Entropy reduction in language change and language acquisition

Sabine Stoll[1,2], Michele Loporcaro[1,3], Jekaterina Mažara[1,2], and Balthasar Bickel[1,2]

[1]Department of Comparative Language Science, University of Zurich, Zurich, Switzerland
[2]Center for the Interdisciplinary Study of Language Evolution (ISLE), University of Zurich, Zurich, Switzerland
[3]Department of Romance Studies, University of Zurich, Zurich, Switzerland

1 Introduction

Languages are subject to relentless change over time. This dynamic is one of the most specific features of human language compared to other communication systems across animals. How do these changes come about? It is generally recognized that language change rests on two mutually dependent and simultaneous, but mechanistically distinct processes (e.g. Bybee & Slobin 1982, Ohala 1989, Yang 2000, Croft 2000, Labov 2001, Lightfoot 2006, Niyogi & Berwick 2009), which we call here DIFFUSION and ACQUISITION.

The diffusion process starts with speakers altering language structures by choice, error, or admixture. These alternations are then replicated by others, i.e. they diffuse in the populaton. This in turn results in a shift and diversification of probability distributions of language structures. The second process, acquisition, consists in children acquiring their language from these shifted and diverse distributions. If these shifts affect their grammar, a change becomes established and characterizes the next generation.

Both processes are required for change. No change can occur unless an alteration has spread and changed the probability distribution from which the next generation acquires the language. And the spread of an alteration cannot endure unless it is inherited by subsequent generations. The processes occur simultaneously because a population of speakers typically consists of mixed ages, and so there will always be diffusion processes and acquisition processes at the same time.

The distinction between the two procceses is evidenced by the fact that only a subset of adult alterations lead to language change, e.g. not every fancy slang expression makes it to the next generation. A second piece of evidence is that children are highly resilient learners and typically only adopt changes in language usage that are sufficiently noticeable in their environment (Yang 2000, Lightfoot 2006).

Theories differ in how to model diffusion and acquisition in detail, but they agree that both processes involve a core mechanism of social learning, whereby structures are taken over by the output of others. This raises the question

whether fundamental principles of social learning are shared between diffusion and acquisition. Early attempts to answer this question have looked at concrete structures and how exactly or in which order they are learned. Parallels turned out to be limited (Vihman 1980, Bybee & Slobin 1982). More recent research has turned to general principles of language learning, such as probabilistic functions for deciding between rules and lists (Yang 2016) or general pathways of conceptual mapping (Diessel 2011). Even for these principles, the evidence is ambiguous; at least some applications and aspects seem limited to either diffusion or acquisition (Ringe & Yang 2022, Diessel 2012).

It is possible, however, that parallels might be found in principles that are even more general, holding of any type of social learning, perhaps indeed any kind of learning. One candidate for such a general principle that suggests itself from learning theory is ENTROPY REDUCTION. Consistent with a general simplicity principle in cognition (Chater & Vitányi 2003) or the Free Energy Principle in neural systems (Friston 2010), the core idea is that learning becomes more efficient by seeking the simplest mechanism that generates the data and by under-sampling counter-examples against this mechanism. What is learned as a result is patterns that are more regular, more systematic, and more coherent than they could be, given the full distribution of the data. This reduces entropy, i.e. it makes patterns more compressible in memory.

Effects of entropy reduction are widely established in diffusion processes as studied in artificial language learning experiments with adults (e.g. Kirby et al. 2008, Raviv et al. 2019, Berdicevskis & Semenuks 2022, Mansfield et al. 2022), and they are consistent with several trends in linguistic evolution

Here we contribute towards filling this gap with a case study on ANALOGICAL LEVELING. Analogical leveling is a well established process of language change whereby extra distinctions are lost, regularizing the system and reducing its entropy. A typical example is the loss of the *was* vs. *were* distinction of *to be* in some varieties of English (*they was nice*), in analogy to the fact that no other verb makes such a distinction.

