THE SOCIO-GENETIC SOLUTION: A NEW LOOK AT LANGUAGE GENESIS THROUGH SWARM MODELING

TERESA SATTERFIELD, University of Michigan, Ann Arbor, Michigan*

ABSTRACT

Creole languages have long been a point of contention in linguistic circles. Broadly defined, creoles are grammars resulting from an amalgamation of two or more languages, such as when speakers of differing mother tongues find a need for rudimentary communication during economic or social transactions. Creolization occurs if the "invented" system becomes the native language of the speech community. There are several hypotheses for how linguistic properties and social contact each came to bear on the formation of creole languages in past centuries; however until recently no reliable method for testing these complex interactions existed. By building agent-based 'societies,' it becomes possible to examine premises of linguistic theory and to reconstruct historical contexts, with an eye to isolating patterns and factors that are most relevant to the acquisition and transmission of languages. Implementing SWARM 2.1.1, the current model consists of a multi-agent population drawn from historical records of Surinamese sugar cane plantations (Arends 1995). Each agent is endowed with a demographic profile and linguistic parameters. Linguistic features include a set of genetic constraints stipulating the environmental conditions required for successful analysis and acquisition of any language. Three experiments using the Swarm model are described. The results provide viable motivation for advancing a "socio-genetic" solution for the emergence of prototypical creole languages.

INTRODUCTION

This paper simulates a language formation process known as creolization, using a computer program that tracks 'speaker-agents' who enter and emerge from the learning environment. Throughout history, creole languages have arisen only in restricted social contexts, during trade transactions, or under circumstances of upheaval such as slavery. In such cases, speakers of diverse languages must bridge a communication gap when there is little opportunity or will to learn others' language(s). The situation results in the use of a simplified communication system known as a *pidgin*.[1] In time, this newly formed code may nativize (become the native tongue in a speech community), at which point we say that the pidgin has *creolized.* The resultant language is called a *creole*.

---

* I would like to thank Wei Chen and Rick Riolo at the Center for Study of Complex Systems (CSCS), University of Michigan, for valuable assistance on this project. *Correspondence to*: Teresa Satterfield, Department of Linguistics, University of Michigan, 1076 Frieze Bldg., Ann Arbor, Michigan, 48109-1255; e-mail: tsatter@umich.edu

[1] I use the term *'pidgin'* not in the detrimental sense attributed to it in many standard dictionaries, but in the technical linguistic sense, as a type of "auxiliary" language or strategy that is called upon by speakers in addition to their own native languages to enable communication to take place.

Where the controversy lies, and where motivation for agent-based modeling begins, is that unlike pidgins that confine themselves to a few simple structures, the creole language regularizes and expands into complex grammatical forms. These syntactic innovations are akin to “normally” transmitted native languages, yet unlike the ambient mother tongues or pidgin elements which function as the creole’s primary linguistic input. Moreover, some argue that creoles worldwide pattern uniformly, such that among themselves they share far-reaching structural similarities, which make them distinguishable from “conventional” languages.

Since pidgins never constitute a mother tongue for any speaker, how is a full and nativized creole language acquired given ‘difficult’ environments? Most researchers agree that both ‘nature’ (biological capacity for humans to acquire language) and ‘nurture’ (external features shaping human development) play a part; however, this consensus is not integrated into research programs. The majority of creolization investigations make strong arguments for either the genetic component of language acquisition or the social aspect. Below, I sketch the dichotomy that holds between these frameworks.

The *Language Bioprogram Hypothesis (LBH) (*Bickerton 1984a, 1992), at its most extreme, depicts creolization as a “bet-hedging” reflex related solely to biological (first) language acquisition processes. Given a “chaotic” linguistic environment, innate “blueprints” yield the most generic structures possible for natural language. To the extent that children generate creoles as mother tongues only in language contact situations, there is a basis for the related claim that creole languages emerge abruptly within one generation of speakers.

