Multi agent study for infuluence of immigration on regularization of irregular verbs

Hiroaki Suzuki (1310034)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 12, 2015

Keywords: irregular verb, multi agent, langeage change, s-curve, social linguistics.

The purpose of this study is to reproduce the preceding work (Lieberman 2007), by multi agent simulation. Lieberman showed that irregular verbs of English had tended to regularize over time.

Human language comprises various inflectional system, such as German and Latin (these language called inflectional language). Inflectional system enables speaker to communicate with another, and simplifies these communication. On the contrary, an agglutinative language such as Japanese also exists. These languages can not be classified completely. For example, language A has only inflectional feature, on the other hand language B has only agglutinative feature (Saito 2001). Tense in English is also one of the inflection. We can generate the past tense by attaching a suffix *ed* to verb stem. This rule can be applied to new English word without exception. For example, a word *google* means that *search by Google* and we use *google* as verb in our conversation naturally. Similarly we use past tense *googled*.

However, irregular verbs generate the past tense by irregular conjugation (sing - sang). In other words, we can not use *ed* rule to generate the past tense. Therefore, irregular conjugation considered to be an exception of rules. The origine of irregular verbs (currently we use) is *strong verbs*. Strong verbs had been used in Old English era (A.D 800) and they had some patterns of irregular conjugation. This feature (irregular conjugation) is transmitted to today's irregular verbs, and they also have a lot of conjugation patterns. That is, it is clear that irregular verbs have *No Rules*.

According to (Saito 2001), we need a lot of burden to learn the complex inflectional system (in this study, *complex* means that there exists a lot of conjugation patterns). For this reason, irregular verbs (strong verbs) regularization is not unnatural. But irregular verbs still remain in English. One of the reasons is that irregular verbs have *high frequency*. For example, we use *be, have, come,* many times in our daily conversation, and all of them are irregular. Actually, the proportion of irregular verbs in all English verbs is 3%, but in case of only conversation, the proportion of irregular verbs is 70% (Pinker 2000).

Lieberman showed that in his statistical study such irregular verbs had tended to regularize over time. Moreover, this phenomena began with low frequet irregular verbs. He also showed it's regularization speed function. This function consists of *time* and *frequency*, and can calculate irregular verbs distribution on any time point (past to future).

On the other hand, when we study such a historical language change, it is important for us to take *communication* into consideration because historical language change is a complex phenomenon of close interaction of features of the language itself and features (purpose) of communication. The purpose of communication is *to get the meaning across with a minimum physical effort on the part of both speaker and hearer*, and *convey the meaning as clearly as possible* (Jackendoff 2002).

In this study, we propose a multi agent model including the above two features, and simulate the historical language change in the preceding work. First, we explain the knowledge of agents and learning system concerning the artificial verbs made by (Pinker 1988). Agents have the frequency information of each verb, however at the initial state of the past tense form of each verb has not been fixed. Agents also have the learning system is implemented by Genetic Algorithm (GA), the genes of which representes the choices between the irregular conjugation and regular suffix *ed*. It means that agents have multiple choices for generating the past tense form, and which is different for each verb. Therefore, the learning process is to determine the best choice for generating the past tense form. Next, we explain the connection (relation) between agents. We define the agents connection as a complex network which consists of nodes and edges (same as the graph). In our model, a node means an agent and an edge means a connection. An agent can communicate with connected agents. After that, agents learn *what is the best past tence form in our community?* Therefore, the agents continue to learn through communication over time.

Features of the language in our model is *verb frequency*, and refers to the following sit-

uations. Agents may *keep inflection (generate irregular past tense)* or may *lose inflection (generate regular past tense)*. Losing inflection means that language change to agglutination. Dependent on the verb frequency used by agent communication, and keeping or losing infection is determined by agent's learning through communication.

We have performed two experiments by proposed model.

- The purpose of the first experiment is to reproduce the preceding work. To do this, we gave the following condition. When an agent utters in the regular inflection (ex. *singed* for sing), then we need to restrict the fitness of genes according to a *parameter*, which expresses the ratio between keeping irregularity or regularization. Actually, human may not want to utter longer words. But, if human can not communicate with each other, then human may use regular past tense form. We impose this condition only on those verbs with high frequency.
- The purpose of the second experiment is to validate the effect of the network size (agent size) in the regularization speed. Because we use complex networks, we must investigate the effect of network structure. The condition on the verbs with high frequency is the same as experiment1.

The condition of high frequent verbs is based on the purpose of human's communication (Jackendoff 2002). This condition controls the regularization speed of the high frequency verb.

As a result, experiment1 was not possible to reproduce the preceding work. Experiment2 showed that our model captured characteristics of complex networks. Furthermore, both experiments depend on the initial state of agent's knowledge. But our model contributes to the social linguistic simulation in terms of containing the language and communication features. Future direction of this study is betterment these issue, and make the parameter settings better.