We test analogical leveling in a unique situation of a system in change for which we have a clear projected outcome as already evidenced in sister languages based on historical development. At the same time, we have naturalistic longitudinal acquisition data (age 2;0 – 4;3) in which we can track whether children apply the same leveling as observable in history. The case at hand comes from agreement patterns in Sursilvan, a Rhaeto-Romance language of Switzerland. The Sursilvan agreement system has reached an intermediate stage between Latin and the other Romance languages. It has preserved a three-way contrast in gender agreement in predicative but not in attributive function Loporcaro (2017). In attributive function it has already turned into a bipartite system similar to all other modern Romance languages. The current system and frequency distributions are such that a single step of analogical leveling can be expected to happen in predicative agreement of Sursilvan as well, eliminating the extra gender distinction in predicative function. Such leveling has widely happened in the history of other Romance languages Loporcaro (2017). We test whether children apply analogical leveling as well.

2 Gender agreement in Sursilvan from a Romance perspective

Sursilvan is a Rhaetoromance variant spoken in the Grisons of Switzerland. All adult speakers are bilingual with Swiss German. Children learn both Sursilvan and German while growing up. Sursilvan and it's dialectal variant Tuatschin, from which the data of the present paper comes, are of the few Romance varieties that show remnants of the old Latin tripartite agreement system in adjectives and participles. Latin had three genders (masculine, feminine and neuter) both in attributive and predicative agreement. The neuter form in Latin was used when no controller noun was present. All modern Romance languages (e.g. French, Italian, Spanish) with the exception of three variants, Sursilvan (Loporcaro 2017), Asturian a West-Iberian variant (see Loporcaro,2017), and Maceratese, a South-Italian dialect (Paciaroni 2017, Paciaroni & Loporcaro 2010) have changed into a bipartite system with masculine and feminine gender only (see Fig. 1). In the modern Romance languages the old neuter form was lost and the masculine gender has taken over where the neuter was used in Latin Loporcaro (2017). This is an innovation of the modern Romance languages (see Figure 1 with Italian as an example). The masculine in these languages has also taken over for the default use in predicative agreement, when no controller noun is present. This is where Latin used the neuter. Thus, in the modern Romance languages the masculine form is now used not only for agreement with masculine controllers but also by default for predicative agreement whenever there is no gender-specified controller (Loporcaro 201736). The same binary contrast is found in the plural, which generates a four-cell paradigm for all agreement targets in modern Romance languages. Sursilvan, is differnt. Like all other modern Romance languages it has only masculine and feminine agreement in attributive contexts as illustrated in examples (1) and (2). The masculine is a zero form and the feminine has an -a ending. In predicative contexts, however, Sursilvan has a tripartite system, diverging from the modern Romance languages (see Figure 1). In predicative agreement , however, Sursilvan has a third form which is used when no gender-specific controller noun is present (cf. Loporcaro 2018: 36). This form goes back to the old neuter form in Latin. The form is the same as the masculine form in attributive agreement (zero-form, see Figure 1). We call it neutral in this paper, since no controller noun is present to determine the gender. The neuter from occurs in three conditions: (i) with neuter pronouns as in (1a), (ii)in infinitive constructions (1b), (iii) with place names (1c).

Participles behave similar to adjectives. Agreement in the M.SG. in predicative agreement requires the -s form. When the subject is a neuter pronoun the neuter form is chosen. This is the case when the auxiliary is 'èssar' 'to be' or 'vagni' 'to come'.

The situation for the learner is intricate because the masculine form and the neutral from in predicative agreement are the same (zero-form). The masculine form in predicate agreement has an -s which occurs only in this environment. This might constitute a challenge for the learner. Here we test whether children behave as predicted by historical development or whether they employ analogical leveling (see Fig. 1).

To test how stable the agreement system is, we handcoded a subcorpus of 90k words of adult speech from our language acquisition corpus. Out of these 90 k words, 1625 were adjectives and 1720 participles. Adults made only two errors

in predicative agreement. In both cases they used the neutral form instead of the $M_{\varnothing}$ form. In participles we found three 3 errors, again in the singular only. One error consisted in using the neutral form instead of $M_{\varnothing}$ and two errors were in the opposite direction, i.e. they used M_s instead of the neutral form $M_{\varnothing}$. This suggests that the adult system is still very stable and the few errors are not systematic. However, all errors were made in predicative agreement and they involved the vulnerable parts of the system.

