community - overtly harm you (famous (名) - overtly(明); music (音) - harm (阴); community (里) - you (你))

5. 城心可见 (诚心可见) [the name of a candidate for the election campaign for the student union committee in CityU, 2003]: heart of the city observable - sincerity observable (city (城) - sincerity (诚))

6. 天星小邻 (天星小轮) [the name of a candidate for the election campaign for the astronomy society in CityU, 2003]: neighbor of the star - the star ferry (neighbor (邻) - ferry (轮))

7. 城大戏, 好激戏 (成大器, 好激气) [the name of a candidate for the election campaign for the drama society in CityU, 2004]: city university's drama, good exciting drama - to become a great talent, very annoying (drama (激戏) - annoying (激气))

8. 资事分子 (滋事分子) [the name of a candidate for the election campaign for the student union committee of the Department of Information Sciences in CityU, 2004]: the members of the information science - the trouble makers (information matter (资事) - make trouble (滋事))

9. 直飞您想地 (直飞理想地) [an ad. in a supermarket, promoting some lucky draw

offering free flights, 2002]: fly directly to where you want to go - fly directly to the ideal place (you want (您想) - ideal (理想))

10. 净是新感觉 ((干) 净是新感觉) [an ad for a beauty product in the subway, 2002]: all are new feelings - cleanness is the new feeling (all (净) - cleanness (净))

11. 成工在望 (成功在望) [an ad for government's professional training program, in a TV program, 2002]: hopeful to find a job - hopeful to succeed (成工 (find a job) - succeed (成功))

12. 存为你 (全为你) [an ad of a bank for a new saving service, 2002]: save for you - all for you (save (存) - all (全))

13. 出期智胜 (出奇制胜) [an ad of an investment company for its future market service, 2003]: buy a future deal and win wisely - defeat one's opponent by a surprise move (future (期) - surprise (奇); wise (智) - defeat (制))

14. 好息和你 (好适合你) [an ad of a bank for high interest saving service, 2003]: good interest to fit you - fit you very well ((saving) interest (息) - fit (适))

15. 松未口甘抵 (从未口甘抵) [an ad of a credit card company, a squirrel playing as the actor, 2003]: the squirrel is never cheap - never be so cheap (squirrel (松) - ever (从))

16. 羽众不同 (与众不同) [an ad of a fur coat dress, 2003]: the feather is different - being different from all others (feather (羽) - with (与))

17. 脑地方 (老地方) [the name of a computer shop in a shopping mall, 2003]: the place for the brain - the old place (brain (脑) - old (老))

18. 无发无天 (无法无天) [the name of a salon in Taiwan, 2003]: no hair no sky - without law and order (hair (发) - law (法))

19. 童声童气 (同声同气) [the name of a children dress shop, 2003]: children's voice - harmonize (child (童) - same (同))

20. 粤港越好玩 (越讲越好玩) [the name of a program for tourism in TV, 2004]: Canton and Hong Kong are more and more fun - the more you talk, the more fun (Guangdong province (粤) - the more (越); Hong Kong (港) - talk (讲))

1.2. A collection of ambiguities in actual interactions

The following is a list of 26 cases of ambiguities that happened in daily life communications. In some cases we can see clearly the speakers' effort to repair mis-communication due to the ambiguities. Some ambiguities are caused by homophones or near-homophones (1-23); and some are caused by ambiguities in syntactic structures (22-26). Most of these conversations are in Putonghua, except no. 5, 6, 7, 18 in English, and no. 16 and 19 are in Cantonese.

Lexical ambiguity:

1. 元音 - 原因 (vowel - reason)

[a telephone conversation between a professor and his student, 1999.]

-Professor: 让我们来谈谈元音. (Let's talk about vowel.)

-Student: 我想说的原因是 . . . (The reason is that . . .)

2. 稀 - 细 (sparse - thin)

[a conversation about a graph from a simulation model between a professor and his student, 2002.]

-Student: 这条线太稀了. (The (dots) in the line are too sparse.)

-Professor: 太细了? ((the line is) too thin?)

3. 教给 - 交给 (teach - give)

[a conversation about Matlab program and programming, 2002.]

-A: 你以后教给我. (Please teach me later.)

-B: 好啊, 明天交给您. (Ok, I will give it to you tomorrow.)

4. 早到 - 找到 (early - found)

[a conversation between two professors. W knocked at M's door several minutes before plan.]

-W: 我早到了. (I arrived earlier.)

-M: 你找到什么了? (What did you find?)

5. can't - can

[a misunderstanding between a student and his supervisor.]

-Supervisor: I told you that you can't use that instrument.

-Student: Yes, that is why I used it. You told me I can use it.

-Supervisor: No, I said you can't.

6. fool - full

[a telephone conversation between dad and daughter.]

-Daughter: Dad, I am a fool.

-Dad: What did you eat?

7. not today - not a date

[a chat between friends, 2002.]

-Y: Are you going to dinner?

-J: Oh, I am expecting some one.

-Y: a girl?

-J: not today.

-Y: Not a date?

8. 投诉 - 投宿 (complain - lodge)

[in a hotel, a conversation between the staff and a guest, 2002.]

-Staff: 已经确实知道今天不会有水了. 非常抱歉, 不过我可以带你去投诉. (It has been announced that today won't have water. I am sorry for that. But if you want to complain, I can let you how to do so.)

-Guest: 投宿? 你让我们换个地方住啊? (To lodge elsewhere? Are you suggesting us moving to another place?)

9. 手提(电脑) - (电话) (mobile phone - potable computer)

[a conversation between two friends, 2002.]

-A: 你没把手提带过来啊？ (Didn't you bring your notebook over?)

-B (pointed to the mobile phone on the desk.)

-A: (smile, and shook his head).

-B: Oh, sorry. You mean my notebook? I did not bring it.

10. 笔记本 (computer notebook (laptop) - paper notebook)

[a conversation between a student and her colleague, 2001.]

-C: 你带笔记本了吗？ (Did you bring over your notebook?)

-L (showed him her notebook in her hand)

-C: No, I mean your notebook computer.

11. 辣 - 腊 (spicy - bacon)

[in a restaurant, in the middle of ordering food, 2002.]

-J: 辣这里不太常见的. (Spicy food is not common here.)

-W: 腊肉不常见？ (Bacon is not common here?)

12. 是忘 - 失望 (forget - disappointed)

[a conversation between two friends, 2003.]

-W: 你是忘了？ (Did you forget it?)

-J: 什么失望了？我没有啊. (Disappointed? No, I am not.)

13. 发声 - 发生 (phonation - emergence)

[a conversation between a student and his supervisor, 2003.]

-WF: 我想谈谈语言发声. (I would like to talk about phonation.)

-WW: 语言发生？ (Language emergence?)

14. 笔画 - 鼻化

[a conversation between a student and his supervisor, 2003.]

Student: 关于鼻化的问题 . . . (About "nasalization" . . .)

Supervisor: 你在说笔画？ (Are you talking about strokes?)

15. 猜出来 - 拆出来

[a conversation between two students about rule extraction in the simulation model, 2003.]

-G: 把这个 rule 拆出来. (to break this and extract a rule from it.)

-K: 猜出来？怎么猜？ (Guess? How to guess?)

16. 生 - 三汤

[a scene the two-dish counter in CityU canteen. The waitress (A) said 'sam tong' ("three soup"), but her colleague (B) heard she was saying 'sang tong' ("kill-alive"). B nudged A, and A realized, so she changed her words to 'sam go tong' ("three-CL-soups"). 2003. Note: in Hong Kong Cantonese, the velar nasal [-ŋ] is merging with alveolar nasal [-n]]

17. 言中 - 言重

[a conversation about the news which just came out that Taiwan decided to forbid hongkong's flights going into Taiwan to avoid SARS, 2003/4/28.]

-J: 不幸言中了. (Unfortunately it was predicted.)

-W: 言重？谁言重了？ (Flattering? Who is flattering?)

-J: Prof. Myers 说过可能一段时间香港的飞机不受台湾欢迎，果然言中了. (Prof. Myers said the flights from Hong Kong to Taiwan will be blocked soon. His prediction came true.)

18. scan

[a phone conversation between a student and her supervisor, 2004. Early before the conversation, he gave her some pictures to scan to digital files, and also a paper to read.]

-J: I will scan the paper.

-W: When will you scan them?

-J: Oh, you meant the pictures. I will scan them right now. In fact I almost forgot. I was telling you that I will read the paper you gave me today.

19. 爱国报纸 - 外国报纸 (patriotic newspaper - foreign newspaper)

[a dialogue in a phone-in radio program in Hong Kong, 2004.]

-caller: 有几家报纸已经慢慢成为爱国报纸. (A number of newspapers have gradually shifted to be patriot.)

-host: 什么外国报纸？ (What foreign newspapers?)

20. 红人 (American Indian - celebrated people)

[in a middle of a group discussion, 2004.]

-W: 那是一种红人的语言 (It is an American Indian language.)

-G: 红人？ (celebrated people?)

-W: 印地安人 (American Indian.)

-G: 我还以为您说名人呢. (I thought you meant some celebrated people's language.)

21. 三级 - 三集 (Grade-III - three parts)

[a conversation about a recent movie, 2004.]

-K: 那是个三级的片子 (That is a Grade-III movie.)

-W: 三集？ (Three parts?)

Syntactical ambiguity:

22. 他最想你 (He misses you the most (instead of missing others) - He misses you more than any other misses you)

[a phone conversation between two friends, 2003.]

A: 他最想你了. (He misses you the most (instead of missing others)).

B: 我想不是吧. 也许还有别人更想我. (I don't think so. Maybe there is someone else who misses me more than you.)

23. 你怎么能这么想我呢？ (How can you interpret me in this way? - How can you miss me so much?)

[a conversation between two labmates saying goodbye, 2003.]

-W: 我要回家了. (I am going home now.)

-J: 怎么这么早？ (Why so early?)

-W: 你心里肯定在想怎么不快点走. (You must be talking to yourself why this guy did not leave as quickly as he can.)

-J: 你怎么能这么想我呢？ (How can you interpret me in this way?)

-W: 我才不想你呢. (I don't miss you.)

24. 我借你个箱子 (lend - borrow)

[a conversation between two roommates, 2003.]

-T: 我借你个箱子.

-Y: 唔？借我个箱子？为什么？我不用？

-T: 我借你个箱子用.

