How the SIR Model Can Be Adapted to Simulate the Spread of a Slang Word

Module:51701000 Method-oriented Approaches to Natural Sciences

Seminar: Applied Algorithms in Estimation and in Control of Technical, Economical, and Biological Dynamical Systems

Winter Semester 2024/2025

Leuphana Universität Lüneburg

Project Report

by

Sahadsawat Khamsuk

4000595

submitted to

Prof. Dr. Paolo Mercorelli

submitted on

23.04.2025

Introduction

Language change is a phenomenon that grabs the interest of many linguists who passionately question how a language evolves. The findings in the field so far have shown that the process involves many complex factors, making it a tiresome task for researchers of language evolution.

Thanks to the advancement of computational technology, computer simulation has gradually become another approach in studying languages. As every method inevitably comes with its own limitations, this recent method, however, has given many advantages and become needed in investigating language evolution (Gong & Shuai, 2013; Cangelosi & Parisi, 2002). For instance, De Oliveira et al. (2008) exhibit simulations of language families' populations. Sevenants and Speelman (2020) present how to model the divergence of the Standard Dutch pronunciations in the Netherlands and Belgium. It could then be claimed that linguistics has benefited from computer simulations.

Provided the advantage of computational approach in linguistics, this project attempts to develop a model that can simulate and predict how language change occurs. To give a clear subject of study and make this goal more concrete, I specify that the spread of language change I mean in this project is the acceptance of slang word *x* in language *L*. Additionally, given the analogous dynamics of language spread and disease dispersal, I want to apply the SIR model (Susceptible-Infected-Recovered) to simulation of language change. This objective then can be formulated into a research question as follows: How can the SIR model be adapted to simulate how a slang word spreads?

In the next section, I reviewed theories relevant to executing the simulation. Then, I explained how the model is designed and developed from the principles employed. Next, I provided details of the implementation and results of the simulation process. Following this, the results as well as limitations and suggestions for future research were further discussed. Lastly, in the Conclusion, I summarized the findings, answering the question of this research.

Theoretical Background

The simulation conducted in this research project is informed by two theoretical frameworks: Susceptible-Infected-Recovered model (SIR model) and Social Impact Theory. The former theory usually accounts for how an epidemic spreads through the cycle of infectious disease, while the latter takes into account social factors influencing language diffusion.

Susceptible-Infected-Recovered Model

As “a contribution to the mathematical theory of epidemics”, Kermack and McKendrick (1927) contends that an individual person suspicious of contracting a disease is a factor prominently determining the spread of the contagious disease and differentiates three stages of the infection: 1) Susceptible, uninfected individuals who risk of being infected; 2) Infective, those individuals who contracted the disease; and 3) Recovered, those removed from the infected population, either from recovery or decease. Their contribution has since then majorly benefited the field of epidemiology (cf. Talukder, 2020; Bettencourt & Ribeiro, 2008) as well as applications of the theory to other subject matters (cf. Woo & Chen, 2016; Bettencourt et al., 2008).

This mathematical theory of disease spread can be summarized and derived into a set of differential equations as displayed below:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

The three letters S , I and R stand for the three compartments defined above. Each population group is specified in percentage. Other than these compartments, there are then two parameters: 1) beta β, the rate of disease transmission; and 2) gamma γ, the rate of recovery. While there are other adjusted or extended versions of the model that can capture nuances and complexity of the spread, the algorithm used in this research relies on this standard SIR model.

Social Impact Theory

Originally developed by Latané (1981), Social Impact Theory is used for modelling impact on an individual of other individuals surrounding that person. It is specified that social impact should be formed as a multiplicative function of these following social forces: 1) strength, power or salience of the influential group which is usually determined by their socioeconomic status; 2) immediacy, closeness or social distance between individuals; and 3) number of other people. The theory then has been applied and employed as a ground for computer simulations.

Among these applications of the theory, Nettle (1999) is one of the researchers who appropriate the principle of social impact for simulating language change. In his study, he makes a case where learners are exposed in their social groups to speakers who use a variant of language. Then, to a point of time, the new feature will be adopted and used commonly among the group. Given this scenario, he then formulates the impact into a formular as follows:

$$i = b \cdot f(S, D, N)$$

The impact i is a product of the function of social impact and constant b, which represents the bias of being acquainted with the new feature. Parameter S refers to difference of influence determined by social status between individuals in a group. The parameter D, denotes social distance between individuals within the group. Some may be closer and probably have a greater effect. The variable N is added to include influence of number of the individuals in the group. The actual equation Nettle employs in his research is more complex[1], but, for the simplicity of this project, this function is kept in this simplified form and will be soon adapted.

Adjusting SIR Model with Social Impact Theory

As the standard SIR model, which is primarily used for predicting dispersal of disease, may be too simplistic for predicting language diffusion, I modified the model by extending the formula with Nettle's (1999) function of social impact.

Firstly, before adapting the SIR model for simulating the spread of a novel linguistic item, I drew an analogy of language and disease, basing this study on the following assumption:

A new language characteristic spreads in a social group
similarly to how an infectious disease disperses.