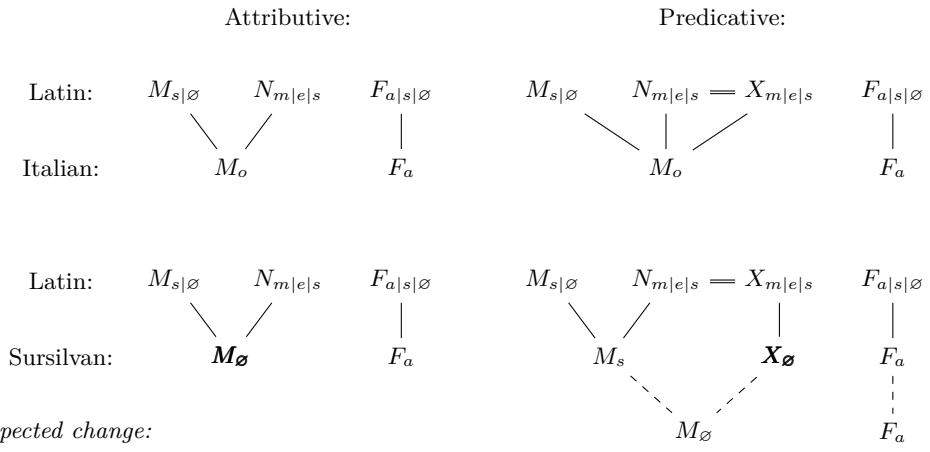

Figure 1: Evolution of adjectival and participial gender marking in most Romance languages, illustrated by Italian, and in Sursilvan. Subscripts denote the characteristic desinences across inflection classes. 'X' denotes contexts where there is no intrinsic gender in the agreement trigger, e.g. with clausal subjects. In Latin, the 'X' context is indistinguishable ('=') from neuter gender.

(1) M and F in attributive contexts:
   a. in bian magnùc
      INDEF.M.SG good.**Ø**
      'A good cheeese'
   b. ina dùna ampernaivla
      INDEF.F.SG woman pleasant-F.SG
      'A pleasant woman'

(2) M and F in predicative contexts:
   a. al paun è tgèr-**s**.
      DEF.M.SG bread is expensive-M.SG
      'Bread is expensive.'
   b. la tgarn è tgèr-**a**.
      DEF.F.SG meat is expensive-F.SG
      'Meat is expensive.'

(3) X ("neutral") in predicative contexts:
   a. tùt è stermentus tgèr.
      All.**Ø** is terribly expensive.**Ø**
      'Everything is terribly expensive.'

b. i ò da tgèsa è bi.
INF out of home is nice.∅
'Going out is nice.'

c. fugí è bian è aun mégljar.
Flee.INF be good.∅
'To flee is good and to escape is even better'.

d. Sedrún è ampernaivel.
Sedrun is pleasant.∅
'Sedrun is pleasant.'

This study aims to examine the language acquisition process in children who are learning Sursilvan-Tuatschin as their first language, with a particular focus on understanding whether they exhibit errors that converge with the historical development observed in other modern Romance languages. It is well known that errors in first language acquisition are overall very rare and we therefore need large corpora to detect errors at all (e.g. Maslen et al. 2004, Marcus 1995). In the following we explore the error patterns in a large corpus of Sursilvan-Tuatschin first language acquisition, providing the first acquisition results on agreement in a Romansh variety. We are interested in the errors in predicative vs. attributive agreement in Sursilvan-Tuatschin with respect to the role of gender. We are especially interested in the types of errors children make to compare these errors to the diachronic development of this agreement pattern found in modern Romance languages.

2.1 Predictions

If analogical leveling operates in acquisition, we expect the predicative system to simplify in analogy to the attributive system. Because the ∅-form dominates frequency, we specifically expect that children erroneously use this form where M_s is expected, but not where F_a is expected. In attributive function we expect few errors where F_a is expected, at a similarly small rate as in predicative function. If we represent errors by arrows, the predictions can be summarized as follows:

(4) Predictions under the hypothesis of analogical leveling:

$$P(M_{s[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]}) > P(F_{a[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]})$$
$$P(F_{a[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]}) \sim P(F_{a[\text{attributive}]} \rightarrow M_{\varnothing[\text{attributive}]})$$