A contrasting viewpoint can be labeled the *Social Context Hypothesis (SCH)* (Thomason & Kaufman 1988). Framed in the premise that creoles are indicators of human interaction in unique and varying communicative contexts, they argue against a bioprogram, stating that children possess only enough *a priori* linguistic knowledge for environmental and social conditions to impact acquisition in important and interesting ways. The SCH questions “universal” creole similarities, using comparative historical data as evidence that creoles contain broad structural differences. Finally, the SCH points to nativization as a gradual process requiring generations of participation from the community’s adult speakers.

Researchers tacitly assume that mechanisms formulated in the LBH and the SCH cannot function bilaterally. The proposed model aims for a fuller understanding of the creolization process. Object-oriented programming provides an optimal system for simulating complex and dynamical aspects, such as those in language formation. Following Epstein & Axtell (1996), the proposed setting for language genesis functions as an *artificial society,* “cultivating” creolized syntactic structures *in silico.* The primary goal is to discover the role of certain innate and socially-based mechanisms in generating prototypical creole grammars.

Overview

This model simulates real world multilingual plantations which supplied optimal conditions for growth of pidgins and continued development of existing contact languages. While plantation-based slave labor tended to lead to creole genesis, it is not known if these languages necessarily originated from pidgin stages. Nor is creolized grammar an inevitable outcome of the plantation system. We will focus on one known context which spawned and perpetuated creole language.

Ethnolinguistic and demographic parameters utilized derive from statistics on Sranan Tongo (Surinam Tongue) (Arends 1995, Migge 2000). Sranan is the native language of 200,000 inhabitants of modern Surinam, and functions as *lingua franca* for the majority population. Sranan's earliest histories document mid-17th century English settlers who set up small farms. A Dutch invasion drove out the English by 1680. Until 1690, the colony experienced accelerated growth as Dutch planters expanded farms into large-scale sugar plantations and imported increasing numbers of African slaves. By 1750, every planter-master owned 45 to 60 slaves. Plantation numbers were routinely decimated from slaves' lowered life expectancy, low birth rates, and escape. Records speculate that male slaves outnumbered female counterparts 2 to 1. Children were 15% of the total slave cohort; approximately 40% of these infants died before age five. In short, the population was sustained through the constant influx of new slave labor rather than from natural growth.

Plantations functioned as strict hierarchical organizations. In Surinam, a large social distance existed between different groups on the plantation, causing considerable reduction in social networks. Social boundaries included: European-African; adult-child; elite slaves, including overseers and house slaves, versus field hands, and to a lesser extent, skilled African laborers. Ethnolinguistically, imported populations contributing to the formation of Sranan link to language families on the western African coast. Within 70 years, the population was linguistically homogeneous, with three individual languages as most prominent (Arends 1995).

Implementation of Demographic and Linguistic Data

Conforming closely to the previous statistics, each artificial society inhabitant is assigned a demographic profile designating age, race, health, death, cultural identity, and social status. Planters have the highest social standing. Slaves receive a status index based on their function in the plantation. Overseers have high status among slaves, field hands and infirm slaves receive the lowest index. Upon reaching age 12, the child's status is assigned through random 'inheritance' of one of his parents' indices.[2] The status index is essential, as it drives movement and subsequent language transmission, following the SCH.[3] Slaves' language variables are randomly assigned tags constituting different *A*(African)-*language* families. The planter language variable is one *E*(European-based)-*language*.

Per LBH specifications, individuals have innate linguistic capacities, characterized by a lexicon, represented as a list of values corresponding to each native and non-native word stored by the agent; a storage unit for accumulating new vocabulary items; lists of derivational morphemes (building blocks of word formation) and grammatical morphemes (tense and plurality affixes on lexical items, etc.). A morphology storage unit accumulates morphemes as a native language. Sentence structure, or syntactic constraints, are represented as links mapping words and morphology in various ratios depending on language specified. A-languages are mapped three morphemes to one word, E-languages have one morpheme per word.

