Learning and Linguistic Change

Anton Karl Ingason

University of Pennsylvania

Centre for Behaviour & Evolution
Newcastle University, Oct 29, 2013

Outline

- Part I: Predicting the evolutionary trajectory of linguistic change over generations.
- Part II: Understanding what happens when a generation fails to identify a subset context for a linguistic rule.

Part I

The Evolutionary Trajectory of a New Passive in Icelandic

Joint work with Julie Anne Legate and Charles Yang

Enterprise

- This study is situated within the greater enterprise of (trying to explain everything in) historical syntax
- How can we contribute to that enterprise by applying a predictive mathematical model of language acquisition (Yang 2002; see also Berwick and Niyogi 1997; 2007) to a major syntactic change in progress – the Icelandic New Passive?

The Icelandic New Passive

Passive-like, but the object of the active (theme/patient) does not become a subject in the passive (Maling & Sigurjonsdottir 2002).

(1) a. The elf beat the **boy**$_{Object}$. (Active)
b. **The boy**$_{Subject}$ was beaten$_{Pass}$. (Canonical Passive)
c. There$_{Expl}$ was beaten$_{Pass}$ **the boy**$_{Object}$. (New Passive)

(Icelandic examples given as English glosses)

History:

- (1c) type first attested around 1950
- No difference in meaning between (1b) and (1c)
- Ongoing change from type (1b) to (1c)
- Generations of speakers differ greatly (Thrainsson et al. 2010)

Natural Selection

Two variants, A and a, compete for the same space in nature and one is better (more fit) than the other.

- 2 types of genetic material (e.g. Red vs. Green eyes)
- 2 types of grammars (e.g. Canoncial vs. New Passive)

Current generation:

- p is the proportion of A in the population
- q is the proportion of a in the population

Consider a case where A is better than a: A has fitness 1 and a has a lower fitness value, (1-s), where s reflects the fitness difference between A and a:

$$p_{nextGeneration} = p' = \frac{p}{p + q(1 - s)}$$

Goals

- Explain the spread of syntactic change in terms of language acquisition as natural selection (Yang 2002)
- Derive a specific S-shaped curve of change (ca. 1950-2050), as opposed to fitting an arbitrary S-curve to observed data
- Show how favorable conditions emerged for the Actuation of the New Passive change (early 1900s)

Voyager

- We know that the 3-year-old voyager in the linguistic learning space uses linguistic input to select from possible grammars (wherever these possibilities come from)
- Can we use natural selection as a mechanism to select the grammar most compatible with the input?

The Variational Model (Yang 2002)

Models how a child gradually makes a binary choice (e.g. **A** vs. **a**) based on evidence from linguistic input:

- When the child hears an input token/sentence, the **A** grammar is selected with probability **p** (and **a** with prob. **q**)
- **A** gets a reward/penalty based on whether **A** is compatible with the input; **p** and **q** are updated for future interactions

The child hears a sentence and selects A ...

If **A is compatible**  with input, $\begin{cases} p' = p + \gamma q \\ q' = (1-\gamma)q \end{cases}$

If **A is not compatible**  with input, $\begin{cases} p' = (1-\gamma)p \\ q' = q + \gamma p \end{cases}$

(cf. Bush & Mosteller 1951)

The Next Generation

- We know that language change introduces conflicting evidence in the input to children
- What does our model predict about acquisition in the next generation during change?

Acquisition under heterogenous input

Consider language change:

- Evidence of a new grammar (e.g. **A**) is introduced
- The child faces conflicting evidence

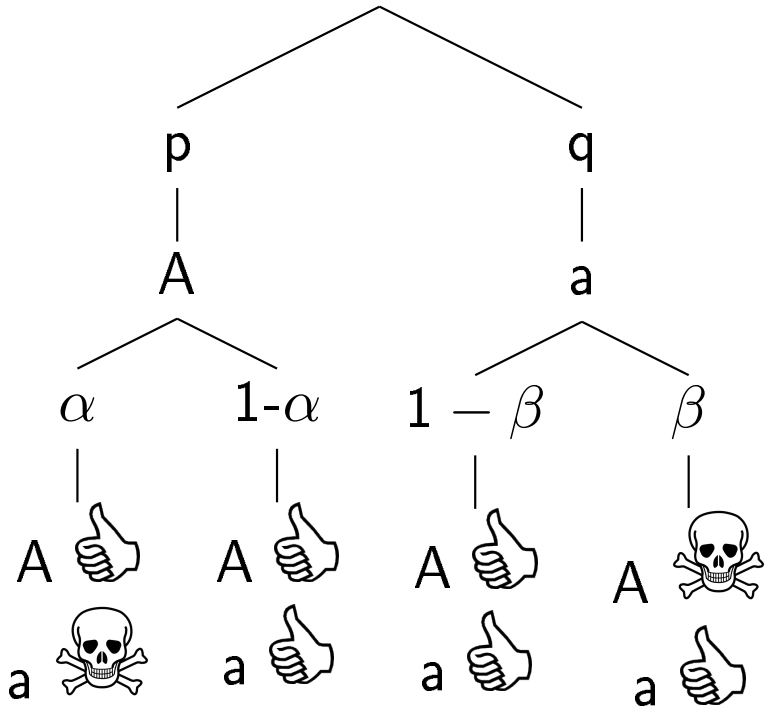
Penalty probability (PP): The probability with which a grammar is not compatible with input: PP can be measured in a corpus