-Y: 什么？

-T: 我的箱子坏了. 借了你的箱子.

-Y: 噢，我还以为你要借给我一个箱子呢. 没问题.

25. 不用还我

[a conversation between a student and her supervisor, 2004.]

-Supervisor: 不用还我. (You may return me those you don't need.)

-Student: 不用还你了？ (I don't need to return them to you?)

26. 没有消息是好消息

[a conversation between a student and her supervisor, 2004.]

-W: 没有消息是好消息. (No news is good news.)

-Y: 没有消息是好消息？ (No news is good news?)

-W: 没有消息就是好消息呀. (If there is no further news, then it means good news.)

-Y: 我以为您说一个好消息都没有呢. (I thought you meant there is no good news at all.)

1.3. Lists of homophones in the first 5000 words in three Germanic languages

The sets of homophones which have at least one word in the first 5000 word list, sorted according to the frequency of the first word. The number within the parenthesis is the frequency given in CELEX, followed by the possible part of speech.

A list of 151 homophone sets in Dutch

index	word 1	word 2	word 3
1	hij(470924,pron)	hei(419,n,int)	
2	bij(160583,prep)	bei(8,n)	
3	wat(152600,pron,adv)	wad(30,n)	watt(54,n)
4	nog(149233,adv)	noch(5562,conj)	
5	tot(126926,prep,conj)	tod(27,n)	
6	zij(112279,pron,v)	zei(84742,v)	
7	mij(67877,pron)	mei(2639,n)	

A list of 151 homophone sets in Dutch

8	kon(52291,v)	con(69,prep)	
9	wij(47266,pron)	wei(283,n)	
10	grote(40650,a,n)	grootte(1304,n)	
11	want(34656,conj)	wand(1043,n)	
12	keek(24581,v)	cake(116,n)	
13	wist(19986,v)	whist(9,n)	
14	liet(16405,v)	lied(841,n)	
15	moet(15891,v)	moed(2101,n)	
16	dood(14626,a,n)	doodt(87,v)	
17	grond(13608,n)	grondt(2,v)	
18	hield(13090,v)	hielt(2,v)	
19	licht(12940,v,a,n)	ligt(5693,v)	
20	vond(12712,v)	vont(7,n)	
21	land(12640,n)	landt(26,v)	
22	lag(12366,v)	lach(1622,n,v)	
23	bed(12052,n)	bet(12,v)	
24	reeds(11779,adv)	raids(11,n)	
25	pas(11737,n,a,n)	pass(94,n)	
26	geld(11691,n)	geldt(2425,v)	
27	stad(10540,n)	stat(3,v)	
28	recht(10305,a,n)	regt(32,v)	
29	laat(10143,v,a)	laad(11,v)	laadt(12,v)
30	mond(9313,n)	mondt(33,v)	
31	feit(8592,n)	fijt(13,n)	
32	hart(7773,n)	hard(5410,a)	hardt(2,v)
33	gebied(7766,n)	gebiedt(42,v)	
34	aarde(6492,n)	aardde(4,v)	
35	o(6287,int,n)	eau(102,n)	
36	wet(6171,n)	wed(58,v,n)	wedt(7,v)
37	druk(5675,v,n,a)	drug(61,n)	
38	wilt(5497,v)	wild(1631,a,n)	
39	raad(5492,n)	raadt(71,v)	raat(12,n)
40	beleid(5138,n)	belijd(2,v)	belijdt(22,v)
41	gauw(5107,a)	gouw(26,n)	
42	doch(5053,conj)	dog(73,n)	
43	oud(4897,a,n)	out(152,a)	

A list of 151 homophone sets in Dutch

44	oom(4499,n)	ohm(12,n)	
45	wind(4475,n)	windt(40,v)	wint(227,v)
46	voet(4069,n)	voed(9,v)	voedt(69,v)
47	niks(3907,adv,pron)	nix(8,n)	
48	rijk(3899,a,n)	reik(19,v)	
49	houdt(3794,v)	hout(2056,n,n)	houd(1193,v)
50	mag(3611,v)	mach(31,n)	
51	rood(3575,a,n)	root(2,n)	
52	reis(3493,n)	rijs(2,v)	
53	wou(3488,v)	wauw(10,a)	wouw(9,n)
54	koud(3250,a)	kout(16,n)	
55	links(3226,a)	lynx(16,n)	
56	maand(3147,n)	maant(16,v)	
57	i(3123,n)	ie(1643,pron)	
58	wisten(3115,v)	whisten(3,v)	
59	a(3066,n,prep)	ah(463,int)	
60	gij(2832,pron)	gei(215,n,v)	
61	reed(2822,v)	raid(12,n)	reet(146,n,v)
62	waard(2717,a)	waart(58,v)	
63	bos(2453,n,n)	boss(38,n)	
64	hals(2383,n)	halls(9,n)	
65	zand(2362,n)	zandt(2,v)	
66	rij(2127,v,n)	rei(11,n,v)	
67	lijden(2091,n,v)	leidden(445,v)	leiden(1974,v)
68	boot(2085,n)	bood(1418,v)	
69	afscheid(2082,n)	afscheidt(14,v)	
70	zeiden(2077,v)	zijden(1126,n,a)	
71	pad(2048,n,n)	pat(4,n)	
72	lezer(2018,n)	laser(41,n)	
73	strand(1999,n)	strandt(5,v)	
74	neiging(1985,n)	nijging(2,n)	
75	boord(1949,n)	boort(23,v)	
76	stijl(1947,n)	steil(347,a)	
77	band(1843,n)	bant(6,v)	
78	mis(1826,a,n,v)	miss(472,n)	
79	termen(1792,n)	thermen(232,n)	

A list of 151 homophone sets in Dutch

80	leider(1750,n)	lijder(45,n)	
81	veld(1732,n)	velt(18,v)	
82	trek(1726,n,v)	track(5,n)	
83	oor(1710,n)	oir(22,n)	
84	schade(1676,n)	schaadde(15,v)	
85	top(1648,n)	tob(7,v)	
86	luid(1641,a)	luidt(435,v)	luit(84,n)
87	papa(1639,n)	pappa(206,n)	
88	moord(1609,n)	moordt(6,v)	
89	kast(1518,n)	cast(4,n)	
90	steek(1509,n,v)	steak(25,n)	
91	bericht(1429,n,v)	beringd(2,v)	
92	mes(1396,n)	mess(80,n)	
93	bad(1316,v,n)	bat(6,n)	
94	keus(1275,n)	queues(4,n)	
95	leiders(1254,n)	lijders(76,n)	
96	ijs(1196,n)	eis(1048,n,v)	
97	hal(1182,n)	hall(453,n)	
98	ei(1175,int,n)	ij(120,n)	y(128,n)
99	blok(1144,n,n)	bloc(38,exp)	
100	nood(1125,n)	noodt(17,v)	noot(379,n)
101	wijd(1108,a,v)	weidt(6,v)	weit(14,n)
102	historisch(1044,a)	histories(20,n)	
103	reizen(1029,n,v)	rijzen(175,v)	
104	genoot(998,v,n)	genood(11,v)	
105	stok(988,n)	stock(53,n)	
106	biedt(977,v)	beat(34,n)	
107	meid(944,n)	mijt(11,n)	mijdt(7,v) mijd(5,v)
108	held(924,n)	helt(12,v)	
109	kostte(915,v)	koste(717,n)	
110	rok(903,n)	rock(66,n)	
111	gewijd(892,a,v)	geweid(7,v)	
112	koers(880,n)	cours(21,n)	
113	wijde(870,a)	weidde(193,n,v)	
114	lied(841,n)	liet(16405,v)	
115	lezers(830,n)	lasers(20,n)	

A list of 151 homophone sets in Dutch

116	onderhoud(826,n)	onderhout(18,n)
117	baard(825,n)	baart(49,v)
118	dubbel(807,n,a)	double(5,n)
119	bek(805,n)	back(3,n)
120	gewekt(762,v)	geweckt(3,v)
121	cent(756,n)	sent(5,n)
122	puk(734,n)	puck(3,n)
123	bond(727,v,n)	bont(333,n,a)
124	dek(716,n,v)	deck(10,n)
125	pols(675,n,v)	polls(2,n)
126	ad(659,prep)	at(644,v)
127	slap(648,a)	slab(5,n)
128	gelach(634,n)	gelag(35,n)
129	vermijden(629,v)	vermeiden(2,v)
130	hoede(628,n)	hoedde(28,v)
131	gebed(624,n)	gebet(9,v)
132	natie(593,n)	nazi(79,n)
133	mouw(589,n)	mauw(4,v)
134	wekken(587,v)	wecken(2,v)
135	fort(586,n,n)	ford(414,n)
136	aangelegd(580,a,v)	aangelengd(55,v)
137	lies(573,n)	lease(2,n)
138	rijst(569,v,n)	reist(69,v)
139	peil(569,n)	pijl(393,n)
140	pond(564,n)	pont(164,n)
141	unie(545,n)	uni(3,a)
142	graad(542,n)	graat(46,n)
143	meiden(517,n)	mijden(56,v)
144	forse(513,a)	force(88,n)

A list of 357 homophone sets in English

1	the(1093547,adv,art)	thee(166,pron)	
2	a(844671,art,n)	aye(128,adv,scon)	eh(547,scon)
3	to(483428,prep)	too(18686,adv)	two(24552,num)
4	in(337995,adv,prep)	inn(169,n)	
5	I(198140,pron)	aye(102,adv)	eye(2284,n)

