Computer Simulation of Grammatical Change

Makoto Nakamura
School of Information Science
Japan Advanced Institute of Science and Technology

Table of Contents

Introduction

- A computer simulation is a program that attempts to simulate an abstract model of a particular system.
- Computer simulations can be considered as virtual experimental laboratories that allow researchers to run realistic, impossible, and counterfactual experiments.
  - Flight simulators to train pilots
  - Weather forecasting
  - Robot simulation for the design of robots and robot control algorithms
  - Grammatical change(?)

3

Computer Simulation of the Evolution of Language

- The study of language origins and evolution has widely been studied for more than a decade.
- Computer simulations can help us to overcome some very problematic aspects of the study of language evolution.
- A simulation is the implementation of a theory in a computer.

e.g. syntax, symbol grounding, grammaticalization, etc

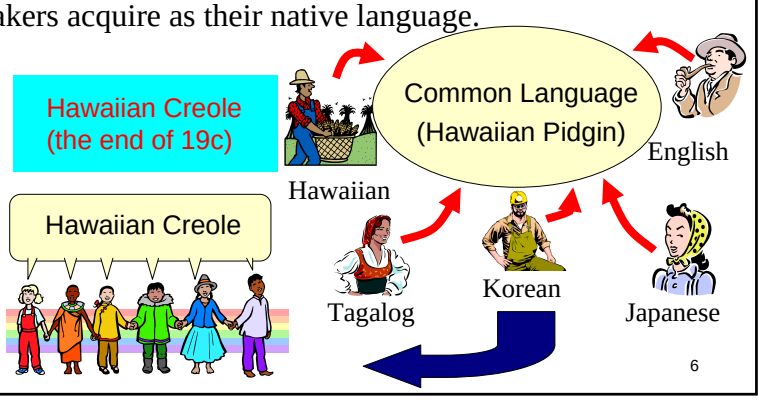
4

Table of Contents

5

Pidgins and Creoles

- Pidgins are simplified tentative languages spoken in multi-lingual communities.
- Creoles are full-fledged new languages which children of the pidgin speakers acquire as their native language.



6

Examples of utterance (Bickerton, 1990)

- Hawaiian Pidgin:

  Ifu laik meiki, mo beta *make* time, mani no kaen *hapai*.

  If like make, more better die time, money no can carry.

  If you want to build (a temple), you should do it just before you die -- you can't take it with you!

- Hawaiian Creole:

  They wen go up there early in the morning *e* go plant.

  They went up there early in the morning in order to plant (crops).

7

Purpose

- To investigate a process of the emergence of creoles
  - Population Dynamics (Nowak et al., 2001)
  - Multi-Agent Models (Briscoe, 2001; Kirby, 2001)
- To obtain some conditions for creolization
  - Relationship between creoles and other languages
  - Learning environments for infants

8

Language and Population Dynamics

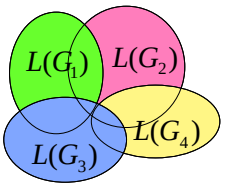
[Nowak 2001]

- Universal Grammar (Chomsky,81)
  - A finite set of grammars $\{G_1,\cdots,G_n\}$
  - A plausible grammar for each person is one of them
  - The proportion of population $\{x_1,\cdots,x_n\}$
- Similarity among Languages

  The similarity matrix $S=\{s_{ij}\}$: the probability that a sentence of G_i is accepted by G_j .
- Transitivity of population among grammars

  The transition matrix $Q=\{q_{ij}\}$: the probability that children of G_i speakers fail to acquire G_i but come to speak G_j

Venn diagram of shared sentences



9

Modified Language Dynamics Equation

(Nakamura et al., 2004)

$$\frac{dx_i}{dt}=\sum_{j=1}^{n}x_j\overline{Q}_{ji}(t)-x_i \quad (i=1,...,n),$$

where

x_i : the population proportion of G_i, where $\sum_j x_j=1$,

$\overline{Q}(t)=\{\overline{Q}_{ij}(t)\}$: the transition probability between grammars that a child of G_i speakers come to acquire G_j .