The next generation (where C_A/C_a are the PP's of **A**/**a**):

$$\lim_{t\to\infty} p_t = \frac{C_a}{C_A + C_a} \qquad \lim_{t\to\infty} q_t = \frac{C_A}{C_A + C_a}$$

Note: Yang (2002) is not a model of actuation.

Last Generation Speaks

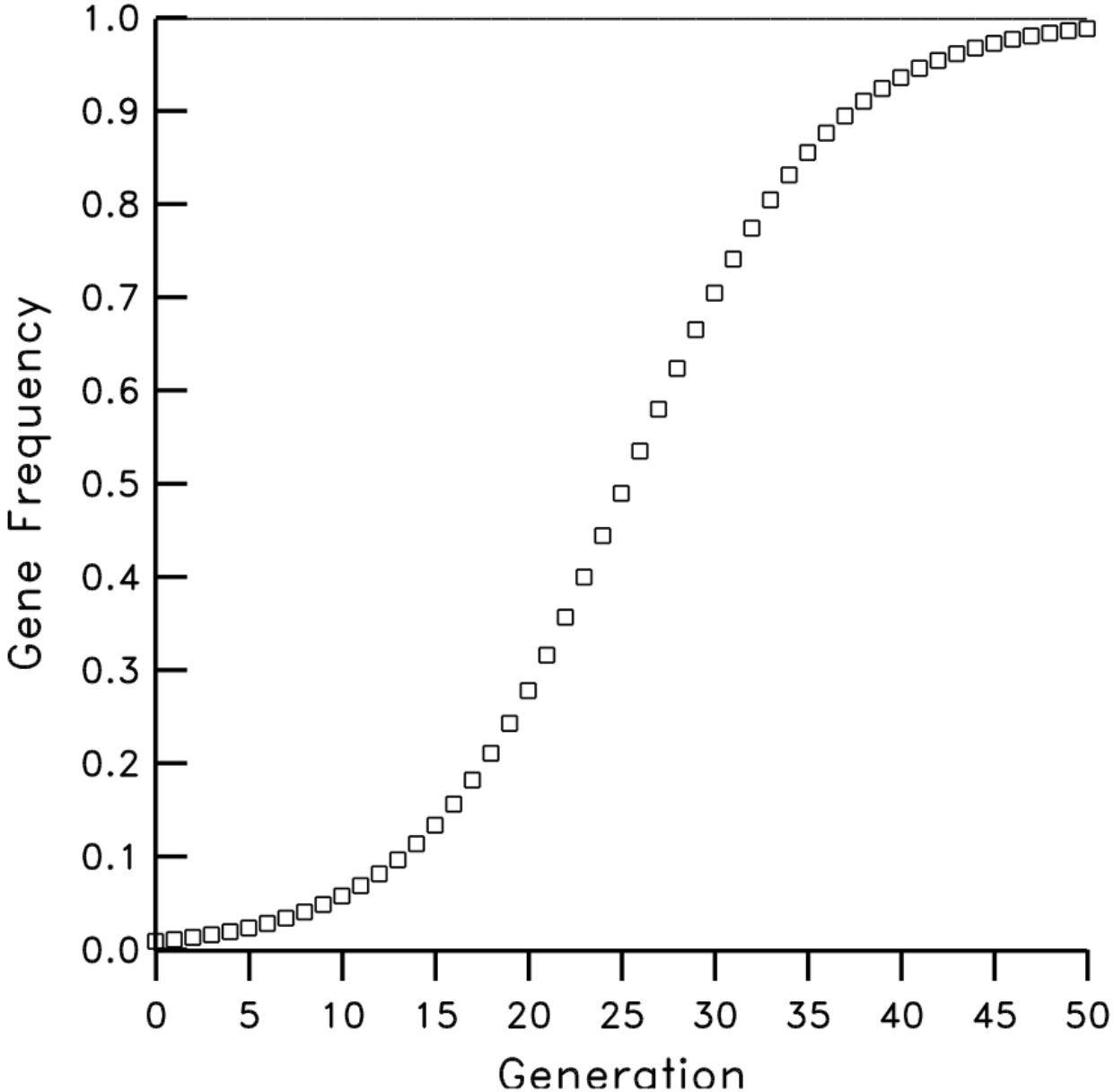


Penalty probabilities:
$C_A = \beta q \quad C_a = \alpha p$

Prediction:
If $\alpha > \beta$, A beats a

The next generation:

$$p' = \frac{\alpha p}{\alpha p + \beta q} = \frac{p}{p + (1 - s)q} \qquad \text{where } s = \frac{\alpha - \beta}{\alpha}$$


Gene Frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
25
30
35
40
45
50
Generation

Advantage of the New Passive, $\alpha \approx 0.5$

Advantage of the New Passive: definite-theme-initial. The New Passive system produces sentences like (2) as (3), which is not compatible with the Canonical Passive:

(2) **The child** was beaten.

(3) There was beaten **the child**.

There-passives and definiteness:

(4) a. There was **a child** beaten.
b. *There was **the child** beaten.

We converted the IcePaHC corpus (Wallenberg et al. 2011) from a monolingual Canonical Passive corpus to a monolingual New Passive corpus to esitmate this value.
Estimate: $\alpha \approx 0.5$

Advantage of the Canonical Passive (CanP), $\beta \approx 0.05$

Advantage of CanP: When the underlying object appears in subject position – XP aux indefinite-Theme (theme must stay low in NewP). This high subject position is not compatible with NewP:

(5) Sometimes was **a child** beaten.

(Because Icelandic is a verb second language, the subject appears after the auxiliary when any other element appears in first position)
Estimate: $\beta \approx 0.05$

Predicting spread

Prediction:
Since $\alpha > \beta$, namely $0.5 > 0.05$, the New Passive is predicted to beat the Canonical Passive and rise to complete acceptance, which is consistent with reports of its rapid spread.