Agents process linguistic input via bounded cognitive resources (*e.g.*, memory and time). Following the LBH, children and adults possess different computational limitations for acquisition and storage of lexical and grammatical information.

[2] Offspring can only receive planter status if both parents are planters; if one parent is a planter, child receives slave index.

[3] Given sociolinguistic studies (Gumperz 1971), socio-economic status appears to play a role in language transmission.

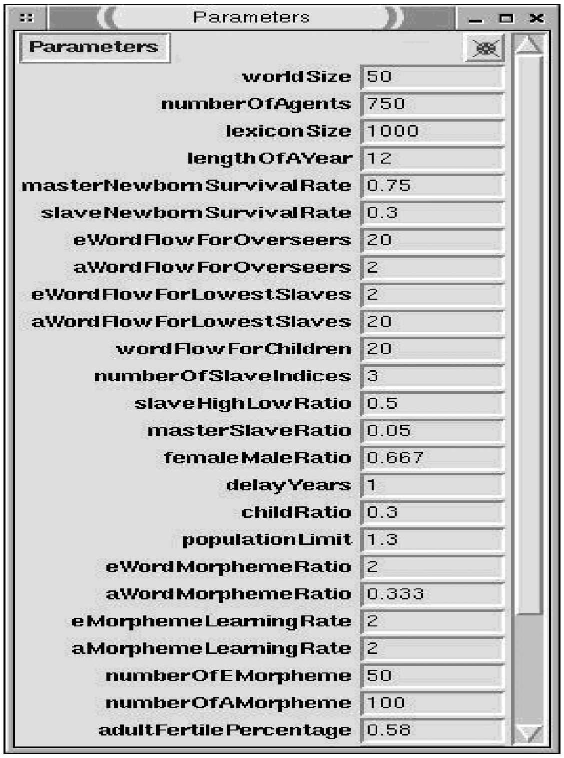


FIGURE 1 Model Parameters

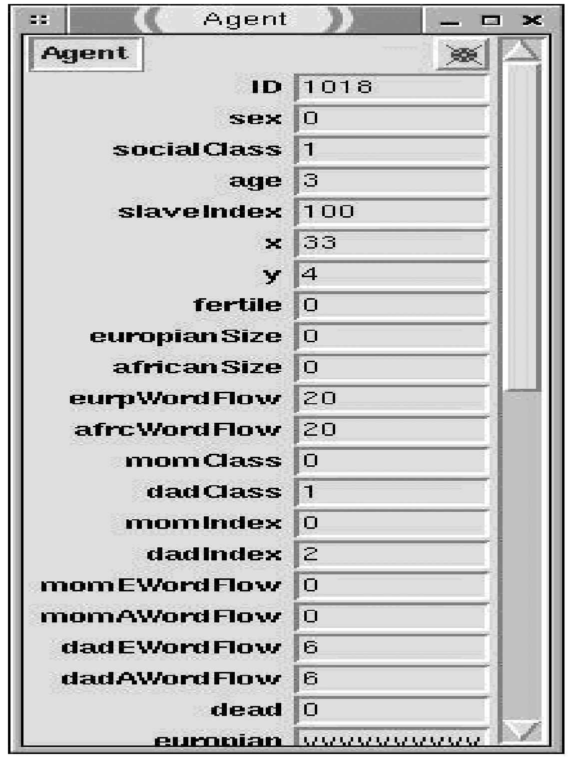


FIGURE 2 Agent Profile

DESCRIPTION OF THE SWARM MODEL

SWARM experiments generally involve three components: agents, environment space, and rules. Basic creolized systems are hypothesized to emerge as by-products of rule-governed interaction between agents operating in the established social environment. Distributions of innate linguistic knowledge, lexicons, and demographics are entered to initialize the population of agents. The Temporary Memory Buffer prohibits adult learners from acquiring an unchecked amount of vocabulary, on analogy with the restricted computational resources of adult-learners. Child-agents enter the environment with no linguistic affiliation, but possess inherent capacities to store and generate any language. Children acquire both words and grammatical morphemes at higher rates than adults, following the LBH. Steps performed by the model are listed in Table 1.

