

Time's arrow reversed? The (a)symmetry of language change

Terttu Nevalainen^[0000-0003-3088-4903]

University of Helsinki, Finland
terttu.nevalainen@helsinki.fi

Abstract. This paper analyses language users' participation in real-time grammatical change. The question addressed is the extent to which individuals continue using both the incoming form and the recessive, outgoing form as opposed to using one of them categorically. Variable grammars are related to the socio-linguistic discussion of whether language change is a generational or a communal process. Ultimately, they also raise the question of the predictability of real-time language change.

Keywords: language change, variable grammar, S-curve model.

1 Introduction

The title of this paper was inspired by the distinction between some time-symmetrical physical phenomena and the asymmetry or irreversibility of time – “time’s arrow” – associated with many others [1]. Applied to the study of long-term language change, the distinction raises the hypothetical question of whether we could project the trajectory of a language change back to its beginning once we know its outcome. For an empiricist this would be a tall order indeed, not least because of the paucity of historical data, but from a modelling perspective the answer can be unequivocal: the trajectory of a completed language change is expected to pattern symmetrically over time and hence in principle be reversible. The perspective adopted in this paper is that of an empiricist working with data collected from a language community that extends from the early 15th century to the late 1600s.

This study analyses individual language users' participation in morphological and syntactic changes in English as they unfold in real time over three centuries. The data comes from the *Corpus of Early English Correspondence* [2], which enables a diachronic comparison of the linguistic output of the total of 778 individuals. The study of real-time processes of language change allows a closer examination of the question of whether linguistic change in progress always follows a symmetrical path of transmission. This path is expected to form an S-shaped pattern, with time plotted on the horizontal x-axis and frequency of use on the vertical y-axis, showing how the incoming variant starts off as a minority form, and then gains momentum, but takes some time to

become categorical. This model captures the spread of innovations through populations from innovators and early adopters to late adopters to laggards (see Figure 1).

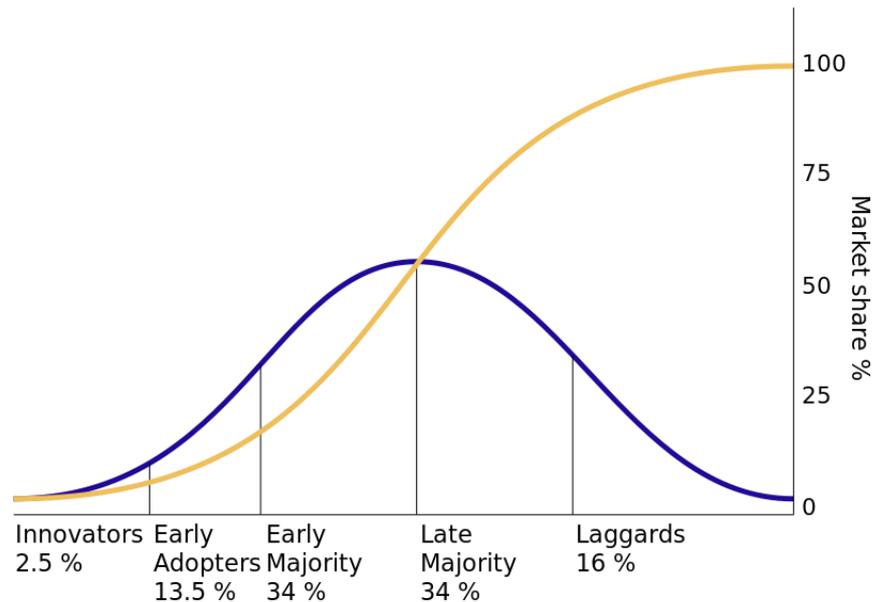


Fig. 1. The diffusion of innovations. With successive groups of consumers/speakers adopting the new technology/linguistic feature (in blue), its market share (yellow) will eventually reach the saturation level. (Source: Wikimedia Commons. Based on E. Rogers: Diffusion of innovations. Free Press, London, NY, USA, 1962).