(Note: We are still working on more precise values for α and β so they may not be exactly 0.5 and 0.05. Slight changes in these values do not affect the general prediction, crucially, $\alpha > \beta$.)

Generations

- Now we know that our acquisition model predicts how usage probabilities change from one generation to the next.
- How does the change evolve in subsequent generations?

Rate of Change

Next generation:

$$p' = \frac{p}{p + (1-s)q}$$

$$q' = \frac{(1-s)q}{p + (1-s)q}$$

$$\frac{q'}{p'} = (1-s)\frac{q}{p}$$

Multiply last gen by (1-s):

$$\frac{q_t}{p_t} = (1-s)^t \frac{q_0}{p_0}$$

$$t = \frac{\log \frac{q_t}{p_t} - \log \frac{q_0}{p_0}}{\log(1-s)}$$

s from α and β:

$$s = \frac{\alpha - \beta}{\alpha} \approx \frac{0.5 - 0.05}{0.5} \approx 0.9$$

Let p rise from 1% to 99%

$$t \approx -4.0/\log(1-s)$$

$$t \approx -4.0/\log(0.1) = 4$$

Ca. 4 generations

Generations and years

- 4 generations $\approx$ 100 years
- New Passive should be at just over 50% now ($\approx$60%)
- Consistent with recent studies (Thrainsson 2010)
- Acquisition of the old system predicted to die ca. 2050

First Contact

- Now we know that our model makes predictions about spread and derives a specific S-shaped curve (as opposed to fitting a curve to observed data)
- Can we say anything about why the change started when it did and not at some other time in history? Actuation event?

Actuation

- (Yang 2002) is not a model of actuation
- Still, by observing other independent changes, we can pinpoint the time when the New Passive gains its advantage
- When we get $\alpha > \beta$, conditions for spread of the new grammar are met in the input data to children
- Curiously, and apparently casting doubt on our model, the advantage of the New Passive appears to remain stable if the grammar of Icelandic is otherwise stable

Whe did the spread not start earlier?