1. Initialize **adult** (>15.0 years) Agent's distribution of language features and demographic features.

2. **Language Contact.** Initialize Random Walk. (*1* chronological year = *12* cycles) Select unoccupied space within *X* Von Neumann neighbor-agents.

3. **Language Learning.** Talk to *X* neighbors. If input language = *[+African]*, form probability of encountering lexical and morphological information based on population ratio of African languages: Language *A*, Language *B*, Language *C*, Other African Languages.

   A. Adult:
   If neighbor has *[*> status index], look at *X* words in his lexicon. If you do not find those words in your lexicon, select those *X* items and copy into your Temporary Memory Buffer. Talk to *X* neighbors during this step. Consult Temporary Memory Buffer. Add new vocabulary item(s) to your lexicon only if: you encounter *>2.0* instances of those new words.

   B. Child:
   If neighbor has [> status index], or if he is a child *>5.0* years, look at *X* words in his lexicon and *Y* morphemes in the morphology listing. If you do not find those words and morphemes in your lexicon, select items and copy them into your respective lexical and morphological storage units. Talk to *X* neighbors during this step. Add new vocabulary item(s) and morphemes to storage areas immediately when you encounter *1.0* instance of new word *W* and new morpheme *M*.

4. Language learning for current cycle completed.

TABLE 1 Language Learning Rules for Agents

SAMPLE RUNS AND RESULTS

Three experiments examine the roles of 'nature' and 'nurture' in creole genesis. For each test, the number of 'years' of contact is given, as well as the linguistic profile of the community and an average of individual language patterns. Population and age distributions were monitored. Primary experimental variables are child language acquisition mechanisms (LBH) and rates of social contact (SCH).