If analogical leveling does not operate in acquisition, we expect children to make errors only as a function of probability distributions in the opportunities of making errors. In predicative functions, the ∅-form heavily dominates over of both M_s- and F_a-forms, and so we expect similarly frequent errors in the two genders. We know this distribution from analyzing the adjectives and participles used by the children in both types of agreement. In attributive functions, by contrast, the ∅-form is similarly frequent as the F_a form, and so we expect much less errors that would substitute an F_a-form by a $M_\varnothing$-form. In summary form:

(5) Predictions under the hypothesis of no analogical leveling:

$$P(M_{s[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]}) \sim P(F_{a[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]})$$
$$P(F_{a[\text{predicative}]} \rightarrow X_{\varnothing[\text{predicative}]}) > P(F_{a[\text{attributive}]} \rightarrow M_{\varnothing[\text{attributive}]})$$

3 Materials and methods

3.1 Participants

Our data come from a longitudinal audio-visual corpus of 6 children learning Sursilvan-Tuatschin as their first language (Mažara et al. 2020). The recordings were conducted in Val Tujetsch, as well as with Sursilvan-Tuatschin-speaking families in the Swiss diaspora between 2016 and 2019. Two children were recorded for two years from age 2;3 to age 4;3, two children were recorded from age 2;0 to 3;1, and another two were recorded from age 3;0 to 4;1. Both the children's linguistic situation at home and their place of residence vary. All mothers are native speakers of Sursilvan-Tuatschin. Two of the fathers are also native speakers of Sursilvan-Tuatschin, two are Sursilvan speakers, and two speak another language (Swiss-German and Italian and German). Only one of the participating families lives in Val Tujetsch, two more live in a predominantly Sursilvan speaking environment (and with a Sursilvan speaking father), while three live in a Swiss-German dominated town (two in Bonaduz, one in Zurich). Thus, we have a rather heterogenoeus samples of language backgrounds but in all family contexts the main language spoken by the mother is Sursilvan-Tuatschin.

3.2 Data

Children were video-recorded in their natural environment at home. Recordings took place once a month within a predetermined week to space recording intervals equally. The recordings were conducted by the parents of the target children. Within the recording week, parents were instructed to video-tape natural interactions in the family context, amounting to at least 4.5 hours per child divided usually into several recording sessions. OVerall, the child corpus consists of approximately 450 hours of recordings. For the present paper we used 280 h of recordings. The utterances of all speakers in this subcorpus were transcribed, translated and annotated for morphosyntax, part-of-speech, and lemmas according to the Leipzig Glossing Rules. For this study we extracted all uses of adjectives and participles used by the children and a trained linguist, who is also a native speaker of Sursilvan annotated all adjectives used by the children for attributive and predicative agreement. All agreement target forms were annotated both for contextually required as well as actually realized forms. Further the annotator marked all errors and their context.

3.3 Methods

We operationalise error as the probability of producing a ∅-form (i.e. $X_{\varnothing}$ in predicative, $M_{\varnothing}$ in attributive) in lieu of the required non-∅-form (i.e. M_s and F_a in predicative, F_a in attributive function). To estimate this probability we fit a multilevel Bernoulli model with a logit link function (also known as mixed-effects logistic regression). The predictor of interest is the required gender, "treatment"-coded so that the reference level is the gender required for the ∅-form. The coefficients of the other genders are then direct estimators of the production error.

When the predicate is a participle, we expect error probabilities to be potentially confounded by the choice of auxiliaries because one of them (*vaj* 'have')

is associated with ∅-forms. We set this form as the reference level of auxiliary choice and assess the extent to which the other auxiliaries interact with gender requirement.

We allow for group-level variation among children ('random effects') in the estimators of all population-level predictors ('fixed effects'). In order to directly capture the uncertainty of our estimates, we fit the models in a Bayesian framework, using the `brms` v. 2.18 (Bürkner 2017) interface to `Stan` (Carpenter et al. 2017). We use a flat prior on the population level, which corresponds to narrow normal distribution on the logit scale (concretely, $\mathcal{N}(0, 1.5)$), and an Exponential(1) prior on the group level.

3.4 Results

Figure 2 shows the marginal error probability across children, after controling for auxiliary choice. On average, childern substitute M_s and F_a forms by $X_\varnothing$ forms in predicative use with a very similar probability of about $P = .17$. Individual children occasionally deviate from this, with some showing more errors for M_s-forms, others more error for F_a-forms, but there is no discernible trend. Accordingly, the group-level variance among children is smaller than the population-level coefficients (see Supporting Information).