- The grammar of Icelandic wasn't stable, so the assumption of a modern grammar in earlier times is not valid
- Crucially, there was no categorical definiteness effect (DE) until the 20th century

(6) There was beaten **the boy**

... used to be fine, until the early 20th century

- The rise of a definiteness effect allowed the New Passive to produce output not compatible with the Canoncial Passive

Part I Conclusion

- The study demonstrates that the spread of syntactic change can be explained in terms of mechanical principles
- Importantly, we derive the S-shaped curve of linguistic change from a predictive model of language acquisition (no curve-fitting)
- The actuation of the New Passive change may be rooted in the rise of a Definiteness Effect in Icelandic in the early 1900s

Part II

A Death Rattle Effect in linguistic change

(Parallels from other evolution scenarios?)

How does a rule die?

- **What happens when the context of a rule is lost?**
- **A Death Rattle Hypothesis:**
  During a period of inconsistent input, when the context of a specific minority rule has been lost from the grammar of some speakers, rule competition will result in speakers who extend the dying pattern to all items of a more general fallback context.

Icelandic experiencer subjects

English:

(7) a. **I** feel ticklish.
b. **I** like Newcastle.

Icelandic experiencers (historically):

(8) a. **Mig** kitlar.
Me.Acc.Exp-Discomfort tickles.
'I feel ticklish'
b. **Mér** líkar Newcastle.
Me.Dat.Exp likes Newcastle.
'I like Newcastle'

- Special realization of experiencers (dative case)
- Special realization of experiencers of physical discomfort (accusative case)

Blocking (Elsewhere/Specificity Hierarchy)

- Blocking in a specificity hierarchy: A rule R_1 which applies in context C_1 blocks rule R_2 in context C_2 if C_1 is a subset of C_2
- In other words: Specific rules block general rules

(9)
IF [+A, +B]
THEN apply R_1
ELSE IF [+A]
THEN apply R_2
ELSE
apply $R_{default}$

Stochastic blocking (Yang 2002)

- Stochastic blocking assumes that rules can apply with a probability lower than 100%
- Any rule based grammar + **Weights**
- Acquisition involves updating weights

Explaining overregularization

(10) Continuous account of rule acquisition

a) no specific rule
IF [+A]
THEN apply R_1
(weight≈1.00)
ELSE
apply $R_{default}$

b) overregularization
IF [+A, +B]
THEN apply R_1
(weight=0.30)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

c) adult grammar
IF [+A, +B]
THEN apply R_1
(weight≈1.00)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

Inconsistent input and mixed adult grammars

(11) Continuity in historical change

a) before change
IF [+A]
THEN apply R_1
(weight≈1.00)
ELSE
apply $R_{default}$

b) during change
IF [+A, +B]
THEN apply R_1
(weight=0.30)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

c) after change
IF [+A, +B]
THEN apply R_1
(weight≈1.00)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

A Death Rattle Hypothesis

- **What happens when the context of a rule is lost?**

How does a rule stop being productive?

- Children fail to identify [+A, +B] as a context for a productive rule
- Why?
  - For some reason (language contact, etc.) the [+A, +B] context now contains (more) items that don't undergo the rule and the exceptions derail its productivity
  - Yang (2005) suggest a tolerance threshold for listed exceptions: $N/\ln(N)$
  - Regardless of details, a context can cease to support a productive rule
- What happens then?

Losing a rule from a specific context

(12) If acquisition of rules per context is all-or-nothing, the rule "simply" disappears from the specific context [+A, +B]

a) before change

IF [+A, +B]
  THEN apply R_1
ELSE IF [+A]
  THEN apply R_2
ELSE
  apply $R_{default}$

b) after change

IF [+A]
  THEN apply R_1
ELSE
  apply $R_{default}$

Inconsistent input during the loss of a rule

- If we assume that inconsistent input leads to rule competition where weights are adjusted in reaction to input sentences – the predictions are different
- If, for whichever reason, children of some generation do not identify the specific context [+A, +B] but get conflicting evidence because other speakers do, they have to analyze that somehow

Resolving the conflict

- Some items in the specific set [+A, +B] might be demoted to per-item rules (independently of whether we assume categorical or stochastic blocking)
  IF item #512
  THEN apply R_1
  (weight=0.30)
  ELSE
  apply $R_{default}$
- But clearly, this does not happen all the time. In fact, being saved by a per-item rule correlates with high token frequency, as in the massively irregular inflection of the verb *to be*, even in English
- What if the items in [+A, +B] have low token frequency and really need a context for the rule to survive?