A list of 357 homophone sets in English

6	for(148356,prep,conj)	fore(44,a,adv,n)	four(5843,num)	
7	you(128688,pron)	ewe(22,n)	u(358,n)	yew(28,n)
8	be(111471,v)	bee(118,n)		
9	but(96889,conj)	butt(115,n,v)		
10	not(91464,adv)	knot(136,n,v)		
11	by(79382,adv,prep)	buy(565,v)	bye(197,scon)	
12	or(76563,conj)	oar(7,n)	ore(55,n)	
13	we(64846,pron)	wee(111,a,n)		
14	all(64395,adv,pron)	awl(3,n)		
15	one(61857,n,pron,num)	won(251,v)		
16	which(61399,pron)	witch(279,n)		
17	so(53887,adv,conj)	sew(12,v)	sow(20,v)	
18	what(50116,pron)	watt(60,n)		
19	been(48589,v)	bean(68,n)		
20	no(46469,adv,pron)	know(6088,v)		
21	when(46311,conj,pron)	wen(10,n)		
22	him(45024,pron)	hymn(114,n)		
23	more(43746,pron,adv)	moor(35,n)		
24	me(43071,pron)	mi(19,n)		
25	some(34232,adv,pron)	sum(568,n)		
26	time(32093,n)	thyme(54,n)		
27	your(28923,pron)	yore(5,n)		
28	our(23029,pron)	hour(2867,n)		
29	way(21570,n,adv)	weigh(39,v)	whey(38,n)	
30	were(21107,v)	whirr(13,n)		
31	new(19011,a)	knew(2038,v)		
32	where(18971,adv,conj,pron)	ware(19,n)	wear(412,n,v)	
33	through(17007,adv,prep)	threw(212,v)		
34	right(14779,n,a,adv)	rite(57,n)	wright(127,n)	write(568,v)
35	here(14061,adv)	hear(836,v)		
36	world(13345,n)	whirled(13,v)	whorled(8,a)	
37	would(13221,v,v)	wood(1166,n)		
38	away(12155,adv)	aweigh(2,adv)		
39	course(11914,n)	coarse(196,a)	corse(8,n)	
40	great(11693,a)	grate(33,n,v)		
41	why(11103,adv,scon,pron,conj)	y(896,n)		

A list of 357 homophone sets in English

42	while(11005,n,conj)	wile(7,n)	
43	few(10473,pron,n)	phew(17,scon)	
44	cent(10038,n,)	scent(254,n)	sent(551,v)
45	place(9474,n,v)	plaice(11,n)	
46	room(8249,n)	rheum(3,n)	
47	oh(8158,scon)	o(1779,n,scon)	owe(56,v)
48	night(7671,n)	knight(133,n)	
49	whole(7546,n,a)	hole(1015,n)	
50	done(7282,a,v)	dun(28,a,n)	
51	eyes(7100,n)	ayes(5,n)	
52	days(6999,n,n)	daze(10,v)	
53	side(6943,n)	sighed(90,v)	
54	seen(6750,v)	scene(1591,n)	
55	white(6594,n,a)	wight(42,n)	
56	high(6516,a,adv,n)	hi(153,scon)	
57	per(6515,prep)	purr(15,n)	
58	least(6427,n,pron,adv)	leased(5,v)	
59	war(6082,n)	wore(271,v)	
60	whether(5978,conj)	weather(1167,n,v)	wether(7,n)
61	mind(5530,n,v)	mined(13,v)	
62	sort(5529,n,v)	sought(132,v)	
63	sure(5439,a,adv)	shore(408,n,v)	
64	morning(5406,n)	mourning(114,n,v)	
65	see(5242,v)	sea(2872,n)	si(73,n)
66	past(4922,n,a,adv,prep,pron)	passed(498,v)	
67	father(4872,n)	farther(436,a,adv)	
68	week(4853,n)	weak(839,a)	
69	air(4496,n)	ere(12,conj)	heir(121,n)
70	seems(4342,v)	seams(66,n)	
71	need(4337,n,v)	knead(5,v)	
72	feet(4104,n)	feat(97,n)	
73	poor(4039,a,n)	pore(10,n)	pour(110,v)
74	word(3950,n)	whirred(2,v)	
75	hours(3881,n)	ours(458,pron)	
76	boy(3876,n,scon)	buoy(8,n)	
77	road(3791,n)	rode(70,v)	rowed(6,v)

A list of 357 homophone sets in English

78	rest(3703,n,v)	wrest(4,v)	
79	made(3572,v)	maid(227,n)	
80	main(3459,a)	mane(45,n)	
81	hair(3420,n)	hare(40,n)	
82	must(3304,v)	mussed(2,v)	
83	might(3025,v,v)	mite(27,n)	
84	story(2992,n)	storey(36,n)	
85	red(2889,n,a)	read(583,v)	
86	son(2858,n)	sun(2689,n)	
87	ways(2793,n,n)	weighs(57,v)	
88	arms(2668,n,n)	alms(13,n)	
89	boys(2599,n,n)	buoys(4,n)	buoys(2,v)
90	heart(2597,n)	hart(80,n)	
91	greater(2542,a)	grater(4,n)	
92	low(2536,n,a,adv)	lo(337,scon)	
93	set(2527,n,v,a)	sett(11,n)	
94	waiting(2491,v)	weighting(10,n,v)	
95	higher(2411,a,adv,a)	hire(139,n,v)	
96	mean(2339,n,v,a)	mien(14,n)	
97	turn(2324,n,v)	tern(19,n)	
98	find(2318,v)	fined(17,v)	
99	knows(2306,v)	noes(4,n)	nose(1307,n)
100	blue(2304,n,a)	blew(96,v)	
101	court(2240,n,v)	caught(396,v)	
102	sex(2216,n)	secs(50,n)	
103	straight(2198,a,adv)	strait(30,n,a)	
104	dear(2173,n,a,scon)	deer(105,n)	
105	none(2165,adv,pron)	nun(96,n)	
106	hall(2160,n)	haul(54,n,v)	
107	please(2119,v,scon)	pleas(50,n)	
108	role(2102,n)	roll(281,n,v)	
109	piece(2082,n)	peace(1596,n)	
110	writing(2030,n,v)	righting(6,v)	
111	size(2015,n)	sighs(49,v,n)	
112	tax(1949,n)	tacks(11,n)	
113	council(1818,n)	counsel(186,n,v)	

A list of 357 homophone sets in English

114	meet(1792,v,a)	meat(1280,n)	
115	told(1763,v)	tolled(3,v)	
116	talk(1728,n,v)	torque(37,n)	
117	due(1696,n,a,prep)	dew(93,n)	
118	sight(1685,n)	cite(11,v)	site(936,n)
119	ball(1664,n)	bawl(4,v)	
120	weight(1664,n)	wait(628,v,n)	
121	born(1634,a)	bourn(17,n)	
122	sign(1621,n,v)	sine(49,n)	
123	colonel(1610,n)	kernel(19,n)	
124	tea(1589,n)	tee(56,n,v)	ti(3,n)
125	call(1582,n,v)	caul(3,n)	
126	current(1582,n,a)	currant(19,n)	
127	cause(1553,n,v)	cores(20,n)	
128	style(1550,n,v)	stile(7,n)	
129	male(1546,n,a)	mail(446,n,v)	
130	board(1534,n,v)	bored(89,v)	
131	fair(1527,n,a,adv)	fare(156,n,v)	
132	horse(1518,n)	hoarse(74,a)	
133	seemed(1492,v)	seamed(4,v)	
134	source(1376,n)	sauce(272,n)	
135	manner(1366,n)	manor(119,n)	
136	pain(1354,n)	pane(31,n)	
137	bread(1327,n)	bred(31,v)	
138	rights(1317,n)	rites(87,n)	writes(348,v)
139	wine(1313,n)	whine(47,n,v)	
140	key(1281,n,n)	quay(58,n)	
141	rain(1261,n,v)	reign(111,n,v)	rein(40,n)
142	steps(1248,n,v,n)	steppes(11,n)	
143	iron(1234,n,v)	ion(36,n)	
144	step(1232,n,v)	steppe(10,n)	
145	leader(1227,n)	lieder(4,n)	
146	sick(1226,a,n)	sic(43,adv)	
147	use(1218,n,v)	ewes(23,n)	yews(12,n)
148	slow(1211,v,a,adv)	sloe(14,n)	
149	flowers(1198,n)	flours(23,v)	

A list of 357 homophone sets in English

150	base(1187,n,a)	bass(75,n)	
151	principle(1107,n)	principal(619,n,a)	
152	stories(1098,n)	storeys(34,n)	
153	hold(1082,n,v)	holed(5,v)	
154	worst(1075,n,adv,a)	wurst(3,n)	
155	beach(1060,n)	beech(185,n)	
156	birth(1052,n)	berth(68,n)	
157	pale(1030,n,a)	pail(48,n)	
158	muscles(1015,n)	mussels(50,n)	
159	grown(1015,a,v)	groan(40,n,v)	
160	heard(1014,v)	herd(162,n,v)	
161	tears(995,n)	tiers(33,n)	
162	waste(994,n,v,a)	waist(366,n)	
163	seem(987,v)	seam(40,n)	
164	aid(986,n,v)	aide(100,n)	
165	forty(958,n,num)	forte(8,n)	
166	plain(945,n,a,adv)	plane(815,n)	
167	signs(940,n)	sines(4,n)	
168	sources(918,n)	sauces(48,n)	
169	cash(918,n,v)	cache(15,n)	
170	coat(909,n,v)	cote(19,n)	
171	gate(878,n)	gait(47,n)	
172	lay(868,v,a,v)	ley(23,n)	
173	turns(857,n,v)	terns(10,n)	
174	sweet(840,n,a)	suite(191,n)	
175	beer(832,n)	bier(12,n)	
176	sons(831,n)	suns(39,n)	
177	break(816,n,v)	brake(68,n,v)	
178	forth(813,adv)	fourth(805,n,num)	
179	stairs(805,n,n)	stares(147,n,v)	
180	trust(805,n,v)	trussed(4,v)	
181	ring(804,n,v,v,n)	wring(9,v)	
182	metal(803,n)	mettle(10,n)	
183	troops(798,n,n)	troupes(4,n)	
184	sees(790,v)	seas(230,n)	seize(44,v)
185	dust(774,n,v)	dost(11,v)	