In attributive uses, errors are much smaller, with a median estimate of $P = .03$ and with limited group-level variance.

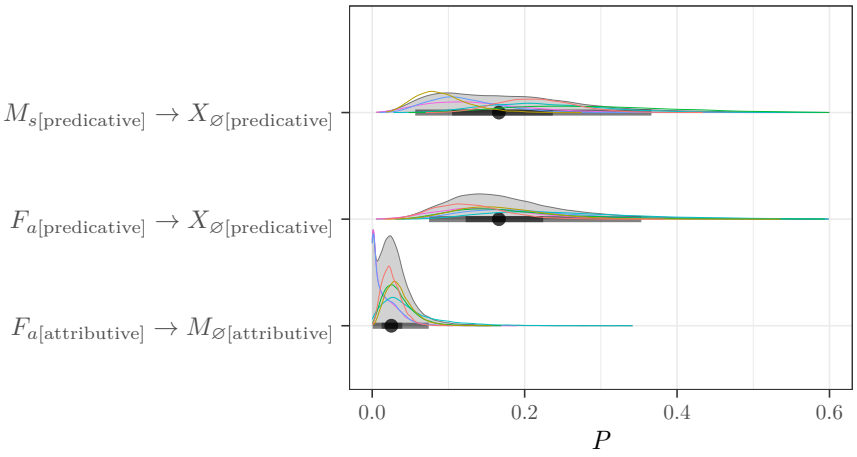

Figure 2: Marginal posterior probability of errors, after controling for auxiliary choice. Dots represent medians; horizontal bars indicate 50% and 90% credible intervals. Colored thin lines represent the estimates for individual children. For model details, see the Supporting Information.

This supports the prediction that analogical leveling is not substantiated in child data. Children make errors based on probability distributions in then opportunities of making errors. As predicted errors with feminine and masculine forms in predicative contexts are equally likely. This is presumably the case because of the preponderance of neutral forms in this context. In attributive function by contrast the two forms are approximately equally distributed (∅-form as the F_a form) and as expected the errors are much smaller.

4 Discussion

We tested whether the social learning during acquisition is characterized by the same mechanism of analogical leveling that also characterizes the social learning during diffusion in adults. Our aim was to evaluate the extent to which these two learning processes differ from each other. We used corpus data to compare the errors children make with the predicted outcome informed by the sister languages. The agreement system in Sursilvan-Tuatschin allowed to make clear predictions which could be tested empirically in the naturalistic data at hand.

Our results are consistent with previous work that showed little or no parallels between diffusion and acquisition (e.g. Vihman 1980, Bybee & Slobin 1982, Blythe & Croft 2021). The errors made by children are not subject to analogical leveling and differ substantiall from what we would expect if their errors were directing toward the projected historical development. Children's errors are rather driven by frequency distributions of the forms they use. The rarer a form the more prone to errors. This makes masculine and feminines in the predicative context equally vulnerable and explains the overuse of ∅. This corresponds to the findings in the literature on errors (e.g. Maslen et al. 2004). Assuming that children keep learning and using Sursilvan-Tuatschin in a native context, we assume that these errors will fade out after children become more productive with the agreement system.

Because entropy reduction is critically involved in both processes, the differences must have a different cause. A likely cause is that children learn language piecemeal relying strongly on frequency distributions in a situation-specific manner. Adults by contrast use new variants in adapting an underlying, already existing system. The errors conducted by adults in the present study are extremely rare, suggesting a still stable agreement syste. However, all of the errors made made by adults occured in predicative contexts which would conform to the expected historic development.

Acquisition shows in fact more entropy reduction than diffusion, but it's local, situation/function-specific: in the predicative function, they overgeneralize everything, also the F_a forms. This minimizes entropy. In the attributive function, there is nothing to minimize, so they don't.

An alternative explanation of the overuse of the ∅ form could be contact with German where there is no agreement with gender in predicative contexts. However, this explanation is extremely unlikely since children in our sample had little exposure to German. Results of this study ask for further investigations in the mechanisms underlying entropy reduction in diffusion and acquisition.

References

Berdicevskis, A. & A. Semenuks, 2022. Imperfect Language Learning Reduces Morphological Overspecification: Experimental Evidence. *PLOS ONE* 17, e0262876.