Modification from the previous work (Nowak, 2001)

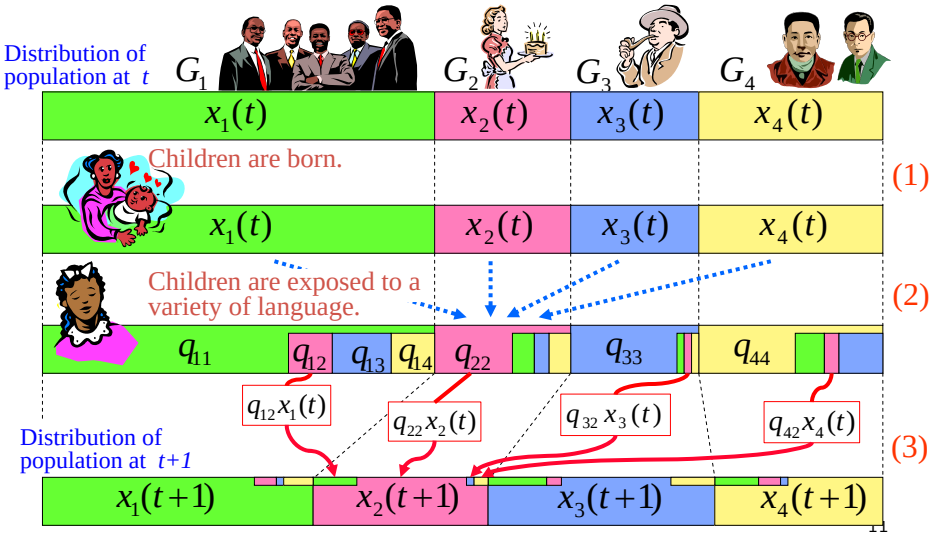
(1) to exclude a fitness term concerned with the biological evolution