The question addressed here is how people who can be identified as categorical users of either the incoming or the outgoing form compare and contrast with those who cannot be categorized as either, i.e., speakers with variable grammars, who are not directly accounted for in the S-curve model of change. More specifically, we would like to know how many people who participate in an ongoing change can be identified as variable language users, as is typically the case with Present-day English grammatical changes. The empirical task is then to show how these people pattern with respect to the ongoing change along the time axis.

2 Variable vs. categorical grammar

My topic of interest is connected with the theoretical issue of variable grammar in the course of a language change in progress, and generational as opposed to communal change. Sociolinguists would agree that if a change occurs in a single generation (i.e., in the childhood of a generation of speakers), it should be abrupt, but the empirical evidence we have speaks of a gradual process, although its rate of diffusion may vary

a great deal. Moreover, the argument that speakers have either a ‘new’ grammar or an ‘old’ grammar, but not a variable one, has been shown to be false [3]. Although linguistic forms are apt to become associated with semantic polarities such as formal/informal, older/younger, higher/lower, and local/outsider, we can expect non-polarized, variable use with ongoing language change as most speakers can be identified as in-betweens with respect to these polarities [4]. At the macro-level, these sociolinguistic concepts also translate into computational evolutionary models of language change as outlined, for example, in [5].

These observations have implications for the heuristic value of the S-curve model in accounting for empirical processes of change. We may ask how dominant speakers with variable grammars are in the successive stages of an ongoing change as opposed to those who have categorical grammars, either ‘new’ or ‘old’, as suggested by the logistic model. The abstract S-curve is fully symmetrical and hence in theory reversible, but we may wonder whether it is fit a description of the trajectories of change based on empirical data.

3 Data: five processes of change

The five linguistic processes to be focused on were extracted from a larger dataset and consist of morphological and syntactic changes that were in progress in English between c. 1450 and 1680. They represent different linguistic subsystems and diffused throughout the literate social section of the Early Modern English language community. As they ran their course during this period, their spread can be visualized by various approximations to an S-shaped curve at the aggregate community level [6, 7].

The following five changes are analyzed in more detail in this paper. Examples (3) and (4) illustrate the use of competing variant forms in the same context.

(my) variable: loss of the nasal in first- and second-person possessive determiners

- (1) *Myn* lord Chanselere come not here sone I come to Lundun (William Paston II, 1454; PASTON, I, 155)
- (2) Ir~es patents concernyng the Creation of *my* Erledom, (Edward Stanley, Earl of Derby, 1537; DERBY, 130)

(you) variable: replacement of the second-person subject pronoun *ye* by the object form *you*

- (3) *you* knowe for a certenty and a thinge without doubt, that *you* be bownden to obey your souerain lorde your Kyng. And therefore are *ye* bounden to leaue of the doute of your vnsure conscience in refusinge the othe, (Thomas More, 1534; MORE, 505)

(s) variable: generalization of the third-person singular present indicative suffix *-s* and loss of *-th*

- (4) lyckewisse your Joyner *comendes* hime vnto you & *sayes* he will mack you such good stufe & suche good peneworthes as he *hoopeth* shall weall licke you & contente you... (Philip Henslowe, 1593; HENSLOWE, 279)

(neg) variable: shift from multiple to single negation with indefinites

- (5) he woll *not* in *no* maner wise lese your favour (Richard Page, 1482; STONOR II, 153)
- (6) it hath bene for that I haue *not* hade *anything* to wryt of to your aduaancement. (Thomas Cromwell, 1523; CROMWELL, I, 313)

(ing) variable: verbalization of the gerund: loss of the preposition *of* in the object of the gerund

- (7) heyr is dyveres sent to proron for *byeng of grayn* (Richard Preston, 1552; JOHNSON, 1541)
- (8) I promis myselfe the contentment of *meeting you*; (Lucy Russell, 1614; CORNWALLIS, 23)

4 Method and analysis

This study draws on the “Quantifying change” project, where we focused on individual writers’ participation in real-time linguistic changes in the *Corpus of Early English Correspondence* [8]. The numerical data was retrieved from a database containing the 778 individual writers’ metadata and frequencies of use of all the linguistic variables analyzed. The linguistic information included in this master database had been previously assembled by members of the project “Sociolinguistics and language change” by using regular corpus linguistic techniques [6].