To put it simply, a new slang word is contagious among speakers like a disease. When a person is exposed to a new term and then uses that word with another person who is unaware of it, that 'infected' person is likely to adopt the word and share it with other people, spreading the slang in their society. And as time passes, they later 'recover' and gradually use the word less.

[1] $i = \mathrm{b} \cdot f(\mathrm{S}, \mathrm{D}, \mathrm{N}) = \mathrm{b}\ \mathrm{N}^{a}[\Sigma(\mathrm{s}/\mathrm{d}^2)/\mathrm{N}]$, where the alphabet a is an exponent of the number

As the rationale of utilizing the SIR model for modelling the spread of language change is now grounded, I then proceeded to recontextualize this mathematical model to fit the context of language evolution. The Susceptible are those individuals who are likely to use a slang word but have not yet adopted it. The Infected are the individuals who have acquired and used the term, further transmitting the item. Later, at some point of time, the term may become out of fashion, and there are more in a group of the Recovered who stop using the term.

That we have been informed of the Social Impact Theory, it makes sense to incorporate the effect of social impact into the rate β, the likelihood of someone adopting and transmitting a new language trait after learning about it, by taking into account the influences of social status, social distance and number of users of the new feature. The relationship of the spread rate β and these variables introduced can be formulated as follows:

$$\beta \propto S \cdot \frac{1}{D} \cdot N$$

$$\beta_i(t) = \beta_0 \cdot S(t) \cdot I(t) \cdot I^k$$

Where:

$$S(t) = S_0 + f(t)$$

$$I(t) = I_0 + g(t)$$

The social impact function as employed by Nettle (1999) can be simplified as the interaction between the social factors as shown in the variation formula. The rate β increases when the feature is adopted by highly important people and used by many people. The social status $S(t)$ can be considered in terms of baseline social influence S_0 with a change in strength over time $f(t)$. The transmission rate however decreases when the social gap is large. To align this variable of social distance with the others, I reversed it back to social intimacy $I(t)$, which can be accounted for by baseline social immediacy I_0 with a change in the quality over time $g(t)$. When formulating this relationship into an equation of social-factor-informed rate called $\beta_i(t)$, the baseline diffusion rate β_0 was added as a constant. The number of users N can be substituted by the number of the infected I. The scaling exponent k was included to control the impact of number of users of the slang term adopted. With this formulation of spreading likelihood, the estimating algorithm is now informed of social factors.

Given this new conceptualization of the transmission rate $\beta_i(t)$, the standard SIR model can be adjusted to simulate the adoption of a slang item in a community. The adjusted version of the model is illustrated as follows:

$$\frac{dS}{dt} = -\beta_i(t)SI$$
$$\frac{dI}{dt} = \beta_i(t)SI - \gamma I$$
$$\frac{dR}{dt} = \gamma I$$

Where:

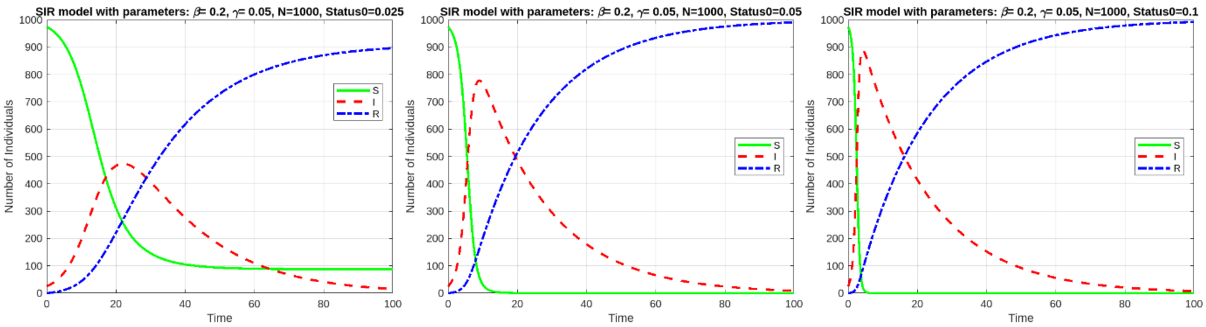
$$\beta_i(t) = \beta_0 \cdot S(t) \cdot I(t) \cdot I^k$$
$$S(t) = S_0 + f(t)\ ;\ I(t) = I_0 + g(t)$$

Simulation and Result

The SIR algorithm adjusted with Social Impact Theory was implemented in MATLAB code by adapting from Alexandre Celestino's (2021) code. Following this, by using influence of social status as a treatment, I tested the model and demonstrated how the algorithm estimates a slang term is adopted by a community.

Hera are control conditions of this simulation. Given a community of 1000 individuals, a new slang word 'lulu' is initially introduced to 25 people. The baseline transmission rate of the term is relatively mild ($\beta_0 = 20\%$) while the rate of abandoning the term is four times lower ($\gamma = 5\%$). The baseline for the influence of social intimacy I_0 is 2.5%. The scaling exponential is at 0.25. These values were kept constant throughout the simulation.