(2) to expand its domain into cultural transmission

(3) to change the transition matrix Q to a function

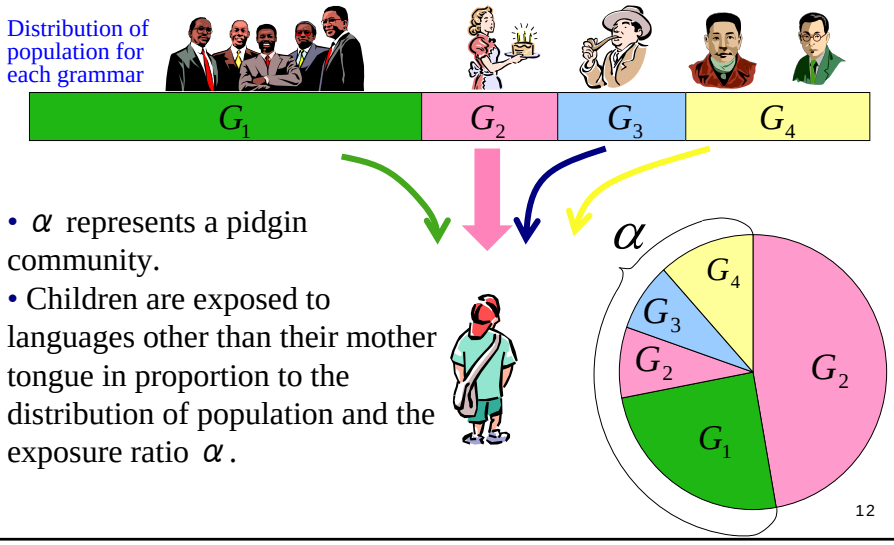
10

Illustration of Our Model



11

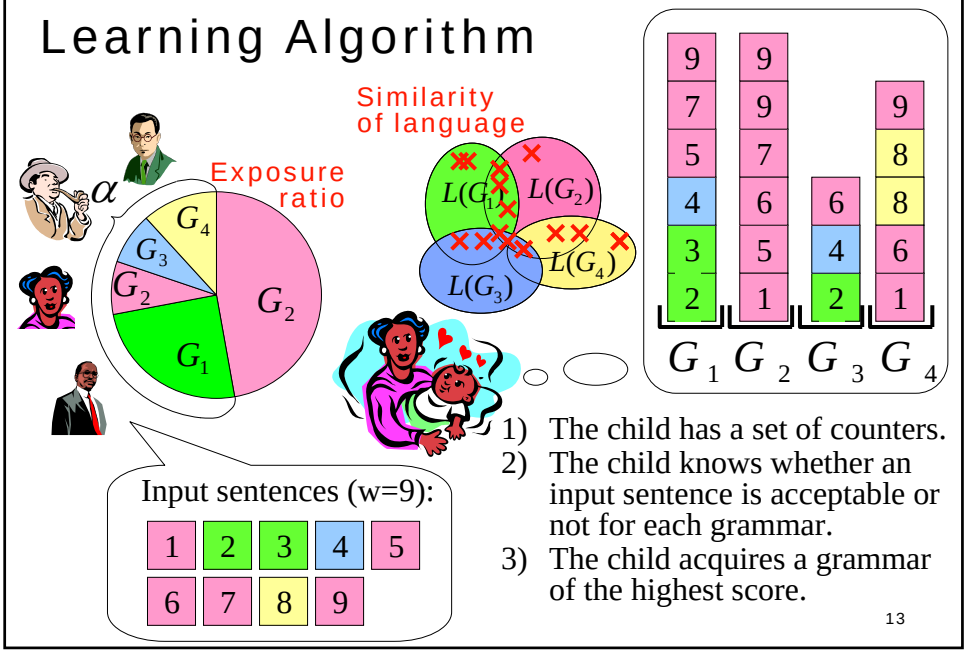
Exposure Ratio α



- α represents a pidgin community.
- Children are exposed to languages other than their mother tongue in proportion to the distribution of population and the exposure ratio α.

12

Learning Algorithm



1) The child has a set of counters.
2) The child knows whether an input sentence is acceptable or not for each grammar.
3) The child acquires a grammar of the highest score.

13

Definition of Creoles in LDE

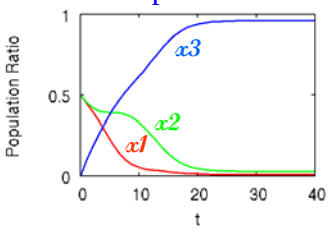
- Creoles are also included in the set of predefined languages UG:

$$\{G_1,\cdots,G_c,\cdots,G_n\}$$

- A creole is defined as a grammar G_c in LDE as follows.

$$\begin{cases} x_c(0)=0 \\ x_c(t)\geq\theta_c\ (=0.9) \end{cases}$$

Example of Creole



14

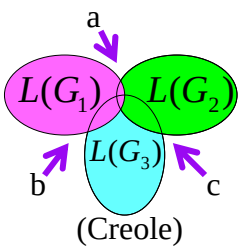
Experiments

- 3 grammars
  - G1, G2: preexisting grammars
  - G3: creole grammar

$$\begin{cases} x_1(0) = x_2(0) = 0.5 \\ x_3(0) = 0 \end{cases}$$

- Similarity among languages

$$S = \begin{bmatrix} 1 & a & b \\ a & 1 & c \\ b & c & 1 \end{bmatrix}$$

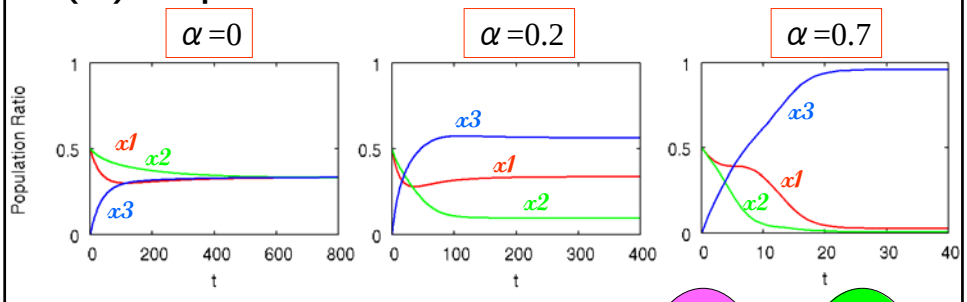


(1) The Exposure Ratio α and creolization

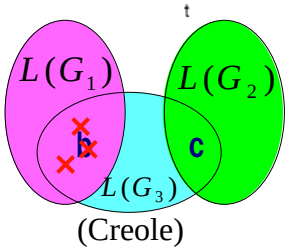
(2) Similarity among languages and creolization

