

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.

# Table of Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Theoretical framework: Self-organization and language evolution</b>	<b>1</b>
1.1 Three perspectives on language evolution – biological, cultural, or both?	2
1.1.1 Biological evolution of language . . . . .	2
1.1.2 Cultural evolution of language . . . . .	7
1.1.2.1 Functional explanations . . . . .	9
1.1.2.2 Controversies on functional explanations . . . . .	13
1.1.3 Co-evolution of language and brain . . . . .	14
1.2 Self-organization: A new scientific paradigm . . . . .	16
1.2.1 A brief introduction to the theory of self-organization . . . . .	16
1.2.2 Honeybee comb as a studied case and its implications . . . . .	20
1.3 Self-organization and language evolution . . . . .	26
1.3.1 Existing proposals of self-organization theory applied in linguistics	26
1.3.2 Self-organization in the language system . . . . .	29
1.3.3 Self-organization in the population and in the individual . . . . .	32
1.3.4 Two levels of existence and three levels of analysis . . . . .	38
1.3.5 A language-ready body - properties of the parts . . . . .	40
1.3.6 Competing motivations in interactions . . . . .	44
1.3.7 Universals, grammars and exceptions . . . . .	49
1.3.8 Rapid evolution as a result of self-organization . . . . .	52
1.4 Organization of the thesis . . . . .	54
<b>2 Self-organization in the system: A case study on homophony</b>	<b>58</b>
2.1 Some background of homophony . . . . .	59
2.1.1 What is “homophony”? . . . . .	59
2.1.2 The rise of homophony . . . . .	60
2.1.2.1 Homophones from sound change . . . . .	60
2.1.2.2 Homophones from borrowing . . . . .	61
2.1.3 The fall of homophony . . . . .	61
2.1.3.1 Split from sound change . . . . .	61
2.1.3.2 Replacement by synonyms . . . . .	63

2.1.4	The exaptive usage of homophony . . . . .	64
2.2	Cross-language comparison of the degree of homophony . . . . .	64
2.2.1	Methodology of comparison and data description . . . . .	65
2.2.2	Degrees of homophony in Chinese dialects . . . . .	66
2.2.2.1	Degrees of homophony in three Germanic languages . . . . .	70
2.3	Homophony and phonological resource . . . . .	73
2.3.1	Measuring phonological resource . . . . .	73
2.3.2	Prediction of the degree of homophony . . . . .	76
2.4	Self-organization in homophony . . . . .	80
2.4.1	Disyllabification in Chinese . . . . .	81
2.4.1.1	Measure of degree of disyllabification . . . . .	83
2.4.1.2	Disyllabification and homophony avoidance . . . . .	85
2.4.2	Grammatical differentiation between homophones . . . . .	95
2.5	Self-organization in the lexicon: Monosyllabicity and lexicalization . . . . .	98
<b>3</b>	<b>Self-organization in the population and individuals: A case study on sound change</b>	<b>103</b>
3.1	Why individual language users? . . . . .	103
3.2	The empirical case study of Cantonese /n-/→/l-/ . . . . .	107
3.2.1	Historical background of the change . . . . .	107
3.2.2	Methodology for data collection and processing . . . . .	112
3.2.3	Summary of collected data . . . . .	115
3.3	Synchronic variation in the on-going change . . . . .	116
3.3.1	Three types of variation . . . . .	116
3.3.2	Variation in the population: VT-I . . . . .	121
3.3.3	Variation in the lexicon: VT-II . . . . .	122
3.3.4	The puzzle of ‘li4’ (“grain”) . . . . .	126
3.3.5	Variation in the individual: VT-III . . . . .	128
3.4	Disparity between production and perception . . . . .	132
3.4.1	Variations VT-I and VT-II in questionnaire data . . . . .	133
3.4.2	Cases of near-merger . . . . .	134
3.5	Individual differences and two hypothetical lexical learning styles . . . . .	136
3.5.1	Different types of speakers . . . . .	136
3.5.2	Explanations for individual differences . . . . .	137
3.5.3	Two hypothetical learning styles . . . . .	139
3.6	Interpretation of the diachronic change: Dialectal influence or internal change . . . . .	142
3.6.1	A uni-modal pattern . . . . .	142
3.6.2	Influence from PTH on Cantonese . . . . .	143
3.6.3	Explanation for the uni-modal pattern . . . . .	146
3.6.4	Differences in actuation: Dialectal contact or internal change . . . . .	149
3.6.5	More on <i>coexistence of variants</i> . . . . .	150
<b>4</b>	<b>A review of computational modeling in language evolution</b>	<b>153</b>
4.1	Modeling at different time scales . . . . .	154
4.1.1	Language acquisition . . . . .	156
4.1.2	Language change . . . . .	160
4.1.3	Language origin . . . . .	166