Blythe, R. A. & W. Croft, 2021. How Individuals Change Language. *PLOS ONE* 16, e0252582.

Bybee, J. L. & D. I. Slobin, 1982. Why Small Children Cannot Change Language on Their Own: Suggestions from the English Past Tense. In Ahlqvist, A. (ed.) *Current Issues in Linguistic Theory*, 29–37. Amsterdam: John Benjamins Publishing Company.

Bürkner, P.-C., 2017. Brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80, 1–28.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, & A. Riddell, 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* 76.

Chater, N. & P. Vitányi, 2003. Simplicity: A Unifying Principle in Cognitive Science? *Trends in Cognitive Sciences* 7, 19–22.

Croft, W., 2000. *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Diessel, H., 2011. Grammaticalization and Language Acquisition. In Heine, B. & H. Narrog (eds.) *The Oxford Handbook of Grammaticalization*, 130–141. Oxford: Oxford University Press.

Diessel, H., 2012. New Perspectives, Theories and Methods: Diachronic Change and Language Acquisition. In Bergs, A. & L. J. Brinton (eds.) *English Historical Linguistics 2*, 1599–1613. Berlin: De Gruyter Mouton.

Friston, K., 2010. The Free-Energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience* 11, 127–138.

Kirby, S., H. Cornish, & K. Smith, 2008. Cumulative Cultural Evolution in the Laboratory: An Experimental Approach to the Origins of Structure in Human Language. *Proceedings of the National Academy of Sciences of the United States of America* 105, 10681–10686.

Labov, W., 2001. *Principles of Linguistic Change: Social Factors*. Oxford: Wiley-Blackwell.

Lightfoot, D., 2006. *How New Languages Emerge*. Cambridge: Cambridge University Press.

Loporcaro, M., 2017. *Gender from Latin to Romance: History, Geography, Typology*. Oxford: Oxford University Press Oxford.

Mansfield, J., C. Saldana, P. Hurst, R. Nordlinger, S. Stoll, B. Bickel, & A. Perfors, 2022. Category Clustering and Morphological Learning. *Cognitive Science* 46.

Marcus, G. F., 1995. Children's Overregularization of English Plurals: A Quantitative Analysis. *Journal of Child Language* 22, 447–459.

Maslen, R. J. C., A. L. Theakston, E. V. M. Lieven, & M. Tomasello, 2004. A Dense Corpus Study of Past Tense and Plural Overregularization in English. *Journal of Speech Language and Hearing Research* 47, 1319–1333.

Niyogi, P. & R. C. Berwick, 2009. The Proper Treatment of Language Acquisition and Change in a Population Setting. *Proceedings of the National Academy of Sciences* 106, 10124–10129.

Ohala, J. J., 1989. Sound Change Is Drawn from a Pool of Synchronic Variation. In Breivik, L. E. & E. H. Jahr (eds.) *Language Change: Contributions to the Study of Its Causes*, 171–198. New York: Mouton de Gruyter.

Paciaroni, T., 2017. *I Dialetti Del Maceratese: Fonetica e Morphologia.* Zürich: University of Zurich.

Paciaroni, T. & M. Loporcaro, 2010. Funzioni Morfologiche Della Distinzione Fra-u e-o Nei Dialetti Del Maceratese .

Raviv, L., A. Meyer, & S. Lev-Ari, 2019. Compositional Structure Can Emerge without Generational Transmission. *Cognition* 182, 151–164.

Ringe, D. & C. Yang, 2022. Chapter 4. The Threshold of Productivity and the 'Irregularization' of Verbs in Early Modern English. In Los, B., C. Cowie, P. Honeybone, & G. Trousdale (eds.) *English Historical Linguistics: Change in Structure and Meaning.Papers from the XXth ICEHL*, 91–112. Current Issues in Linguistic Theory, John Benjamins Publishing Company.

Vihman, MM., 1980. Sound Change and Child Language. In *Papers from the 4th*

International Conference on Historical Linguistics, 303–320. Amsterdam: Benjamins.

Yang, C., 2016. *Thee Price of Linguistic Productivity: How Children Learn to Break Rules of Language.* Cambridge, MA: MIT Press.

Yang, C. D., 2000. Internal and External Forces in Language Change. *Language Variation and Change* 12, 231–250.