15

(1)Exposure Ratio α and Creole

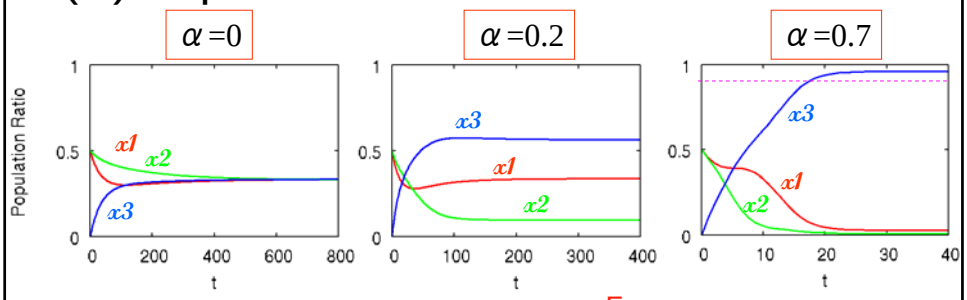


- # of input sentences: w=3
- (a,b,c)=(0,0.3,0.2)
- G1 and G3 are likely to transfer each other.
- α makes population transition active.

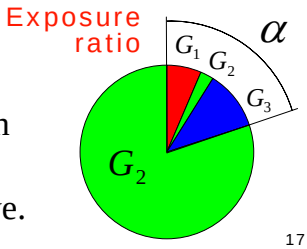


16

(1)Exposure Ratio α and Creole



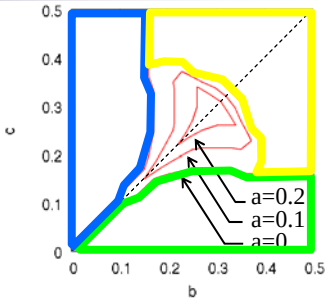
- # of input sentences: w=3
- (a,b,c)=(0,0.3,0.2)
- G1 and G3 are likely to transfer each other.
- α makes population transition active.

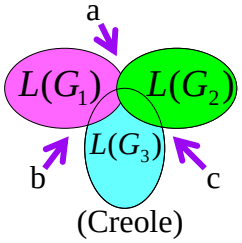


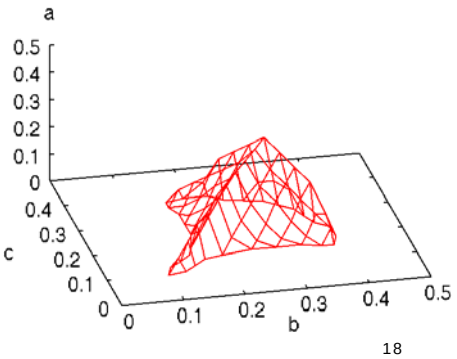
17

(2) Conditions for Creolization (1/6)

- Conditions for creolization on similarity among languages
- #input sentences: w=3
- Exposure Ratio: α=0.7
- Symmetric along with b=c



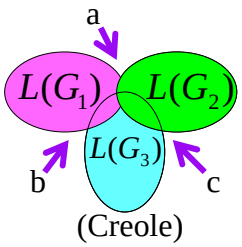


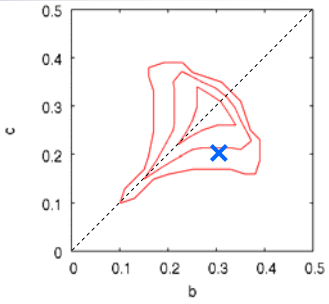


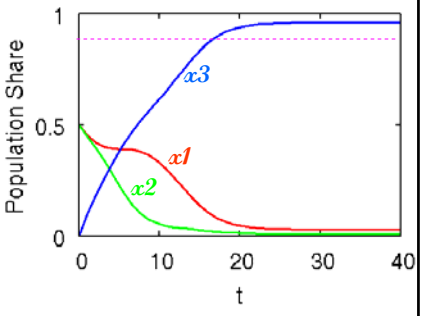
18

(2) Conditions for Creolization (2/6)

- w=3, α=0.7
- (a,b,c)=(0,0.3,0.2)
- Active transition between G1 and G3



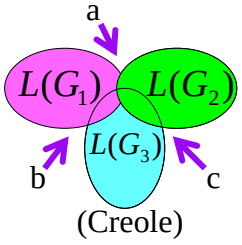


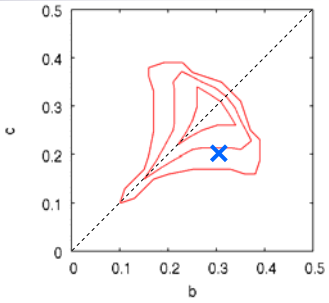


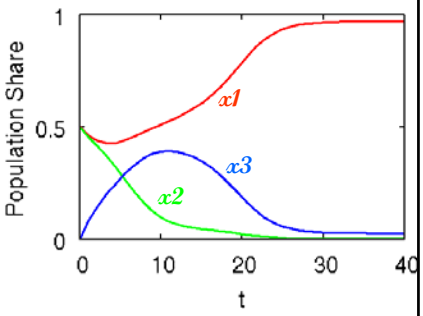
19

(2) Conditions for Creolization (3/6)