4.1.3.1	Functional approach . . . . .	168
4.1.3.2	Emergent approach . . . . .	172
4.2	Modeling at different levels of resolution . . . . .	179
4.2.1	The first level – language as a synthetic whole . . . . .	179
4.2.2	The second level – subsystem as an independent whole . . . . .	182
4.2.3	The third level – language embodied in use . . . . .	183
4.2.4	The fourth level – meaning and/or form embodiment . . . . .	184
4.3	Agent-based modeling . . . . .	186
4.4	Relationship between production and perception . . . . .	188
4.4.1	Evidence of the disparity . . . . .	188
4.4.2	Role in language evolution . . . . .	189
4.4.3	Modeling the disparity . . . . .	191
4.5	Pros and cons of computational modeling . . . . .	193
<b>5</b>	<b>Some simulation models of language origin and change</b>	<b>197</b>
5.1	Some models of language origin . . . . .	198
5.1.1	The emergence of vocabulary . . . . .	198
5.1.1.1	The imitation model . . . . .	200
5.1.1.2	The self-organization model . . . . .	201
5.1.1.3	Phase transition . . . . .	206
5.1.2	The evolution of homonyms . . . . .	209
5.2	Simulation models of language change . . . . .	213
5.2.1	Language change as a diffusion process . . . . .	214
5.2.2	The threshold problem of language change . . . . .	215
5.2.3	The S-curve dynamical pattern of language change . . . . .	216
5.2.4	Importance of social networks . . . . .	218
5.2.5	The baseline model . . . . .	224
5.2.6	Model with age structure . . . . .	226
5.2.7	Different functional biases toward the innovation . . . . .	228
5.2.8	Effect of population size . . . . .	232
5.2.9	Effect of two types of learners . . . . .	233
<b>6</b>	<b>Conclusions, discussions and future work</b>	<b>239</b>
6.1	Self-organization as a unifying framework . . . . .	239
6.2	Investigations at three levels: system, population, and individual . . . . .	240
6.3	Heterogeneity as the norm . . . . .	241
6.4	Computational modeling as a promising methodology . . . . .	243
6.5	Future directions . . . . .	244
6.5.1	Cross-dialectal investigation of /n-/ $\longleftrightarrow$ /l-/ changes . . . . .	244
6.5.2	Psychological reality of the phonological rules . . . . .	246
6.5.3	Further development of the model of language change . . . . .	248
6.5.4	More interactions between modeling and empirical studies . . . . .	250
	<b>Bibliography</b>	<b>253</b>

<b>Appendices</b>	<b>284</b>
1 Homophony . . . . .	284
1.1 A collection of artistic usage of homophones . . . . .	284
1.2 A collection of ambiguities in actual interactions . . . . .	285
1.3 Lists of homophones in the first 5000 words in three Germanic languages . . . . .	289
2 Material used in the fieldwork for the study of the on-going sound change	308
2.1 The questionnaire . . . . .	309
2.2 The word-list reading sheet . . . . .	310
2.3 Data from the two questionnaire tests by 122 subjects . . . . .	311
2.4 Data from the production tests by 42 subjects . . . . .	314

# List of Figures

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). . . . .	21
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. . . . .	25
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. . . . .	34
1.4	Three levels of analysis of language seen at two points of time: System, population (communal language) and individual (idiolect). . . . .	39
2.1	The relationship between “homophone”, “homograph”, and “homonym”. . . . .	60
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. . . . .	69
2.3	The correlation between the size of syllable inventory and the degree of homophony in terms of percentage of homophone pairs. . . . .	69
2.4	The correlation between the size of syllable inventory and the degree of homophony in terms of average number of homophones per syllable. . . . .	70
2.5	Degree of homophony in the first 5000 frequent words in three Germanic languages. . . . .	72

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. . . . .	77
2.7	Distribution of the frequency of syllable types in the Dutch lexicon. . .	77
2.8	Distribution of the frequency of syllable types in the German lexicon. .	78
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. . . . .	79
2.10	Degree of monosyllabicity in the first 5000 words in three languages. .	80
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. .	95
2.12	The degrees of homophony versus the degrees of disyllabification in 20 Chinese dialects. The curve-fitting function is shown in the figure. . .	95
2.13	Proportions of words in English, Dutch, and German in different lengths among the first 20,000 frequent words. . . . .	101
2.14	Proportions of Chinese words in different lengths among the first 5000 frequent words. . . . .	102
3.1	The number of immigrants to and emigrants from Guangzhou during the year of 1931-1990 (adapted from Guangzhou Guangzhou Renkou Zhi, 1995). . . . .	112
3.2	The distribution of states of target morphemes in speakers of three young age groups. . . . .	122
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. . . . .	123
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. . . . .	124
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. . . . .	128