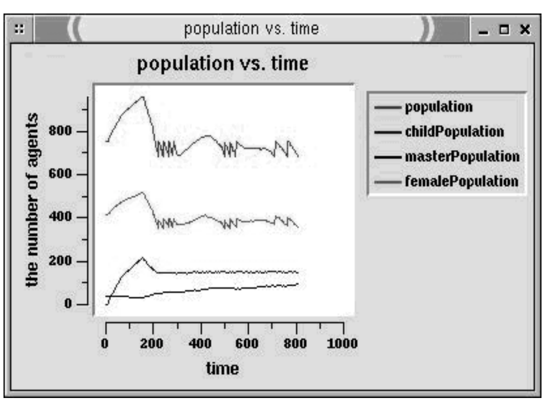


FIGURE 3 Population Distribution

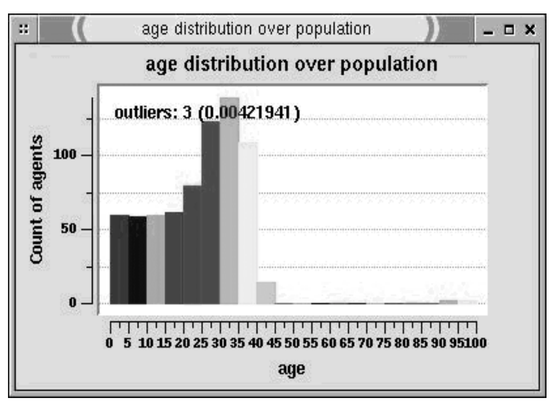


FIGURE 4 Age Distribution

Experiment 1

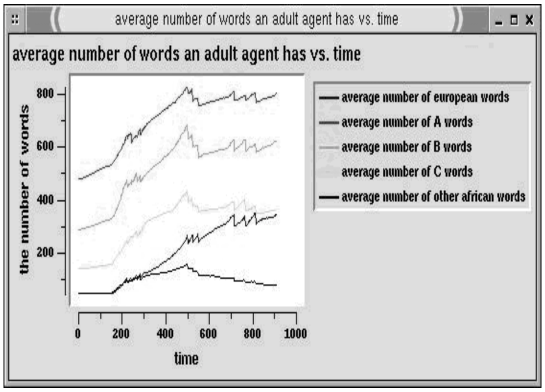


FIGURE 5 Adult Lexical Storage

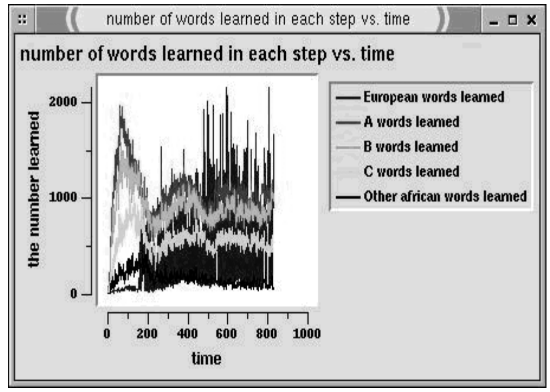


FIGURE 6 Lexical Items Acquired

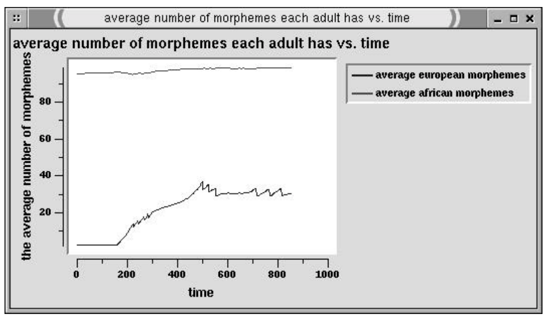
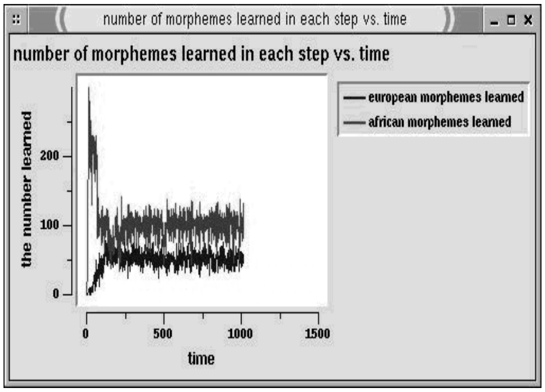


FIGURE 7 Adult Morphemes

All variables were kept constant in Experiment 1; subsequently, elements of the SCH (translated as high frequency and high quality social contact) and the LBH (translated as biologically endowed linguistic mechanisms) are assumed maximally operative. The end-result is indicative of a setting composed of a 15% minimum of children and unlimited contact (per appropriate social hierarchies) between neighbor-agents. Figures 5-7 display results.

Experiment 2

Experiment 2 manipulates the social contact variable to examine the SCH's impact. Instead of unlimited access to linguistic information from all surrounding neighbors of appropriate status, agents interact with one neighbor per cycle. Results may predict whether creole genesis is possible with minimal agent interaction. Graphs shown in Figures 8-9.



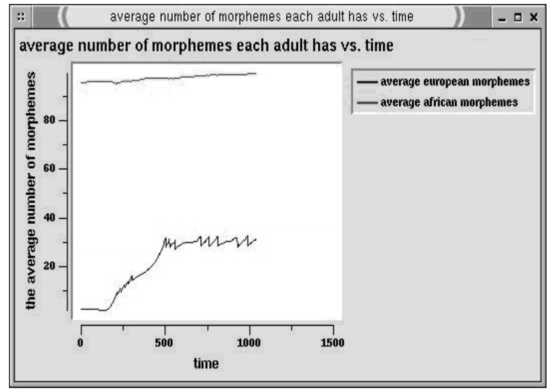
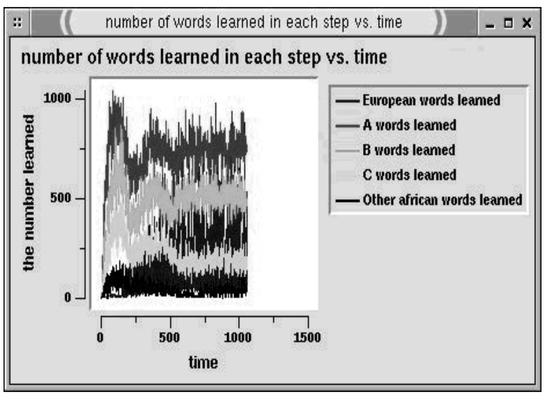