- w=3, α=0.7
- (a,b,c)=(0.1,0.3,0.2)
- Active transition between G1 and G2



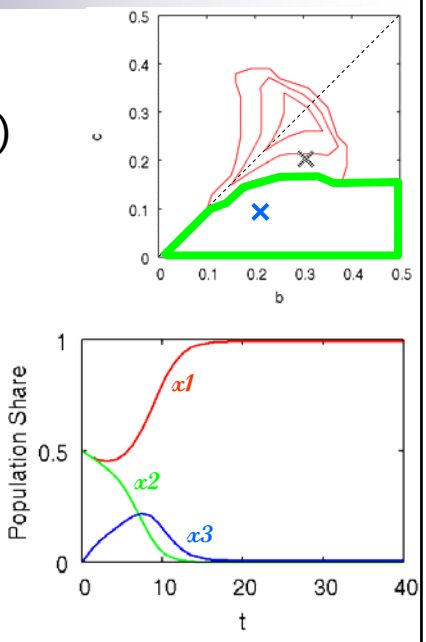




20

(2) Conditions for Creolization (4/6)

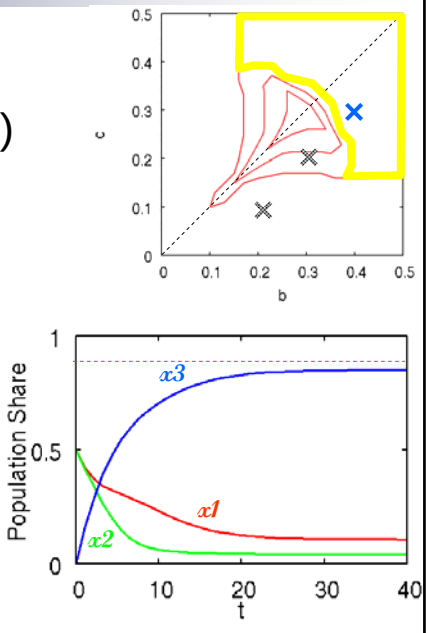
- w=3, α=0.7
- (a,b,c)=(0,0.2,0.1)
- Inactive transition because of low similarity
- G1 becomes dominant since the initial population.



21

(2) Conditions for Creolization (5/6)

- w=3, α=0.7
- (a,b,c)=(0,04,0.3)
- G1 and G3 coexist because of high similarity.
- G3 is the most populous language, but it is not a dominant creole. $(\theta_c = 0.9)$

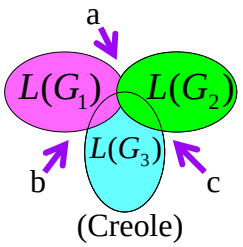


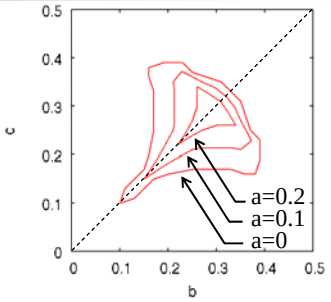
22

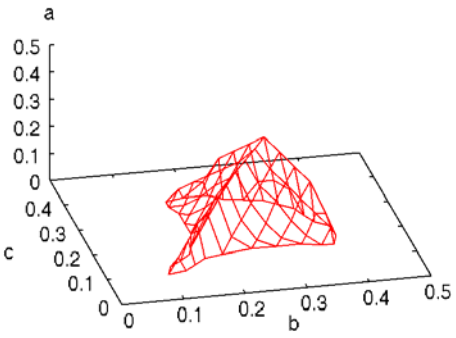
(2) Conditions for Creolization (6/6)

Pre-existing languages are:

- Dissimilar with each other
- equally similar with the creole
- Not too similar with the creole.