A list of 357 homophone sets in English

186	dying(765,a,v)	dyeing(28,v)	
187	rough(762,v,n,a)	ruff(12,n)	
188	steel(755,n,v)	steal(59,n,v)	
189	coal(750,n,v)	kohl(27,n)	
190	rocks(742,n,n)	rocs(2,n)	
191	wave(739,n,v)	waive(2,v)	
192	thrown(735,v)	throne(175,n)	
193	soul(731,n)	sole(290,a,n)	
194	breaking(730,v)	braking(30,v)	
195	minor(712,n,a)	miner(41,n)	
196	route(710,n)	root(377,n,v)	
197	principles(696,n,n)	principals(28,n)	
198	pride(688,n)	pried(3,v)	
199	sets(680,n,v)	setts(15,n)	
200	cross(677,n,v,a)	crosse(4,n)	
201	bare(660,v,a)	bear(329,n,v)	
202	guilt(653,n)	gilt(120,n)	
203	flow(652,n,v)	floe(6,n)	
204	nights(650,n,adv)	knights(102,n)	
205	cell(650,n)	sell(259,n,v)	
206	martin(649,n)	marten(5,n)	
207	profits(646,n)	prophets(101,n)	
208	roots(636,n,v,n)	routes(214,n)	
209	lane(635,n)	lain(109,v)	
210	border(632,n)	boarder(5,n)	
211	sales(629,n,a)	sails(56,n,v)	
212	holes(628,n)	wholes(4,n)	
213	worn(627,v)	warn(51,v)	
214	profit(615,n,v)	prophet(178,n)	
215	wales(613,n)	wails(18,n,v)	whales(95,n)
216	sale(609,n)	sail(48,n,v)	
217	mark(607,n,v)	marc(26,n)	marque(5,n)
218	wise(591,n,a)	whys(10,n)	
219	fate(585,n)	fete(22,n)	
220	block(584,n,v)	bloc(112,n)	
221	handsome(583,a)	hansom(4,n)	

A list of 357 homophone sets in English

222	read(583,v,v)	reed(98,n)	
223	band(579,n,v)	banned(46,v)	
224	dependent(578,a)	dependant(23,n)	
225	wrote(575,v)	rote(16,n)	
226	overseas(569,a,adv)	oversees(2,v)	
227	bay(567,n,v,a)	bey(4,n)	
228	muscle(566,n)	mussel(10,n)	
229	cells(566,n)	sells(90,v)	
230	tail(553,n,v)	tale(313,n)	
231	check(553,n,v)	cheque(504,n)	
232	needed(551,v)	kneaded(2,v)	
233	raising(547,v)	razing(4,v)	
234	lessons(545,n)	lessens(15,v)	
235	sheer(525,a,adv)	shear(3,v)	
236	marks(517,n,v)	marques(24,n)	
237	bell(507,v,n)	belle(43,n)	
238	wheel(499,n,v)	weal(8,n)	
239	lie(498,v,v,n)	lye(17,n)	
240	pair(495,n)	pare(3,v)	pear(44,n)
241	led(488,v)	lead(445,n)	
242	flower(481,n,v)	flour(462,n)	
243	lesson(478,n)	lessen(17,v)	
244	bowl(477,n,v)	bole(9,n)	boll(14,n)
245	sites(475,n)	cites(25,v)	sights(136,n,v)
246	allowed(474,v)	aloud(328,adv)	
247	row(467,n,v,n)	roe(12,n)	
248	bearing(466,n,v)	baring(17,v)	
249	ceiling(463,n)	sealing(31,v,n)	
250	session(461,n)	cession(7,n)	
251	faint(461,n,v,a)	feint(10,n)	
252	damn(457,n,v,a,adv,scon)	dam(112,n)	
253	weekly(457,n,a,adv)	weakly(92,a,adv)	
254	blocks(452,n,v)	blocs(21,n)	
255	died(451,v)	dyed(15,v)	
256	cast(439,n,v)	caste(88,n)	
257	tide(438,n)	tied(136,v)	

A list of 357 homophone sets in English

258	guest(436,n)	guessed(71,v)		
259	load(435,n,v)	lowed(6,v)		
260	sonny(428,n)	sunny(203,a)		
261	symbol(420,n)	cymbal(4,n)		
262	fort(411,n)	fought(132,v)		
263	pause(409,n,v)	paws(58,n)	pores(44,n)	pours(39,v)
264	peak(406,n,v)	peek(9,n)	peke(2,n)	pique(18,n)
265	rose(394,n,v,n)	roes(5,n)	rows(363,n)	
266	prize(391,n,v,v,a)	pries(5,v)		
267	build(391,n,v)	billed(5,v)		
268	beaches(389,n)	beeches(18,n)		
269	taxes(387,n)	taxis(95,n)		
270	altar(380,n)	alter(79,v)		
271	add(377,v)	ad(107,n)		
272	manners(377,n,n)	manors(8,n)		
273	lawn(368,n)	lorn(4,a)		
274	wheels(364,n,v,n)	weals(2,n)		
275	die(360,v)	dye(76,n,v)		
276	tons(360,n)	tuns(2,n)		
277	links(359,v,n)	lynx(6,n)		
278	damned(359,adv,v,a)	dammed(2,v)		
279	lose(358,v)	loos(9,n)		
280	gates(353,n)	gaits(6,n)		
281	win(353,n,v)	whin(2,n)		
282	rings(350,n,v)	wrings(2,v)		
283	hey(345,scon)	hay(266,n)		
284	roles(344,n)	rolls(163,n,v)		
285	review(335,n,v)	revue(25,n)		
286	loan(335,n,v)	lone(109,a)		
287	styles(334,n)	stiles(10,n)		
288	guys(333,n)	guise(63,n)		
289	fur(333,n)	fir(41,n)		
290	breaks(333,n,v)	brakes(78,n)		
291	climb(331,n,v)	clime(4,n)		
292	pact(329,n)	packed(238,a,v)		
293	councils(326,n)	counsels(15,n,v)		

A list of 357 homophone sets in English

294	heels(324,n)	heals(22,v)	
295	planes(322,n)	plains(146,n)	
296	raised(318,v)	razed(2,v)	
297	lee(308,n)	lea(67,n)	
298	males(307,n)	mails(18,n,v)	
299	resting(304,v)	wresting(6,v)	
300	core(303,n)	cor(21,scon)	corps(172,n)
301	sink(300,n,v)	sync(4,n)	
302	teams(299,n)	teems(4,v)	
303	naval(297,a)	navel(61,n)	
304	keys(292,n,n)	quays(4,n)	
305	ringing(290,v)	wringing(37,v)	
306	grippe(289,n)	grip(90,n,v)	
307	choose(288,v)	chews(19,n,v)	
308	gaze(287,n,v)	gays(6,n)	
309	petrol(286,n)	petrel(3,n)	
310	leaf(270,n,v)	lief(8,adv)	
311	bases(270,n)	basses(8,n)	
312	prey(269,n,v)	pray(92,v,adv)	
313	mayor(268,n)	mare(59,n)	
314	floors(263,n)	flaws(35,n)	
315	tales(259,n)	tails(122,n,v)	
316	wastes(258,n,v)	waists(18,n)	
317	souls(256,n)	soles(84,n)	
318	sore(256,n,a)	soar(7,v)	
319	praise(255,n,v)	prays(12,v)	preys(9,v)
320	reads(255,v)	reeds(79,n)	
321	bells(252,n)	belles(4,n)	
322	miners(252,n)	minors(19,n)	
323	pains(252,n,n)	panes(46,n)	
324	shoe(249,n)	shoo(11,v,scon)	
325	adds(247,v)	ads(63,n)	adze(2,n)
326	creek(243,n)	creak(19,n,v)	
327	waving(243,v)	waiving(2,v)	
328	rumours(242,n)	roomers(2,n)	
329	mask(241,n,v)	masque(25,n)	

A list of 357 homophone sets in English

330	taught(241,v)	taut(78,a)	tort(14,n)
331	vain(239,a)	vane(18,n)	vein(93,n)
332	rays(238,n)	raise(204,v)	
333	pa(238,n)	pah(5,scon)	
334	beat(237,n,v)	beet(40,n)	
335	sword(237,n)	sawed(3,v)	soared(11,v)
336	stake(235,n,v)	steak(146,n)	
337	symbols(231,n)	cymbals(13,n)	
338	formerly(230,adv)	formally(216,adv)	
339	mist(229,n)	missed(129,v)	
340	jeans(229,n)	genes(227,n)	
341	waited(228,v)	weighted(17,a,v)	
342	lock(228,n,v)	loch(62,n)	lough(2,n)
343	halls(227,n)	hauls(11,n,v)	
344	rude(227,a)	rood(6,n)	
345	jam(227,n,v)	jamb(5,n)	
346	idle(227,a)	idol(86,n)	
347	pie(225,n)	pi(39,n)	
348	cheques(221,n)	checks(200,n,v)	
349	pairs(220,n,v)	pares(4,v)	pears(68,n)
350	dies(217,v)	dyes(26,n,v)	
351	bite(216,n,v)	bight(3,n)	
352	mode(215,n)	mowed(2,v)	
353	rains(212,n,v,n)	reigns(24,n,v)	reins(52,n)
354	sack(211,n,v)	sac(19,n)	
355	meets(208,n,v)	meats(59,n)	
356	bore(208,n,v,v)	boar(36,n)	
357	cereal(207,n)	serial(78,n,a)	

A list of 149 homophone sets in German

1	das(68427,art,pron,pron)	dass(35188,conj)
2	ist(44361,v)	isst(45,v)
3	sie(41752,pron,pron)	sieh(56,v)
4	als(28842,conj,adv)	Alls(4,n)
5	wie(21448,conj,adv)	Vieh(81,n)
6	war(19410,v)	wahr(585,a)

A list of 149 homophone sets in German

7	wird(18969,v)	Wirt(89,n)	
8	man(16296,pron)	Mann(2883,n)	
9	bis(10293,prep,adv,conj)	biss(46,v)	Biss(11,n)
10	mehr(9710,pron,adv)	Meer(334,n)	
11	ihre(7652,pron,pron,pron,pron)	Ire(10,n)	
12	wieder(7510,adv)	wider(122,prep)	
13	Sie(6952,pron,n)	sieh(56,v)	
14	heute(5272,v,adv)	Haeute(49,n)	
15	waren(5035,v)	wahren(194,a,v)	
16	Uhr(4337,n)	Ur(4,n)	
17	ihren(3948,pron,pron,pron,pron)	Iren(12,n)	
18	seit(3496,prep,adv)	seid(114,v)	
19	Seite(2964,n)	Saite(7,n)	
20	viel(2728,pron,adv)	fiel(611,v)	
21	denen(2498,pron,pron)	dehnen(4,v)	
22	viele(2225,pron)	fiele(5,v)	
23	wer(2214,pron,pron,pron)	Wehr(60,n,v)	
24	mal(2117,adv)	Mahl(12,n)	
25	Stadt(1889,n)	statt(1120,prep,conj)	Statt(10,n)
26	liess(1722,v)	lies(9,v)	
27	recht(1398,v,a)	raecht(3,v)	
28	weit(1355,a)	weiht(2,v)	
29	vielen(1296,pron)	fielen(201,v)	
30	Namen(1210,n)	nahmen(406,v)	
31	Stelle(1199,n)	Staelle(14,n)	
32	her(1177,adv)	Heer(69,n)	
33	sechs(1155,num)	Sex(32,n)	
34	fast(1142,v,adv)	fasst(71,v)	
35	statt(1120,prep,conj)	Stadt(1889,n)	
36	stellen(1077,v)	Staellen(7,n)	
37	Recht(1007,n)	raecht(3,v)	
38	konnten(1000,v)	Konten(21,n)	
39	bald(956,adv)	ballt(3,v)	
40	Seiten(856,n)	Saiten(5,n)	
41	Bad(821,n)	bat(270,v)	
42	Rede(820,n)	Reede(7,n)	