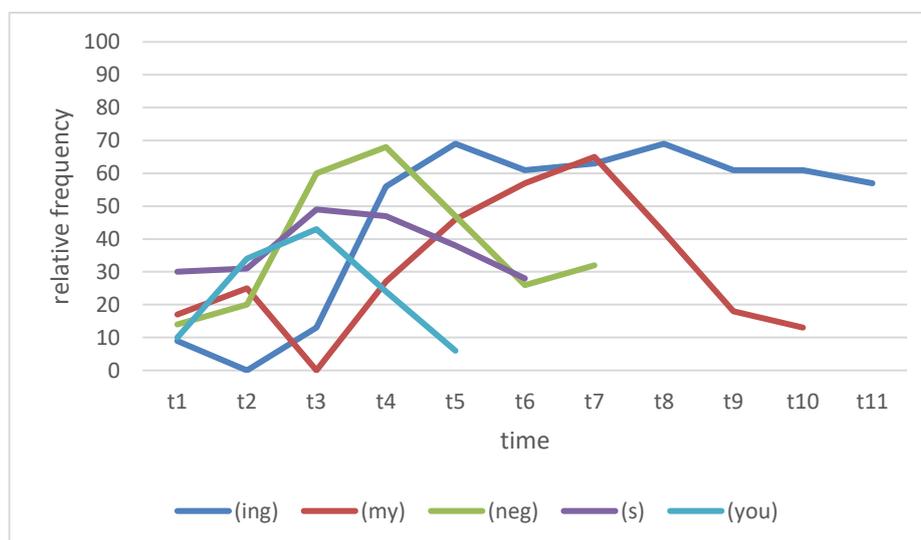
As historical data is often sparse and unevenly distributed over time, only periods with at least five people were included in each phase of each of the linguistic changes analyzed. We then separated potentially variable individuals from those (a) who did not provide enough data for a reliable analysis and (b) whose language use was categorical. Variable use was based on a minimum of six occurrences of the linguistic variable in question: if a person had at least six occurrences, he or she was regarded as a *potential* variable user of the linguistic variable, but a minimum of three instances of both the incoming form and the outgoing, recessive form was required for the person to be categorized as having an *actual* variable grammar (for more details, see [8]).

Table 1 presents the duration of the five processes of change, the number of sliding 40-year subperiods they covered, the number of actual variable users, and the number of periods with a variable user majority in the course of the change in progress. The table shows that the change of the (ing) variable took the longest to run its course, 240 years, while the change of the (you) variable was completed in half that time.

Table 1. Processes of change, their duration, and number of variable users in the CEEC.

Analyses	(ing)	(my)	(neg)	(s)	(you)
Period of change	1440–1680	1440–1650	1460–1620	1540–1680	1480–1600
Sliding 40-year subperiods	11	10	7	6	5
Number of actual variable users	178	71	61	175	76
Variable user majority	in 8/11 periods	in 2/10 periods	in 2/7 periods	in 0/6 periods	in 0/5 periods

Figure 2 diagrams the relative frequency of the number of people with actual variable grammars in the various stages of the five ongoing changes. These findings relate the number of variable users to the duration of the change, ranging from a rapid pronoun change with the fewest variable users (you) to a protracted syntactic change (ing) with a large number of people with variable grammars for much of the duration of the change. In the latter case, variable users constitute the majority from the mid-course of the change onwards.

**Fig. 2.** Proportion of individuals with variable grammars in five processes of change.

The overall trend is similar in all cases: the most people with variable grammars are found in the mid-course of the change in progress. This is also the case with many of the language users classified as progressive or conservative who have a variable grammar and use the incoming and outgoing variants side by side [8]. The results of this small-scale study therefore agree with the work on present-day languages suggesting that variable grammar is of common occurrence in ongoing processes of change at the level of individual language users. Its top frequencies coincide with the early and late majorities adopting an innovation according to the S-curve model of change shown in Figure 1.