23

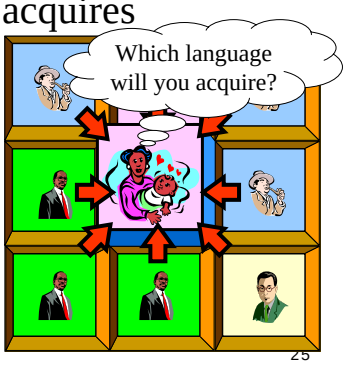
Summary of Creole Simulation

- Creole viewed from population dynamics
  - Since languages are extremely abstracted, similarity can be parameterized.
  - Parameters: Learning environment, Similarity among languages
- Conditions for creolization
  - The exposure ratio: the greater, the easier
  - Pre-existing languages: dissimilar with each other
  - Creoles are not too similar to pre-existing languages
  - Creoles are equally similar to pre-existing languages

24

Introducing a Spatial Structure to Language Dynamics

- How do language learners organize a creole community?
- Each individual independently acquires a language, applying the transition probability.
- Deterministic vs. stochastic

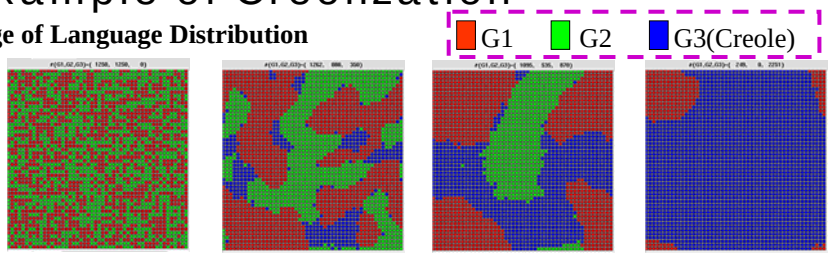


25

Example of Creolization

$$S = \begin{pmatrix} 1 & 0 & 0.3 \\ 0 & 1 & 0.4 \\ 0.3 & 0.4 & 1 \end{pmatrix},\ \alpha = 0.7,\ w = 10$$

Change of Language Distribution



Generation: 0 50 500 1552

- Agents surrounded by both *G*1 and *G*2 neighbors are likely to acquire the creole.
- The creole is the most efficient for accepting input utterances from both languages.
- Some colonies of creole are organized at the early stage.
- Colonies are unlikely to vanish.

26

Table of Contents

- Introduction
- Computer simulation of the emergence of creoles
- Computer simulation of the change in inflection
  - Word order and Inflection
  - Acquisition of a rational grammar
  - Experiments with two corpora
- Conclusion

27

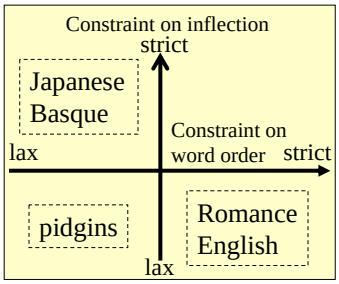
Computer Simulation of the change in Inflection

- Inflection is the way language handles grammatical relations and relational categories such as tense, mood, aspect, person, number, gender, case and so on.
- Under multi language communication, a common language tends to lose [Sebba, 1997]
  - Case marker
  - Agreement of number, gender
  - Tense

28

Word order and Inflection

- The language with rich inflectional system
  ⇒ more flexible in word order
- The language with poor inflectional system
  ⇒ strict word order



English | Word order is fixed: S-V-O

A dog bites a man. | S: a dog O: a man

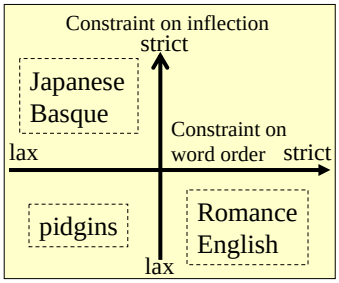
Japanese | Basic word order is S-O-V

Inu-ga otoko-wo kamu
(dog) (man) (bite)

29

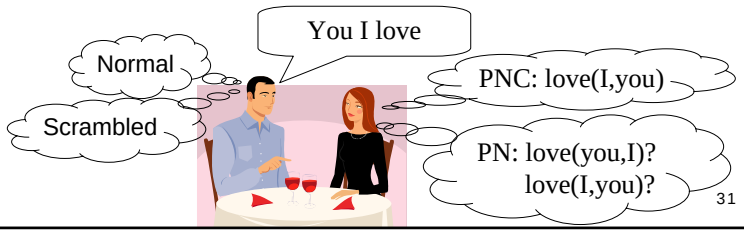
Purpose and Method

- Purpose
  - To acquire grammars adapting for environment
  - To investigate efficiency of inflection and word order
- Method
  1. To make scrambled English sentences from corpora
  2. To construct grammars w/ and w/o the process of inflection
  3. To measure efficiency of grammars using a utility function
  4. To examine the change in inflection with a population dynamics