A list of 149 homophone sets in German

43	haelt(817,v)	Held(105,n)	
44	Wagen(803,n)	vagen(5,a)	Waagen(3,n)
45	Tod(712,n)	tot(293,a)	
46	Ihre(674,pron,pron)	Ire(10,n)	
47	Wahl(640,n)	Wal(2,n)	
48	hast(634,v)	hasst(12,v)	
49	fiel(611,v)	viel(2728,pron,adv)	
50	setzen(604,v)	Saetzen(119,n)	
51	wahr(585,a)	war(19410,v)	
52	Grad(583,n)	Grat(6,n)	
53	Wert(567,n)	wehrt(45,v)	
54	A(563,n)	ah(124,int)	
55	Lehrer(552,n)	leerer(21,a)	
56	waeren(534,v)	waehren(2,v)	
57	Herz(471,n)	Hertz(9,n)	
58	Willen(464,n,n)	Villen(33,n)	
59	Gaeste(463,n)	Geste(66,n)	
60	Rat(445,n)	Rad(151,n)	
61	Mal(435,n)	Mahl(12,n)	
62	faellt(424,v,v)	Feld(345,n)	
63	Faellen(419,n)	Fellen(2,n)	
64	nahmen(406,v)	Namen(1210,n)	
65	Stellen(391,n)	Staellen(7,n)	
66	rein(388,a,adv)	Rain(2,n)	Rhein(249,n)
67	wies(377,v)	Viehs(5,n)	
68	Male(363,n)	Mahle(3,n)	
69	Bund(354,n,n)	bunt(47,a)	
70	Feld(345,n)	faellt(424,v)	
71	Rechte(337,n)	raechte(7,v)	
72	Meer(334,n)	mehr(9710,pron,adv)	Mehr(12,n)
73	erhaelt(322,v)	erhellt(20,v)	
74	Staedte(310,n)	Staette(46,n)	
75	tot(293,a)	Tod(712,n)	
76	Start(285,n)	starrt(25,v)	
77	spaet(279,a)	spaeht(3,v)	
78	kannte(276,v)	Kante(43,n)	

A list of 149 homophone sets in German

79	bat(270,v)	Bad(821,n)	
80	weist(267,v)	weisst(230,v)	
81	Ton(266,n)	Thon(2,n)	
82	halte(266,v)	hallte(9,v)	
83	Werte(255,n)	wehrte(61,v)	
84	Lehre(251,n)	leere(80,a,v)	Leere(63,n)
85	hart(250,a,adv)	harrt(6,v)	
86	Rhein(249,n)	rein(388,a,adv)	Rain(2,n)
87	Staedten(247,n)	Staetten(24,n)	
88	wenden(232,v,v)	Waenden(52,n)	
89	weisst(230,v,v)	weist(267,v)	
90	Ihren(224,pron,pron)	Iren(12,n)	
91	Saetze(221,n)	setze(45,v)	
92	verliess(220,v)	Verlies(3,n)	
93	Waren(220,n)	wahren(194,a,v)	
94	holen(213,v)	hohlen(12,a)	
95	Verband(205,n)	verbannt(15,v)	
96	Faelle(204,n)	Felle(16,n)	
97	bot(203,v)	Boot(86,n)	
98	Stil(201,n)	stiehl(2,v)	Stiel(11,n)
99	wahren(194,v,a)	waren(5035,v)	Waren(220,n)
100	gerecht(188,v,a)	geraecht(2,v)	
101	Kueste(186,n)	kuesste(94,v)	
102	halt(184,int,adv)	hallt(5,v)	
103	rechte(180,v,a)	raechte(7,v)	
104	wirst(179,v)	wirrst(2,a)	
105	wert(170,a)	wehrt(45,v)	
106	Band(159,n,n,n,n)	bannt(2,v)	
107	weiten(156,a,v)	weihten(3,v)	
108	schafft(155,v)	Schaft(8,n)	
109	harte(152,a)	harrte(2,v)	
110	Rad(151,n)	Rat(445,n)	
111	willen(151,prep)	Villen(33,n)	
112	einsetzen(150,v)	Einsaetzen(12,n)	
113	weisen(146,v,a)	Waisen(3,n)	
114	Ware(139,n)	wahre(126,a,v)	

A list of 149 homophone sets in German

115	Zunahme(137,n)	Zuname(5,n)	
116	leeren(134,v,a)	lehren(41,v)	Lehren(98,n)
117	weite(130,a,v)	weihte(12,v)	
118	leid(127,a)	leiht(3,v)	
119	Gaesten(126,n)	Gesten(30,n)	
120	wahre(126,v,a)	Ware(139,n)	
121	ah(124,int)	A(563,n)	
122	wen(123,pron,pron,pron)	wehn(3,v)	
123	wider(122,prep)	wieder(7510,adv)	
124	merkte(121,v)	Maerkte(61,n)	
125	Saetzen(119,n)	setzen(604,v)	
126	wagen(119,v)	vagen(5,a)	Waagen(3,n)
127	Bayer(119,n)	Baier(4,n)	
128	Leib(117,n)	Laib(3,n)	
129	stelle(117,v)	Staelle(14,n)	
130	seid(114,v)	seit(3496,adv,prep)	
131	Last(114,n)	lasst(64,v)	
132	starten(112,v)	starrten(26,v)	
133	Lok(109,n)	Log(2,n)	
134	Held(105,n)	haelt(817,v)	
135	Frist(105,n)	frisst(25,v)	
136	Lieder(103,n)	Lider(15,n)	
137	Hemd(102,n)	hemmt(16,v)	
138	Welle(100,n)	Waelle(6,n)	
139	Lehren(98,n)	leeren(134,a,v)	
140	gewandt(98,v,a)	Gewand(40,n)	
141	wollt(96,v)	Volt(32,n)	
142	kuesste(94,v)	Kueste(186,n)	
143	Jagd(94,n)	jagt(27,v)	
144	Wende(90,n,n)	Waende(56,n)	
145	Wirt(89,n)	wird(18969,v)	
146	packte(89,v)	Pakte(2,n)	
147	Chor(88,n)	Korps(25,n)	
148	Wellen(86,n)	Waellen(7,n)	
149	Boot(86,n)	bot(203,v)	

2. Material used in the fieldwork for the study of the on-going sound change

2.1. The questionnaire

2.2. The word-list reading sheet

2.3. Data from the two questionnaire tests by 122 subjects

2.4. Data from the production tests by 42 subjects

Appendix 2.1

说明：您好！这是香港城市大学的一个语言学关于**广东话**的研究课题的调查问卷。非常谢谢您的参加和合作。问卷大概需要 20 分钟的时间。

下面有 22 组词，请您判断每组词里的三个带下划线的字的读音是否一样。如果您认为三个字中有一个字的读音和其他两个不一样，就请把该字用圆圈上；如果三个字的读音都一样，请圈右边的“ 全同” 两字；如果三个字的读音都不一样，请圈“ 全不同” ；如果您觉得不能肯定，请圈“ 不定” 。如果你不懂念其中某个字，就请在该字旁边写上 “？” 。

例：	广 东	通 过	冬 天	全同	全不同	不定
1	甘 心	今 天	黄 金	全同	全不同	不定
2	年 代	可 怜	连 长	全同	全不同	不定
3	旅 行	女 兵	吕 后	全同	全不同	不定
4	蜡 烛	腊 肠	纳 入	全同	全不同	不定
5	尼 姑	厘 米	离 开	全同	全不同	不定
6	努 力	老 板	俘 虏	全同	全不同	不定
7	怒 气	路 边	露 营	全同	全不同	不定
8	篮 球	蓝 色	南 京	全同	全不同	不定
9	兰 花	阻 拦	难 题	全同	全不同	不定
10	痢 疾	油 腻	利 益	全同	全不同	不定
11	狼 狗	囊 括	郎 君	全同	全不同	不定
12	兴 隆	农 民	龙 袍	全同	全不同	不定
13	尿 桶	廖 晖	料 计	全同	全不同	不定
14	凉 快	娘 亲	粮 食	全同	全不同	不定
15	阅 览	揽 住	腩 肉	全同	全不同	不定
16	落 水	快 乐	诺 言	全同	全不同	不定
17	宁 愿	零 星	灵 活	全同	全不同	不定
18	奴 隶	牢 房	劳 动	全同	全不同	不定
19	黎 族	犁 田	泥 土	全同	全不同	不定
20	法 国	爽 快	忽 然	全同	全不同	不定
21	锣 鼓	挪 动	罗 列	全同	全不同	不定
22	理 发	你 好	李 子	全同	全不同	不定

由于研究分析的需要，问卷后会收集一些与您的语言的使用情况有关的简单资料。请填写以下资料，所有资料只会用作学术研究用途。

1. 姓名：________________
2. 性别: ☐男 ☐女
3. 年龄: ☐14 或以下 ☐15-24 ☐25-34 ☐35-44 ☐45-54 ☐55-64 ☐65 或以上
4. 出生地: ________
5. 学历: ☐小学 ☐中学 ☐中专 ☐大专 ☐大学 ☐大学以上
6. 除了广东话以外，您是否还能熟练地讲其他的方言？
   ☐没有 ☐有, 是________（请在下面选择）
   A. 普通话 B.上海话 C.潮州话 D.客家话 E.湖南话
   F. 东莞话 G.台山话 H.中山话 I.顺德话 J.其他________
7. 您小时候在家中主要听到的方言：父亲：____母亲：____其他：________(请从上面的 A-J 中选)。

谢谢！

调查人：香港城市大学语言工程实验室 柯津云 (jyke@ee.cityu.edu.hk), 2003 年 9 月.