FIGURE 8 Morphemes Acquired no SCH **FIGURE 9** Adult Morphemes no SCH



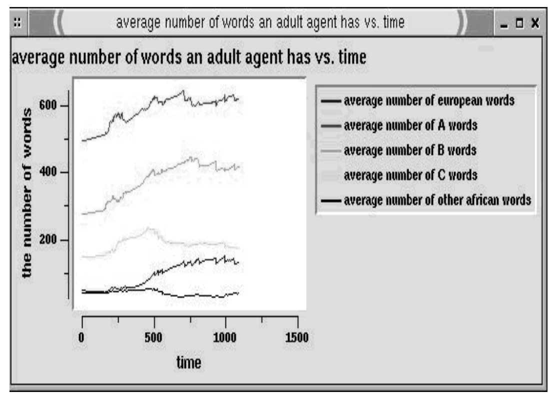
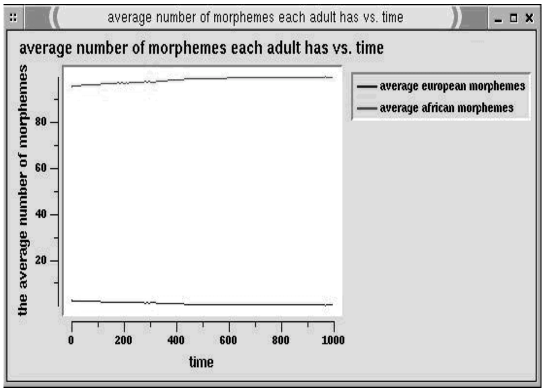


FIGURE 10 Lexical Items Acquired noSCH **FIGURE 11** Adult Lexical Storage noSCH

Experiment 3

Experiment 3 manipulates the child-driven language variable to investigate the LBH's role. In this scenario, no children are born into the population. Adults interact, with numbers maintained through periodical importation of slaves into the contact setting. Findings may predict whether creole genesis is possible in the absence of child learners. Results presented in Figures 12-13.