I have a hypothesis that the slang word will spread faster if it is adopted by a group of more influential status in that society. I provided three scenarios to simply test this presumption. In the first scenario, the term is at the beginning used and shared by popular students in a high school ($S_0 = 2.5\%$). In the second scenario, the term is initially adopted by Instagram influencers who are more famous than the students in the previous case ($S_0 = 5\%$). And, in the last scenario, the term is adopted by those celebrities who usually appear in media ($S_0 = 10\%$).



Comparing the result, at the point where time has passed for 20 units, when $S_0 = 2.5\%$, almost 500 people have adopted the slang term. However, when $S_0 = 5\%$ and $S_0 = 10\%$, the word was adopted so fast that the term becomes less used and that 500 people in the second scenario and 600 people in the third scenario have abandoned the term.

Discussions

From the result described above, it can be inferred that when the social influence of a group who adopt a slang term increases, even slightly, the term spreads and is shared so fast that it can quickly outdated and less used in the speakers' community.

However, it should be reminded that there are a few crucial limitations in this project. The algorithm used for simulating the spread of a slang word does not completely reflect nor entirely represent how the phenomenon occurs in reality. The functional purpose of the SIR model, as used in epidemiology, is to predict the spread of infectious diseases over time, which is by nature different from how human languages diffuse. There are in reality other multiple factors, internal and external, that I did not include in this research. These factors can greatly influence language change, rendering such an estimation a complex task.

Moreover, the data used for testing this model is not based on empirical data nor is it collected from actual language agents. While the result is reasonably considered valid by the virtue of nature of model and simulation, it still leaves a big gap for real-world applications. I therefore suggest that studies in future, in addition to basing research on more appropriate assumptions or developing other mathematical models (e.g. Song et al., 2018), use evidence empirically gathered from real language users to optimize the simulation procedure.

Conclusion

To conclude, in this project, I simulated how a slang term spreads in a community of language users, as a case study of language change. I did so by extending the SIR model with the Social Impact Theory, given the analogy drawn from language spread and disease dispersal. The adjusted model was implemented in MATLAB and then tested by using the algorithm to test my hypothesis on the relationship of language spread and strength of social status. The finding indicates that a slight increase in social influence can accelerate language diffusion. However, the algorithm is not perfect and still leaves room for improvement for simulation research.

References

AlexandreCelestino. (2021). *Epidemic-Models/SIR at main · AlexandreCelestino/Epidemic-Models*. GitHub. https://github.com/AlexandreCelestino/Epidemic-Models/tree/main/SIR

Bettencourt, L. M. A., Kaiser, D. I., Kaur, J., Castillo-Chávez, C., & Wojick, D. E. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics*, *75*(3), 495–518. https://doi.org/10.1007/s11192-007-1888-4

Bettencourt, L. M. A., & Ribeiro, R. M. (2008). Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE*, *3*(5), e2185. https://doi.org/10.1371/journal.pone.0002185

Cangelosi, A., & Parisi, D. (2002). Computer Simulation: a new scientific approach to the study of language evolution. In *Simulating the Evolution of Language* (pp. 3–28). Springer. https://doi.org/10.1007/978-1-4471-0663-0_1

De Oliveira, P. M. C., Stauffer, D., & De Oliveira, S. W. a. S. M. (2008). A computer simulation of language families. *Journal of Linguistics*, *44*(3), 659–675. https://www.jstor.org/stable/40058068

Gong, T., & Shuai, L. (2013). Computer simulation as a scientific approach in evolutionary linguistics. *Language Sciences*, *40*, 12–23. https://doi.org/10.1016/j.langsci.2013.04.002

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical Theory of epidemics. *Proceedings of the Royal Society of London. Series a, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721. https://www.jstor.org/stable/94815

Latané, B. (1981). The psychology of social impact. *American Psychologist*, *36*(4), 343–356. https://doi.org/10.1037/0003-066x.36.4.343

Nettle, D. (1999). Using Social Impact Theory to simulate language change. *Lingua*, *108*(2–3), 95–117. https://doi.org/10.1016/s0024-3841(98)00046-1

Sevenants, A., & Speelman, D. (2020). Keeping up with the Neighbours - An Agent-Based Simulation of the Divergence of the Standard Dutch Pronunciations in the Netherlands and Belgium. *Computational Linguistics in the Netherlands Journal*, *11*, 5–26. https://lirias.kuleuven.be/retrieve/681686

Song, Z., Chen, J., & Han, X. (2018). Simulating the Quantitative Variation of Language Speakers under Globalization. *Proceedings of the 2018 International Conference on Mathematics, Modelling, Simulation and Algorithms (MMSA 2018)*. https://doi.org/10.2991/mmsa-18.2018.102

Talukder, A. (2020). Susceptible-Infectious-Recovered (SIR) model-based forecasting of COVID-19 outbreak in Bangladesh. *International Journal of Clinical Practice*, *74*(11). https://doi.org/10.1111/ijcp.13648

Woo, J., & Chen, H. (2016). Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. *SpringerPlus*, *5*(1). https://doi.org/10.1186/s40064-016-1675-x