Appendix 2.2

说明：您好！非常感谢您刚才参加了关于广东话的研究课题的问卷调查。在问卷之外，我们还需要您的一些广东话的录音材料。请您按平常说话的速度和方式，先说出您的姓名和今天的日期，然后按照顺序号读一遍下面的三组单词。

被访人姓名：____________________ 日期：__________________

第一组

1	囊	括	11	腩	肉	21	粒	子	31	粒	子	41	牛	腩
2	纳	入	12	归	纳	22	发	难	32	聂	耳	42	南	京
3	溺	水	13	姑	娘	23	屎	尿	33	怒	气	43	花	泥
4	怒	气	14	腩	肉	24	囊	括	34	去	年	44	泥	土
5	你	好	15	粒	子	25	奴	隶	35	腻	烦	45	囊	括
6	匿	藏	16	农	民	26	年	代	36	溺	水	46	南	京
7	宁	愿	17	奴	隶	27	奈	何	37	泥	土	47	嫩	绿
8	花	泥	18	牛	腩	28	腩	肉	38	南	京	48	家	奴
9	你	好	19	尿	桶	29	西	南	39	纳	入	49	嫩	绿
10	难	题	20	姑	娘	30	西	宁	40	去	年	50	尼	姑

第二组

1	发	难	11	牛	腩	21	鸟	类	31	诺	言	41	少	女
2	发	怒	12	西	宁	22	你	好	32	挪	动	42	鸟	类
3	腻	烦	13	西	南	23	努	力	33	家	奴	43	匿	藏
4	耐	力	14	扭	伤	24	鸟	类	34	挪	动	44	腻	烦
5	嫩	绿	15	女	兵	25	难	题	35	女	兵	45	内	外
6	哈	尼	16	诺	言	26	许	诺	36	耐	力	46	屎	尿
7	年	代	17	暖	气	27	花	泥	37	归	纳	47	花	农
8	油	腻	18	扭	伤	28	油	腻	38	胆	囊	48	努	力
9	女	兵	19	许	诺	29	娘	亲	39	年	代	49	哈	尼
10	哈	尼	20	扭	伤	30	挪	动	40	朽	木	50	耐	力

第三组

1	归	纳	11	奈	何	21	匿	藏	31	农	民	41	西	宁
2	诺	言	12	难	题	22	娘	亲	32	男	人	42	尿	桶
3	尼	姑	13	农	民	23	花	农	33	发	怒	43	发	难
4	去	年	14	纳	入	24	男	人	34	男	人	44	宁	愿
5	怒	气	15	花	农	25	努	力	35	许	诺	45	朽	木
6	尿	桶	16	发	怒	26	尼	姑	36	内	外	46	聂	耳
7	娘	亲	17	朽	木	27	内	外	37	胆	囊	47	西	南
8	姑	娘	18	泥	土	28	溺	水	38	暖	气	48	暖	气
9	家	奴	19	奴	隶	29	宁	愿	39	少	女	49	奈	何
10	油	腻	20	屎	尿	30	聂	耳	40	胆	囊	50	少	女

辛苦您了。非常感谢您对我们研究的帮助！

调查人：香港城市大学语言工程实验室 柯津云 (jyke@ee.cityu.edu.hk), 2003 年 9 月.

Appendix 2.3. Data from the two questionnaire tests by 122 subjects

subject index	年	宁	尿	农	尼	你	腻	女	娘	挪	奴	努	怒	囊	诺	泥	南	腩	难	纳
2	n	n	n	n	n	n	n	n	n	n	f/l	n	n	n	n	n	n	n	n	n
	n	n	l	n	n	n	n	n	n	n	n	l	n	n	n	n	n	n	n	n
3	f/l	l	l	l	n	l	l	l	l	l	l	n	n	l	n	f/l	l	f/l	n	l
	l	l	l	l	n	l	n	l	l	n	l	l	l	l	n	f/l	l	n	n	l
4	n	n	n	n	n	n	l	f/l	n	n	n	n	n	n	n	n	n	n	n	n
	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
5	l	l	n	n	l	n	l	l	f/l	?	n	l	n	l	n	l	n	f/l	n	l
	l	l	?	l	n	n	n	n	l	n	n	n	n	l	n	l	n	f/l	n	l
6	n	n	n	n	n	n	f	n	n	n	n	n	n	n	n	n	n	n	n	n
	n	n	f	n	l	n	f	n	n	n	n	n	n	n	l	n	n	n	n	n
7	l	f	l	l	f	l	l	l	l	l	l	l	l	f	l	f	f	l	f	f
	l	l	l	l	l	l	l	l	l	l	f	l	l	f	l	f	f	l	f	f
14	l	l	l	l	l	l	n	l	l	l	l	f/l	l	l	l	?	l	f/l	l	l
	l	l	l	l	l	l	n/?	l	l	n/?	l	f/l	l	l	l	f/l	l	f/l	l	l
17	l	l	l	l	l	l	?	l	l	l	l	l	l	l	l	?	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
21	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
22	l	l	l	l	l	l	l	l	l	f/l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f	l	l
27	l	l	l	l	l	f/l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
28	n	l	n	l	n	n	n	l	l	n	l	l	n	n	n	n	l	n	n	n
	n	l	n	l	n	l	n	l	l	n	l	l	n	n	n	n	l	n	n	n
29	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
30	l	l	l	l	l	f/l	l	l	l	l	l	f/l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
35	n	l	l	?	n	n	l	l	l	n	n	l	l	n	n	l	l	l	l	l
	n	n	f/l	n	n	n	n	l	n	n	n	n	n	n	n	f/l	n	n	n	n
36	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
43	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	n	l	?
	f/l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n/?	l	n/?	l	l
44	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	?	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	?	l	l	l	l
45	l	l	f/l	l	l	l	?	l	l	l	l	l	l	?	l	f/l	l	n	l	l
	l	l	l	f/l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	f/l	l	l
47	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
	f/l	?	f/l	?	?	l	l	l	l	?	l	l	l	n	l	l	l	f/l	l	l
48	l	n	?	l	l	l	?	l	l	l	n	f/l	l	l	l	?	l	f/l	l	l
	l	l	l	?	l	l	l	l	l	n	l	l	l	l	l	f/l	l	f	n	l
52	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
54	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	?	l	f/l	l	l
	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	?	l	f	l	l
56	l	l	?	l	l	l	l	l	l	f	l	l	l	l	l	?	l	f	l	l
	l	f	?	l	l	l	l	l	l	f	l	l	l	l	f	?	l	?	l	l
57	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n?	l	f	l	l
58	n	l	f/l	l	l	l	l	l	l	l	?	l	n	l	l	l	n	l	n	n
	f/l	n	?	l	l	f/l	l	f/l	l	l	f/l	l	l	l	l	l	n	l	n	l
59	l	l	l	l	l	l	f/l	l	l	l	l	l	l	l	l	l	l	l	l	l
	l	l	l	l	l	l	f/l	l	l	l	l	f/l	l	l	l	l	l	f/l	l	l

Appendix 2.3. Data from the two questionnaire tests by 122 subjects

60	?	f	?	n	?	n	?	l	l	n	f	f/l	n	n	f/l	n	n	f/l	?	?
	f/l	f/l	l	n	f/l	n	l	f/l	f/l	l	f	f/l	n	f	f	f/l	n	n	n	l
61	l	l	l	l	?	l	?	l	l	?	l	l	l	?	l	?	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	?	l	l
62	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
63	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	n?	l	l	l	l	l	l	l	f/l	l	l
64	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	f/l	l	l
65	l	l	l	l	n	l	?	l	f/l	l	l	l	l	l	l	?	l	f/l	n	n
	n	f/l	l	l	l	l	l	l	f/l	l	l	l	?	l	f/l	?	l	f/l	l	n
66	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
67	l	l	?	l	l	l	l	l	l	?	l	l	l	l	l	?	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
68	l	l	l	l	l	l	?	l	l	f	l	l	l	?	?	l	l	?	l	l
	l	l	?	l	l	l	l	l	l	?	l	f	l	l	l	l	l	l	l	l
69	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	n/?	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
70	l	n	n	n	n	n	l	l	l	l	n	n	l	l	l	?	n	?	n	l
	l	n	n	n	n	n	n	n	l	l	f/l	n	n	l	l	?	n	l	n	n
71	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f	l	l
	l	l	l	l	l	l	l	l	l	n?	l	l	l	l	l	l	l	l	l	l
72	l	l	?	l	l	l	?	l	l	l	l	l	l	?	l	?	l	f/l	l	l
	l	l	l	l	l	l	?	l	l	l	l	l	l	?	l	l	l	f/l	l	l
73	l	l	f/l	l	l	l	n/?	l	l	n/?	l	l	l	l	l	l	l	f/l	l	l
	l	l	f/l	l	l	l	l	l	l	n?	l	l	l	l	l	l	l	f/l	l	l
74	n	f/l	f/l	l	l	l	l	l	l	f	l	l	l	l	l	f/l	l	n	l	n
	f/l	l	l	l	l	l	l	f/l	l	f/l	l	l	n	l	l	l	l	f/l	l	l
75	l	l	?	l	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	l
	l	l	?	l	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	l
78	l	l	l	n	l	n	f/l	l	l	n	f/l	l	l	n	l	f/l	l	n	n	n
	n	l	l	f/l	l	l	f/l	l	l	f/l	f/l	l	l	f/l	l	l	n	l	n	n
79	n	n	l	f	l	f/l	l	n	l	n	n	l	n	n	l	l	l	f/l	n	l
	n	n	l	n	l	n	l	l	l	n	n	l	l	n	l	l	l	n	n	n
80	l	l	l	l	l	l	l	l	l	?	l	l	l	l	l	l	f	?	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	?	l	f	l	l
81	l	l	l	l	l	f	l	l	l	l	l	l	l	l	l	?	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	f	f	l	l	f	l	l	l	l	l
82	l	l	?	l	l	l	l	l	l	l	l	l	l	l	l	?	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	f	l	l	l	l	l	?	l	?	l	l
83	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
	l	l	l	l	l	?	l	l	l	l	l	?	l	l	l	l	l	l	l	l
84	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	?	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
85	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
86	l	?	?	l	l	l	l	l	?	n/?	l	l	l	l	l	f/l	l	f/l	l	l
	l	l	l	l	l	l	l	l	l	n/?	l	l	l	l	l	f/l	l	f/l	l	l
102	n	l	n	n	n	n	n	n	l	n	l	n	n	n	l	n	n	n	n	n
	n	n	n	n	n	n	?	n	l	n	l	n	n	?	l	n	n	?	n	n
112	n	n	n	n	n	n	n	n	n	n	n	n	n	n	l	n	n	n	n	n
	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
123	n	n	l	n	l	n	n	l	n	n	n	n	n	n	n	n	n	n	n	n