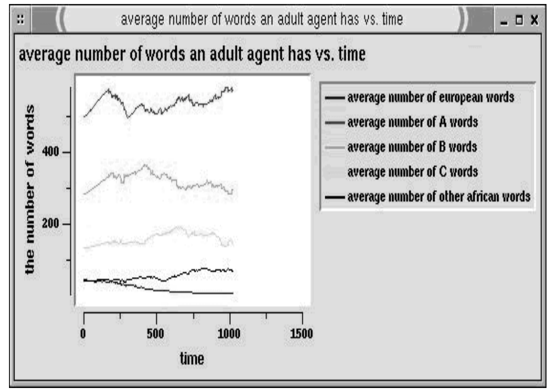


FIGURE 12 Adult Morphemes no LBH **FIGURE 13** Adult Lexical Storage noBCH

Discussion

Experiment 1 produces dynamic characteristics expected with language change. After three generations (approximately 70 years), Figure 5 shows how E-language words come to replace certain A-language varieties, as documented in real world plantation history. In line with historical slave language pidgins, Figure 6 exhibits initial trends of an abrupt overlapping of A-language acquired lexical items. The presence of children shortly after cycle 100 brings an "averaging" effect for the acquisition rate. An explosion of E-language vocabulary emerges in part from the fact that children require less frequency and memory to retain new vocabulary items, but, apparently, also from the influx of slaves from Africa, seen shortly after cycle 200 and again at cycle 450. Stored morphemes of the respective languages are distributed, as shown in Figure 7. Since morphology is acquired entirely as a function of first language, gradual storage of increased non-African morphology might arise if European planter children, post-age eight, have more contact with European adults than do other groups. E-language morphemes acquired in late childhood could still be easily transmitted to younger children. From the data, speakers in this scenario possess a diminishing A-language vocabulary, with substantive, but separate knowledge of both A- and E-language grammatical properties derived from childhood interactions.

Experiment 2 examines creolization under limited social interaction during 1000 cycles. Figure 8 shows that the initial short-lived rise of A-language morpheme acquisition, gives way to E-language morphology acquired at slightly lower rates than A-language forms. The learning pattern is relatively robust despite limited contact, however morphologies of each language do not overlap or combine as the learner acquires stable quantities of A- and E-language morphemes. Figure 9 displays results identical to Figure 7 for morphological storage in an adult speaker. Again, the agent has maximum knowledge of both A and E-language grammatical properties. In the face of restricted interaction, Figure 10 shows that lexical acquisition is 50% less than rates obtained in Experiment 1. A-language (precisely, African language A) forms are acquired at a prominent rate throughout the test, overlapping with E-language activity in the final cycles. Figure 11 tracks adult lexical inventory. Similar to Experiment 1, E-language words gradually replace certain A-language varieties, however A-languages comprise the majority of words stored . An adult's profile in this scenario corresponds to a small and receding A-language vocabulary (slightly more A-dominant than Experiment 1) with substantive, but separate knowledge of both A- and E-language grammatical properties.

Experiment 3 reflects a contact scenario in the absence of child language. Figure 12 shows the adult speaker with a full storage of A-language morphemes; however, contrasting with Figures 7 and 9 involving children, Figure 12 has no E-language morphology present. Figure 13 signals that A-language lexicons completely dominate E-language vocabulary; the latter falling to zero within 1000 cycles. Empirically, A-language words are stored at the same quantities found for Experiment 2 (See Figure 11), with a slight exception for the highest A-language. Speakers in Experiment 3, as adults, are the least influenced by contact, possessing a small A-language-only vocabulary, with full knowledge of solely A-language grammatical properties.

Statistically speaking, Experiment 1 represents the most fertile conditions for hypothetical creole genesis, since active child populations and relatively unrestricted contact trigger the highest incidents of change. Patterns in Experiment 2 did not differ extensively from Experiment 1. Limited contact appears to affect most dramatically the number of lexical items acquired and stored, while having minimal influence on the acquisition of morphology. Experiment 3 deviates greatly from the projected creolization outcome, as the lack of children equated to zero feature transmission across language groups.

CONCLUSION

In attempting to build from formulations of the LBH and SCH, this model offers experiments which duplicate the complex nature of creolization, simultaneously following social and (biological) linguistic development in a plantation setting. Preliminary evidence from the simulations demonstrate that genetic and social mechanisms must function bilaterally if a prototypical creole grammar is to potentially emerge. Overall, our findings suggest that innately specified capacities play a crucial role in organizing disparate input and in generating richer structures in *child* language. However, without significant amounts of social interaction, agents cannot attain native-speaker quantities of lexical inventories, nor do extensive nativized mixtures occur. Notably, striking similarities in morphological acquisition patterns of Experiments 1 and 2 favor the LBH claim that extralinguistic factors and environment do not greatly influence grammatical acquisition. Insofar as this is the case, such results may contradict the conservative "bet-hedging" premise of the LBH, since speakers redundantly retained native levels of morphology in both language groups. Finally, no experiment validated the related LBH claim that creole languages develop within a single generation of speakers. With roughly four generations of agents (1000 runs), few mixed language forms were materialized. Given these factors, we tentatively advance a socio-genetic solution for analyzing questions related to creole genesis.