3.6	The percentages of $V$ in two sessions of word-list reading and the consistency between the two sessions. . . . .	130
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. . . . .	133
3.8	Percentages of $/n-/$ in 17 morphemes from production and questionnaire data. . . . .	134
4.1	Three hypothesized strategies for acquiring the basis of communicative behavior, reproduced from Hurford (1989). . . . .	169
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. . . . .	205
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. . . . .	206
5.3	Three phase transitions in the emergence of a compositional language in a population, reproduced from Kirby (2000), Figure 18.3. . . . .	208
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.) . . . .	212
5.5	Examples of S-curve from two empirical studies. (a) development of the periphrastic <i>do</i> , reproduced from Ogura (1993); (b) diffusion in the apparent time of a sound change in Shanghai, produced using data from Shen (1997). . . . .	216
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), $m_0=m=2$ ); all with the same size ( $N=10$ ) and same average degree( $\langle k \rangle=4$ ). . . .	223
5.7	The degree distribution of four typical networks ( $N=1,000$ , $\langle k \rangle=20$ ). . . .	224

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) $D=50\%$ . . . . .	225
5.9	Diffusion in an age-structured small-world network ( $N = 500$ , $\langle k \rangle = 10$ , $p = 0.01$ , each time step 4000 interactions). (a) adults change ( $\alpha = 0.5$ ) (b) adults rarely change ( $\alpha = 0.001$ ). . . . .	227
5.10	Diffusion dynamics in four types of networks in 10 runs (network size $N = 500$ , average degree $\langle k \rangle = 20$ , functional bias $\beta = 20$ , number of innovators $I = 1$ ). . . . .	229
5.11	Diffusion dynamics in four types of networks in 10 runs. ( $N = 500$ , $\langle k \rangle = 20$ , $\beta = 10$ , $I = 10$ ) . . . . .	230
5.12	Diffusion dynamics in four types of networks in 10 runs. ( $N = 500$ , $\langle k \rangle = 20$ , $\beta = 2$ , $I = 100$ ). . . . .	230
5.13	The probability of successful diffusion under different functional biases, and the average diffusion time in four types of networks over 100 runs. ( $N = 500$ , $\langle k \rangle = 20$ , $I = 1$ ). . . . .	231
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. ( $N = 500$ , $\langle k \rangle = 20$ , $\beta = 4$ ). . . . .	232
5.15	The relation between community size and the rate of change in four types of networks. . . . .	233
5.16	The diffusion dynamics in a population with all probabilistic learners in two networks. ( $N = 500$ , $\langle k \rangle = 20$ , $\beta = 2$ ). . . . .	236
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. . . . .	237
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. ( $N = 500$ , $\langle k \rangle = 20$ , $I = 1$ ) . . . . .	237
6.1	Spectrograms illustrating syllables /ni/ (left) and /li/ (right) of an adult female voice. . . . .	246

# List of Tables

1.1	The oral and nasal vowel patterns in Hauteville before the shift. Taken from Martinet (1952). . . . .	30
1.2	The oral and nasal vowel patterns in Hauteville after the shift. Taken from Martinet (1952). . . . .	30
1.3	The oral vowel patterns at the intermediate stage during the change of the Hauteville vowel system. . . . .	30
1.4	The vowel patterns at a later intermediate stage. . . . .	31
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). . . . .	50
2.1	The degrees of homophony in 20 modern dialects and two historical Chinese (sorted according to the number of occurring syllables <i>Syl.</i> . . . .	68
2.2	Information about the three lexicons from CELEX. . . . .	71
2.3	Segment inventories, number of canonical forms, occurring syllable types, estimated and actual CV combinations, and CV exploitation rates in three Germanic languages. . . . .	75
2.4	A statistical estimate of the percentages of disyllabic words in several books by Jiang (1991), reproduced from Zhang (1997). . . . .	84
2.5	Estimated percentage of disyllabic words across three different historical periods, reproduced from Sproat (2002). . . . .	85
2.6	Comparison of degrees of homophony and degrees of disyllabification in 20 Chinese modern dialects . . . . .	94
2.7	A summary of the experiments in Kelly and Ragade (2000). Homophone pairs are differentiated in usage in terms of grammatical classes. . . . .	96
2.8	Statistical results of the differentiation of homophone pairs in usage in terms of grammatical class in English. . . . .	97
2.9	Statistical results of the differentiation of homophone pairs in usage in terms of grammatical class in Dutch. . . . .	97

3.1	Some evidence of the change /n-/→/l-/ in loan words or translations in Hong Kong . . . . .	108
3.2	Distribution of speakers in age groups and genders . . . . .	116
3.3	Seven types of speakers categorized by different distributions of morphemes in the two phonological categories. . . . .	137
3.4	Comparison of the pronunciations of 50 morphemes of two identical twin Cantonese speakers (HK1 and HK2). . . . .	138
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. . . . .	140
3.6	Comparison of the percentages of /n-/ of speakers from local and non-local families . . . . .	144
3.7	Comparison of the percentages of /n-/ of speakers with high and low level of education. . . . .	148
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. . . . .	175
5.1	The initial and final utterance-meaning mappings, $\Psi$ , of two agents. . . . .	201
5.2	An example of the speaking and listening matrices of one agent in the self-organization model. . . . .	202
5.3	An example of a stable state: All agents have the same speaking and listening matrices. . . . .	203
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. . . . .	203
5.5	Comparison of the four types of networks in probability of successful diffusion and rate of diffusion. . . . .	232