Appendix 2.3. Data from the two questionnaire tests by 122 subjects

	n	n	n	n	n	l	?	l	l	n	n	n	n	n	n	n	n	n	n	n
125	l	l	n	?	n	f	f	n	l	n	n	n	l	n	n	l	l	n	?	n
	l	l	f	l	n	n	n	n	l	n	l	n	n	n	n	n	n	n	?	l
133	l	f/l	f/l	l	n	l	n	l	l	n	n	f/l	n	l	n	n	n	f/l	f/l	l
	n	l	n	n	n	l	n	n	l	n	l	l	n	l	n	n	l	f/l	n	n
137	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l
144	n	l	l	l	n	n	n	n	n	n	l	l	l	n	n	n	n	n	n	n
	n	l	l	l	n	l	n	n	n	n	l	l	l	n	n	n	n	n	n	n
145	n	l	n	l	n	f	l	l	n	l	l	n	l	n	n	n	l	f	l	l
	n	l	n	l	n	n	n	l	n	n	l	n	n	n	l	n	n	n	n	n
150	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n
	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n

1. "f/l" means the subject gave a "failed" response (since he chose a non-target morpheme as the one having a different pronunciation from the other two morphemes), but we can judge based on this that the subject's pronunciation of the target morpheme has changed to /l-/.

2. "n/?" means the subject gave a response which may be interpreted as that the target morpheme has an /n-/ (since the subject chose the target morpheme as the one having a different pronunication from the other two morphomes), but we are not sure of this inferfence, since there are other reasons which would result in the same choice. Especially when such an /n-/ appears in only one or two cases among the set of target morphemes, this /n-/ is dubious.

Appendix 2.4. Data from the production tests by 42 subjects

subject index		1	2	2	3	3	4	4	5	6	6	34	35	35	36	36	37	37	45	60	62	70	74	85	86	86	102	104	112	123	125	132
age group		5	3	3	5	5	3	3	5	3	3	3	2	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2	2	2	2	5
gender		F	F	F	F	F	F	F	F	F	F	F	M	M	M	M	M	M	F	F	M	M	F	F	M	M	M	F	F	F	F	F
1	年 代	2n1f	n	n	l	v	2n1f	n	v	n	n	v	n	n	l	l	l	l	l	l	l	v	v	v	n	n	n	n	n	v	n	n
2	宁 愿	v	n	n	l	l	n	l	l	n	n	v	v	v	l	l	l	l	l	l	l	l	l	v	l	l	v	l	n	v	l	v
3	尿 桶	n	n	n	l	l	n	n	v	n	n	l	n	n	v	v	l	l	l	?	l	v	v	v	n	n	n	n	n	v	v	v
4	农 民	l	n	n	l	l	n	n	l	n	n	l	l	l	l	l	l	l	l	l	l	v	l	l	l	l	n	v	v	v	l	l
5	尼 姑	l	n	n	l	l	n	n	n	n	n	l	v	n	v	n	l	v	l	l	v	n	v	l	v	v	n	n	n	n	v	v
6	你 好	v	n	n	l	l	n	l	v	n	n	v	n	v	v	v	l	l	l	l	l	l	v	l	v	n	n	v	n	v	v	l
7	腻 烦	v	f	2n1?	l	f	n	2n1f	l	n	n	l	v	2f1l	l	l	l	v	2f1l	f	f	f	f	l	v	n	n	n	n	n	f	2n1f
8	女 兵	l	n	n	l	l	1n2f	l	l	n	n	l	l	l	l	l	l	l	l	l	l	l	l	l	v	l	n	v	l	v	l	l
9	娘 亲	l	n	n	l	l	v	n	l	n	n	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	v	v	v	l	v
10	挪 动	l	n	n	l	l	n	n	l	n	n	v	n	n	v	v	l	l	l	f	l	l	l	v	l	n	n	n	n	n	v	v
11	奴 隶	l	n	n	l	l	n	n	l	n	n	l	l	l	v	v	l	l	l	l	l	v	l	l	l	l	2n1f	v	v	v	l	l
12	努 力	l	n	n	l	l	n	v	l	n	n	l	n	v	v	v	l	l	l	l	l	n	v	l	v	v	n	n	n	v	n	l
13	怒 气	l	n	n	l	l	n	n	l	n	n	l	l	l	v	v	l	l	l	l	l	n	l	l	l	v	n	n	n	v	v	v
14	**囊** 括	l	n	n	l	l	n	n	l	n	n	l	l	l	l	l	l	l	l	l	l	v	l	2l1f	l	l	v	l	n	v	f	l
15	诺 言	v	n	n	v	v	2n1	n	l	n	n	l	v	n	v	v	l	l	j	l	l	2l1f	v	l	l	v	l	n	v	n	v	v
16	泥 土	v	n	n	l	l	n	n	l	n	n	l	v	l	v	v	l	l	l	l	l	n	l	l	v	l	n	v	v	v	v	v
17	南 京	v	n	n	l	l	n	l	l	n	n	l	v	l	l	l	l	l	l	l	l	v	v	l	l	l	n	l	n	v	l	v
18	腩 肉	l	n	n	l	l	n	n	l	n	n	l	v	2l1?	l	l	l	l	2l1f	2l1f	l	n	v	l	l	2l1f	v	l	n	v	l	l
19	难 题	l	n	n	v	l	n	n	v	n	n	n	v	l	v	l	l	l	l	l	l	n	l	l	l	v	n	v	n	v	v	l
20	纳 入	l	n	n	l	l	n	n	l	n	n	l	v	v	l	v	l	l	l	2l1f	f	v	l	l	l	l	n	v	n	n	l	v
21	奈 何	l	n	n	l	l	n	n	l	n	n	l	v	n	l	v	l	v	l	l	l	v	l	l	v	n	n	n	n	n	v	v
22	内 外	l	n	n	l	l	n	n	v	n	n	l	v	v	l	l	l	l	l	l	l	v	l	l	l	v	n	n	n	n	l	v
23	耐 力	l	n	n	l	l	n	n	l	n	n	l	n	v	v	l	l	l	l	l	l	l	l	l	v	v	n	n	n	v	l	v
24	男 人	l	n	n	l	l	n	l	l	n	n	v	n	v	l	v	l	l	l	l	l	v	l	l	l	l	n	v	n	n	l	l
25	暖 气	v	n	n	l	l	n	v	v	n	n	l	n	n	l	l	l	l	l	v	l	l	l	l	v	v	n	v	n	v	n	n
26	嫩 绿	n	n	n	l	l	n	n	l	n	n	l	n	n	l	l	l	l	l	2l1?	l	v	l	l	v	n	n	n	n	n	v	v
27	粒 子	v	l	l	l	l	v	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	v	l	l	l	l	l
28	聂 耳	n	v	n	l	l	n	n	v	n	n	v	n	n	l	v	l	v	l	l	l	l	l	f	v	n	n	v	v	v	v	v
29	溺 水	v	n	n	l	v	n	n	v	n	n	l	v	l	l	l	l	v	f	1j2f	f	f	v	l	l	l	n	v	n	v	v	v
30	匿 藏	j	n	n	l	l	n	2n1f	v	n	n	l	v	v	l	l	l	l	f	l	f	f	l	l	l	v	n	n	n	2n1f	v	v
31	鸟 类	n	n	n	l	l	n	n	l	n	n	v	n	n	v	v	l	l	l	l	l	v	v	v	n	n	n	n	n	n	n	v
32	朽 木	v	f	f	l	v	l	l	l	n	n	l	l	l	l	l	l	l	h	f	f	f	2h1f	h	l	n	2f1h	2l1f	f	f	2f1h	l
33	扭 伤	l	n	n	l	l	2n1	n	l	n	n	l	v	v	l	l	l	l	l	l	l	l	l	l	l	l	n	v	v	n	v	v

Appendix 2.4. Data from the production tests by 42 subjects

subject index		1	2	2	3	3	4	4	5	6	6	34	35	35	36	36	37	37	45	60	62	70	74	85	86	86	102	104	112	123	125	132
age group		5	3	3	5	5	3	3	5	3	3	3	2	2	2	2	2	2	1	1	1	1	1	2	2	2	2	2	2	2	2	5
gender		F	F	F	F	F	F	F	F	F	F	F	M	M	M	M	M	M	F	F	M	M	F	F	M	M	M	F	F	F	F	F
34	去 年	l	n	n	l	l	n	v	v	n	n	l	v	n	l	l	l	l	l	l	l	v	l	l	v	v	n	l	v	n	l	v
35	西 宁	l	n	n	l	l	2n1	l	l	n	n	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n	l	n	v	l	l
36	屎 尿	n	n	n	l	l	n	v	n	n	n	l	v	v	v	l	l	l	l	l	l	l	l	v	l	l	n	v	v	l	v	l
37	花 农	l	n	n	l	l	n	n	l	n	n	l	v	v	l	l	l	v	l	l	l	n	l	l	l	l	n	v	n	n	l	l
38	哈 尼	l	n	n	l	l	2n1	n	l	n	n	l	v	v	n	l	l	l	l	l	l	v	l	l	v	v	n	v	n	n	n	l
39	油 腻	l	n	n	v	l	2n1	n	l	n	n	l	l	l	v	v	l	l	l	l	f	v	l	l	l	v	n	v	v	n	n	l
40	少 女	l	n	n	l	l	v	l	l	n	n	l	v	v	l	v	l	l	l	l	l	l	l	l	l	l	n	l	v	v	v	l
41	姑 娘	l	n	n	l	l	v	l	l	n	n	v	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	v	v	l	l
42	家 奴	l	n	n	l	l	v	v	l	n	n	l	n	n	v	l	l	l	l	l	l	n	l	l	l	v	v	v	v	v	n	l
43	发 怒	l	n	n	l	l	n	n	l	n	n	l	v	v	l	v	l	l	2l1f	l	l	v	l	l	l	v	n	n	n	n	l	l
44	胆 囊	l	n	n	l	l	n	n	n	n	n	l	n	n	l	l	l	l	l	l	l	v	n	l	v	v	v	l	n	n	v	v
45	许 诺	l	n	n	v	j	n	n	l	n	n	l	n	v	l	l	l	l	l	2l1f	l	l	v	l	l	v	l	n	l	n	n	l
46	花 泥	l	n	n	l	l	n	n	l	n	n	l	v	l	n	v	l	l	l	l	l	v	l	l	l	l	n	l	v	v	n	l
47	西 南	l	n	n	l	l	2n1	v	l	n	n	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n	l	n	n	l	l
48	牛 腩	l	n	n	l	l	v	l	l	n	n	l	l	l	l	l	l	l	l	l	l	l	l	l	l	l	n	l	v	l	l	l
49	发 难	l	n	n	l	l	n	2n1f	l	n	n	l	v	l	l	l	l	l	l	l	l	n	v	l	l	l	n	n	2n1f	n	l	l
50	归 纳	l	n	n	l	l	n	n	l	n	n	l	v	v	v	v	l	l	l	l	f	v	l	l	l	l	n	v	n	n	l	l
total "n"		5	46	47	0	0	35	30	3	49	49	1	14	13	2	1	0		0	0	0	9	1	0	3	10	38	17	30	21	9	2
total "l"		33	1	1	46	44	1	11	37	1	1	41	13	19	32	31	50		43	40	42	17	35	41	33	23	4	13	3	3	21	26
total "v"		10	1	0	4	4	6	6	10	0	0	8	23	16	16	18	0		0	1	1	19	12	6	14	16	6	19	15	24	17	21
others		2	2	2	0	2	8	3	0	0	0	0	0	2	0	0	0	0	7	9	7	5	2	3	0	1	2	1	2	2	3	1