Motivated by theoretical and practical concerns, modifications are presently being considered, with an eye to providing more explanatory models of language contact. For example, a syntactic formation involving fixed templates of *X* words to*Y* morpheme affixes is not suggested by current linguistic theory, although this strategy efficiently eliminates extensive searches for language family tags. Second, we posited rigid constraints on adult acquisition of non-native words and morphology; whereas the youngest children had few restrictions for storing any morphological or lexical item. It is clear that such stipulations, along with issues of linguistic input, must be coordinated with recent findings in first and second language acquisition theory. Lastly, demographic factors should be fine-tuned to better pinpoint those “critical mass” conditions necessary for triggering individual and population-wide creolization over time.

References

Arends, J., ed., 1995, “Demographic factors in the formation of Sranan,” pp. 233-285 in J. Arends, ed., *The early stages of creolization*, John Benjamins, Amsterdam.

Arends, J., P. Muysken, and N. Smith, eds., 1994, *Pidgins and Creoles: An Introduction,* John Benjamins, Amsterdam.

Bickerton, D., 1999, “How to Acquire Language without Positive Evidence: What Acquisitionists Can Learn from Creoles,” pp.49-74 in M. DeGraff, ed., *Language Creation and Language Change,* MIT Press, Cambridge, MA.

Bickerton, D., 1992, “The sociohistorical matrix of creolization,” *Journal of Pidgin and Creole Languages* 7:307-318.

Bickerton, D., 1988, “Creole languages and the bioprogram,” pp.267-284 in Newmeyer, F., ed., *Linguistics: The Cambridge Survey,* Cambridge University Press, Cambridge.

Bickerton, D., 1984a, “The language bioprogram hypothesis,” *Behavioral and Brain Sciences,* 7:212-218.

Bradshaw, J., ed., 1997, *Software Agents*, MIT Press, Cambridge, MA.

Chomsky, N., 1995, *The Minimalist Program*, MIT Press, Cambridge, MA.

Chomsky, N., 1986, *Knowledge of Language: Its Nature, Origin, and Use,* Praeger, New York.

Epstein, J., and R. Axtell, 1996, *Growing Artificial Societies,* MIT Press/Brookings Institute, Cambridge, MA.

Gumperz, J., 1971, *Language in Social Groups*, Stanford University Press, Stanford, California.

Holm, J., *Pidgins and Creoles, Volume 1: Theory and Structure,* Cambridge University Press, Cambridge.

Migge, B., 2000, “The Origin of the Syntax and Semantics of Property Items in the Surinamese Plantation Creole,” pp.201-234 in J. McWhorter, ed., *Language Change and Language Contact in Pidgins and Creoles,* John Benjamins, Amsterdam.

Rickford, J., and J. McWhorter, 1997, “Language Contact and Language Generation: Pidgins and Creoles,” pp. 238-256 in F. Coulmas, ed., *The Handbook of Sociolinguistics,* Blackwell, Oxford.

Satterfield, T., 1999, *Bilingual Selection of Syntactic Knowledge,* Kluwer, Dordrecht.

Sebba, M., 1998, *Contact Languages: Pidgins and Creoles,* St. Martin's Press, New York.

Siegel, J. , 1999, "Transfer constraints and substrate influence in Melanesian Pidgin," *Journal of Pidgin and Creole Languages,* 14(1):1-44.

Seigel, J., 1987, *Language Contact in a Plantation Environment,* Cambridge University Press, Cambridge.

Thomason, S., 1997, "A Typology of Contact Languages," pp. 71-88 in Spears, A. and D. Winford, eds., *The Structure and Status of Pidgins and Creoles,* John Benjamins, Amsterdam.

Thomason, S. and T. Kaufman, 1988, *Language Contact, Creolization, and Genetic Linguistics,* University of California Press, Berkeley, California.