Strategy

- Speakers' strategy: to avoid ambiguous expression
  - Normal: Normal English sentences (from corpora)
  - Scrambled: Scrambled sentences
- Hearers' strategy: to avoid paying attention to inflection
  - PNC: consideration of person, number, case(pronoun)
  - PN : neglect of inflection for case



31

Corpora

- Wall Street Journal: WSJ
  - Includes many long, complicated sentences

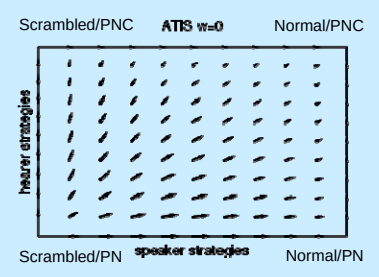
Zenith Data Systems Corp., a subsidiary of Zenith Electronics Corp., received a $534 million Navy contract for software and services of microcomputers over an 84-month period.

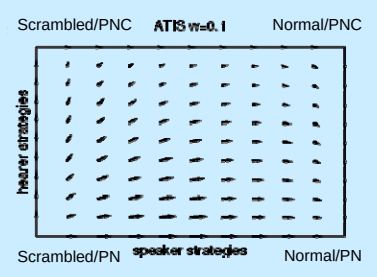
- Air Travel Information System: ATIS
  - Includes many simple sentences

I need to have dinner served.
Show me the flights form Tampa to Baltimore.

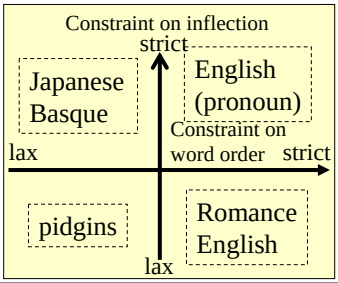
32

Experimental Results (ATIS)

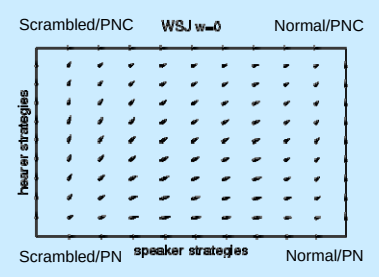


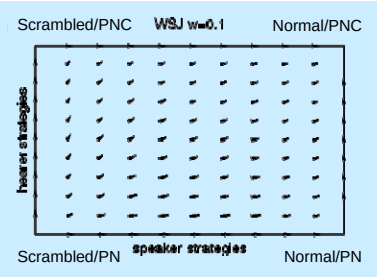


- Phase diagrams represent the grammatical change, corresponding to the distribution map of languages.
- ATIS corpus includes short and simple sentences.
- Hearers as pidgin speakers prefer the grammar in which they do not take the case feature into consideration.

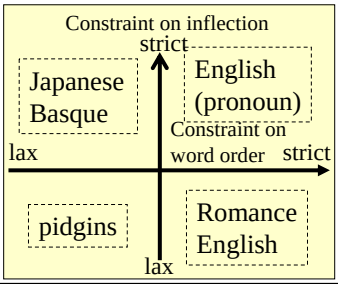


Experimental Results (WSJ)





- WSJ corpus includes long and complicated sentences.
- Case features help them to avoid ambiguous expressions.
- Hearers prefer the grammar in which they take the case feature into consideration, even if it is a burden on the hearers.



Summary of the Change in Inflection

- Rational grammars change depending on environment.
- As long as speakers speak simple sentences, the rational grammar is likely to be simplified.
- Need to reconsider both speakers' and hearers' strategies
- Use Corpora other than English

35

Conclusion

- Simulations of grammatical change
  - Emergence of creoles: conditions for creolization in terms of concerning languages, exposure ratio
  - Learning of inflection: The utility of inflection changes depending on the situation.
- Computer simulations can show how interactions among individuals affect languages spoken throughout the community.
- Thus, hypothesizing about the abilities of individuals or about the learning environment, we can consider what causes a grammar to change.

36