Appendix 2.4. Data from the production tests by 42 subjects

subject index		133	134	134	136	137	137	137	137	137	138	139	140	141	141	142	143	144	145	146	151	152	153	154	155	155	156
age group		3	6	6	5	1	1	3	1		7	3	7	1	1	4	4	6	3	2	5	5	4	4	5	5	5
gender		M	F	F	F	F	F	F	F		F	F	F	F	F	M	F	F	M	M	M	M	M	M	M	M	M
1	年 代	n	v	v	v	v	v	n	n	v	n	n	n	n	n	n	v	n	n	v	v	l	v	l	v	v	n
2	宁 愿	n	v	n	l	v	n	v	n	v	v	v	n	v	n	n	v	l	n	v	v	l	l	l	v	l	n
3	尿 桶	n	v	v	n	n	n	n	n	n	v	n	n	n	n	n	l	n	n	n	v	l	v	v	n	v	n
4	农 民	n	l	l	l	n	v	n	n	v	v	n	v	n	v	n	l	l	n	l	l	l	v	l	l	l	n
5	尼 姑	n	v	n	l	v	v	v	n	v	v	n	n	n	n	n	l	n	v	v	l	l	v	l	l	l	n
6	你 好	n	v	v	l	n	n	n	n	n	v	n	n	n	n	n	v	n	n	n	n	l	n	v	l	l	n
7	腻 烦	n	v	v	f	n	n	v	n	v	f	n	f	n	n	2n1f	v	n	n	l	l	f	v	l	f	l	f
8	女 兵	n	l	l	l	v	v	v	v	v	v	n	v	v	v	n	l	n	v	v	l	l	l	l	v	v	v
9	娘 亲	n	l	l	v	l	v	v	v	v	v	n	v	v	n	n	l	n	v	v	l	l	l	l	v	l	v
10	挪 动	n	l	l	f	n	n	n	n	n	v	n	n	n	n	n	l	n	n	v	l	l	l	l	l	l	n
11	奴 隶	n	l	l	l	n	n	n	v	v	v	n	v	n	v	n	l	l	v	l	l	l	l	l	v	l	n
12	努 力	n	l	l	l	n	n	n	n	n	v	n	n	v	n	n	l	l	l	v	l	l	l	l	v	v	n
13	怒 气	v	l	l	l	n	v	n	n	v	v	n	v	n	n	n	l	v	n	v	l	l	l	l	l	v	n
14	囊 括	n	l	l	2f1l	v	v	v	v	v	l	2n1f	l	v	v	l	l	n	n	l	l	l	l	l	l	l	v
15	诺 言	n	v	l	l	l	v	v	v	v	v	n	v	n	v	n	l	v	n	l	v	j	l	l	v	v	n
16	泥 土	n	l	l	l	n	v	n	n	v	l	n	n	v	n	n	l	n	n	l	l	l	l	l	l	v	n
17	南 京	n	v	v	l	n	v	n	n	v	v	n	v	v	v	n	l	n	v	l	l	v	l	l	l	l	n
18	腩 肉	n	l	v	l	n	n	n	n	n	v	v	l	v	v	n	v	n	v	l	l	l	v	l	v	l	n
19	难 题	n	l	l	l	n	2n1f	v	n	v	l	n	v	n	n	n	v	n	v	l	v	l	v	l	l	l	n
20	纳 入	n	l	v	l	n	n	n	n	n	n	n	v	v	v	n	l	n	n	v	l	l	n	l	v	v	n
21	奈 何	n	l	l	f	n	v	n	n	v	v	n	v	n	n	n	v	n	n	l	l	l	v	l	l	l	n
22	内 外	n	l	l	l	v	n	n	n	v	l	n	v	v	v	n	l	n	n	v	v	l	l	l	l	l	n
23	耐 力	n	l	l	f	n	v	v	n	v	v	n	v	n	n	n	l	n	n	v	l	l	v	l	l	l	n
24	男 人	n	n	v	l	n	v	v	n	v	v	v	n	v	n	n	l	n	v	v	v	l	n	l	l	v	v
25	暖 气	n	n	n	l	n	n	n	n	n	n	n	n	n	n	n	v	n	n	v	v	l	v	l	l	v	n
26	嫩 绿	n	n	n	f	n	v	n	n	v	n	n	n	n	n	n	n	n	n	v	l	f	f	l	v	v	n
27	粒 子	v	l	l	l	l	l	l	v	v	l	l	l	l	v	l	l	n	v	l	l	l	l	l	l	v	l
28	聂 耳	n	v	v	f	v	l	v	v		n	n	n	n	n	n	v	l	n	l	n	l	v	l	v	v	n
29	溺 水	n	n	v	f	j	j	j	j	j	v	n	f	j	j	v	v	n	n	l	j	j	l	j	j	j	n
30	匿 藏	n	l	v	f	l	j	l	l	v	n	n	n	n	n	n	l	v	n	v	j	f	l	l	v	v	n
31	鸟 类	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	n	l	n	v	n	l	v	v	v	v	n
32	朽 木	n	l	l	h	h	f	h	h	h	v	v	n	h	h	v	v	l	n	v	l	f	h	f	h	h	n
33	扭 伤	n	l	l	f	v	l	n	v	v	v	n	n	n	n	n	l	n	n	v	l	l	v	l	v	l	n

Appendix 2.4. Data from the production tests by 42 subjects

subject index		133	134	134	136	137	137	137	137	137	138	139	140	141	141	142	143	144	145	146	151	152	153	154	155	155	156
age group		3	6	6	5	1	1	3	1		7	3	7	1	1	4	4	6	3	2	5	5	4	4	5	5	5
gender		M	F	F	F	F	F	F	F		F	F	F	F	F	M	F	F	M	M	M	M	M	M	M	M	M
34	去 年	n	n	n	l	n	v	n	n	v	l	n	n	v	v	n	l	n	n	v	l	l	l	l	l	l	n
35	西 宁	v	l	l	l	v	v	n	v	v	l	n	n	v	l	n	l	l	n	l	l	l	l	l	l	l	n
36	屎 尿	n	l	l	v	n	n	v	n	v	l	n	n	n	n	n	l	n	n	l	l	l	l	l	l	l	n
37	花 农	n	l	l	l	n	v	n	n	v	l	v	v	l	2l1?	n	l	l	v	l	l	l	l	l	l	l	n
38	哈 尼	n	l	l	l	n	n	n	v	v	l	n	l	n	n	n	l	n	n	v	l	l	l	l	l	l	n
39	油 腻	l	l	l	2f1l	n	v	n	v	v	v	n	f	n	n	n	l	n	n	v	l	l	n	l	l	l	n
40	少 女	n	l	l	l	v	v	n	n	v	l	n	n	n	v	n	l	n	n	l	l	l	l	l	l	l	n
41	姑 娘	v	l	l	l	n	l	v	n	v	l	n	l	l	l	n	l	n	n	l	l	l	l	l	l	l	n
42	家 奴	n	l	l	l	2n1f	v	n	n	v	l	n	l	n	v	n	l	l	n	l	l	l	l	l	l	l	n
43	发 怒	n	l	l	2f1l	n	n	n	n	n	l	n	n	2n1f	2f1n	n	l	v	n	l	l	l	l	l	l	l	n
44	胆 囊	l	l	l	f	v	l	n	v	v	n	n	n	n	v	n	l	n	n	l	l	l	l	l	v	n	l
45	许 诺	n	l	l	2l1f	v	v	n	n	v	l	n	l	v	v	v	l	v	n	l	v	l	l	l	l	l	n
46	花 泥	n	l	l	l	n	v	n	n	v	l	n	v	v	v	n	l	n	n	l	l	l	l	l	l	l	n
47	西 南	v	l	l	l	n	n	n	n	n	l	n	v	v	v	n	l	n	n	l	l	l	l	l	l	l	n
48	牛 腩	v	l	l	l	n	n	n	v	v	l	n	v	l	l	n	l	n	n	l	l	l	l	l	l	l	n
49	发 难	n	l	l	2l1f	n	v	n	n	v	l	n	v	v	v	n	l	n	n	l	l	l	l	l	l	l	n
50	归 纳	v	l	l	l	n	n	v	n	v	l	n	v	n	v	n	l	v	n	l	l	l	l	l	l	l	n
total "n"		41	6	6	2	31	18	32	34		8	43	22	26	24	44	2	34	39	2	3	0	4	0	1	1	43
total "l"		2	35	33	29	4	5	2	1		20	1	7	4	3	2	37	10	1	27	36	43	31	45	31	32	2
total "v"		7	9	11	3	12	23	14	13		21	5	18	17	19	3	11	6	10	21	9	1	13	3	15	15	4
others		0	0	0	16	3	4	2	2	0	1	1	3	3	4	1	0	0	0	0	2	6	2	2	3	4	1