5 Discussion and conclusion

To return to the question asked in the introduction, can the S-curve model account for these empirical findings? The answer would appear to be “yes”, “no”, and “it depends”. It is “no” if the model is interpreted in strictly categorical terms and taken to illustrate the wholesale adoption of incoming forms by an increasing number of language users. Our data show that actual changes in progress are not abrupt and do not progress in a categorical manner. On the other hand, the answer could be in the affirmative if we assumed an S-curve model that could accommodate variable adoption of innovations and the co-occurrence of old and new variants in different proportions. However, the ultimate answer to the question is probably undecided: rather than mixing two levels of abstraction by comparing descriptive empirical findings with symmetric S-curves we should perhaps model the variable reality in its own terms, as is proposed, for example, in [9]. Overall, the details of language change cannot be told in a reverse chronology.

Acknowledgements

I would like to thank the two anonymous reviewers for their helpful feedback on this paper. This study is an offshoot of the project “Quantifying change”, where we developed quantitative techniques for the study of real-time language change. My thanks to Heikki Mannila and Helena Raumolin-Brunberg for inspiring discussions, which led to answers to many old questions and gave rise to quite a few new ones. The quantitative method based on bootstrapping referred to in Section 4 was originally designed by Heikki for one of the team’s joint publications [8].

More recently, the work of the CEEC has advanced in several directions, including the Academy of Finland funded DIGIHUM project STRATAS (2016–2019), which grappled with the dimensions of spelling variation in the corpus, which is one of the major issues with digital collections of early language data. The NATAS subproject of STRATAS proposed, among other things, NLP-based solutions to the issue of spelling normalization of the corpus [10]. My thanks to Jack for his contribution to this research and for broadening our horizons on present-day low-resource languages [11].

References

1. The arrow of time. https://www.informationphilosopher.com/problems/arrow_of_time/, last accessed 2021/3/3.
2. Nevalainen, T., Raumolin-Brunberg, H., Keränen, J., Nevala, M., Nurmi, A., Palander-Collin, M. (comps): *The Corpus of Early English Correspondence*. First version. Helsinki, University of Helsinki, Department of English (1998). <https://varieng.helsinki.fi/CoRD/corpora/CEEC/>.
3. Croft, W.: *Explaining language change: An evolutionary approach*. London, Longman (2000), 45–53.
4. Labov, W.: *Principles of linguistic change: Social factors*. Oxford, Blackwell (2001), 463.
5. Bowern, C.L.: Semantic change and semantic stability: Variation is key. In: *Proceedings of the 1st International Workshop on Computational Approaches to Historical Language Change*, 48–55 (2019), 49–50. <https://www.aclweb.org/anthology/W19-4706/>.
6. Nevalainen, T., Raumolin-Brunberg, H.: *Historical sociolinguistics: Language change in Tudor and Stuart England*. 2nd edn. London, New York, Routledge (2017), 53–82.
7. Nevalainen, T.: Descriptive adequacy of the S-curve model in diachronic studies of language change. In: Sanchez-Stockhammer, C. (ed.) *Can we predict linguistic change? Studies in Variation, Contacts and Change in English 16*. Helsinki, University of Helsinki, VARIENG (2015). <https://varieng.helsinki.fi/series/volumes/16/index.html>.
8. Nevalainen, T., Raumolin-Brunberg, H., Mannila, H.: The diffusion of language change in real time: Progressive and conservative individuals and the time-depth of change. *Language Variation and Change* 23(1), 1–43 (2011).
9. Ghanbarnejad, F., Gerlach M., Miotto J.M., Altmann, E.G.: Extracting information from S-curves of language change. *Journal of the Royal Society Interface* 11, 20141044 (2014). <http://dx.doi.org/10.1098/rsif.2014.1044>.
10. Hämäläinen, M., Säily, T., Rueter, J., Tiedemann, J. & Mäkelä, E.: Revisiting NMT for normalization of early English letters. In: Alex, B., Degaetano-Ortlieb, S., Kazantseva, A., Reiter, N., Szpakowicz, S. (eds.) *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Stroudsburg: The Association for Computational Linguistics (ACL Anthology; no. W19-25), 71–75 (2019). <https://www.aclweb.org/anthology/W19-2509/>.
11. Rueter, J.: Towards a systematic characterization of dialect variation in the Erzya-speaking world: Isoglosses and their reflexes attested in and around the Dubyonki Raion. In: Shagal, K., Arjava H. (eds.) *Mordvin languages in the field*, *Uralica Helsingiensia* 10, 109–148 (2016).