A Death Rattle hypothesis

The dying pattern extends to all items in the general context. The weight of R_1 will be low since most of the evidence is only compatible with the general rule.

a) before change

IF [+A, +B]
THEN apply R_1
(weight=1.00)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

b) during change

IF [+A, +B]
THEN apply R_1
(weight=0.30)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

c) death rattle phase

IF [+A]
THEN apply R_1
(weight=0.03)
ELSE IF [+A]
THEN apply R_2
(weight≈1.00)
ELSE
apply $R_{default}$

The Death Rattle of the *ticklish*-type rule

- **What does the Death Rattle hypothesis say about variable realization of experiencer subjects in Icelandic?**

ACC > DAT, *ticklish*-type merges with *like*-type

- **Old Icelandic:**
  Physical discomfort distinction always maintained

(13) a. Mig kitlar.
me.ACC.EXP-DISCOMFORT tickles.
'I feel ticklish.'

b. Mér líkar Newcastle.
me.DAT.EXP likes Newcastle.
'I like Newcastle.'

- **19th century–now:**
  Physical discomfort distinction probabilistically merged with a more general experiencer class, the *like*-type.

(14) Mér kitlar.
me.**DAT.EXP** tickles.
'I feel ticklish'

DAT>ACC (*like* > *ticklish*) Death Rattle

(15) The special experiencer contexts, of which physical discomfort is one, are subsets of the general experiencer context.

a) during change

IF [+experiencer, +phys.discomfort]
THEN apply ACC
(weight=0.30)
ELSE IF [+experiencer]
THEN apply DAT
(weight≈1.00)
ELSE
apply $R_{default}$

b) death rattle phase

IF [+experiencer]
THEN apply ACC
(weight=0.03)
ELSE IF [+experiencer]
THEN apply DAT
(weight≈1.00)
ELSE
apply $R_{default}$

A recent change

- Historical Icelandic does not seem to show any examples of a DAT to ACC (like > ticklish) extension, with any dative subject verb, until recently.
- In the last generation or two, this seems to change. Numerous ACC examples are found with every DAT subject verb, at a low probability.
- This is consistent with a Death Rattle Effect, but completely mysterious if rule learning is all-or-nothing.

Examples of DAT>ACC (*like*>*ticklish*)

(16) a. **Mig** **sýnist** það líka.
Me.ACC appears it also
'To me it also appears to be like that'

b. En **mig** **finnst** tóbakið ógeðslega gott.
But me.ACC thinks the.tobacco horribly good
'But I think the tobacco is very good'

c. Gaavuuuuuð, **mig** **leiðist** óendanlega mikið.
Gooood, me.ACC is bored infinitely much
'God, I am infinitely bored!'

d. **Mig** **líkar** litirnir.
Me.ACC likes the.colors
'I like the colors'

Predictions

- If the contexts for the accusative subjects have been completely lost in very recent history the Death Rattle Effect provides precise predictions about the language of Icelandic children right now
- **There should be no difference in case marking between verbs that previously took either dative or accusative subjects**
- **For all experiencer verbs in Icelandic, dative should be preferred, but accusative should be sporadically produced, and always at the same rate**

Evidence from language acquisition (Erlingsdóttir 2010)

- Gyða Erlingsdóttir's experiment included 46 children, age 5-7.
- Elicited production experiment, game with pictures
- 6 "DAT"-*like*-verbs
  6 "ACC"-*ticklish*-verbs

Evidence from language acquisition (Erlingsdóttir 2010)

- ACC:DAT usage for DAT-*like*-verbs: 3:192
- ACC:DAT usage for ACC-*ticklish*-verbs: 7:142
- As predicted, the difference is not significant (fischer: p=0.11)
- We could argue that the two most token-frequent ACC verbs (*langa* 'want' and *vanta* 'need') should be excluded since per-rule items might cover their case pattern. This makes the ACC:DAT usage for ACC-verbs 3:80 and p=0.37

Part II Conclusion

- The Death Rattle Hypothesis, as predicted by the model, is supported by the evidence.
- The Dative Substitution phase is over in Icelandic, since children that are growing up right now are in the Death Rattle Phase of ACC (ticklish) subjects

In sum

- We looked at two cases that involved a combination of a learning theory, and a linguistic theory of the structures involved.
- The more abstract side of linguistic structures is an interesting learning problem. Such learning operates on discrete features that we understand to a useful extent, and the consequences of learning surface in a wide range of combinatorial settings, not just a particular word.
- Studying linguistic change in terms of mechanical evolutionary dynamics gets us closer to meaningful predictions about change in linguistics, beyond fitting a curve after we already know what happened.

(The slides can be downloaded from www.linguist.is/papers)