An evolutionary approach to language change,
Part Two

William Croft
University of New Mexico
wcroft@unm.edu
Mechanisms for the generation of variation

Mechanisms and models for propagation

The emergence of New Zealand English

S-curves and the mechanisms of propagation

How do replicators gain weight?

Modeling subtler patterns in language change

Child-based and usage-based approaches
My collaborators (with many thanks!)

Alan McKane
U Manchester, UK

Richard A. Blythe
U Edinburgh, UK

Gareth Baxter
U Aveiro, Portugal
Mechanisms for the generation of variation
Grammar as system

GRAMMAR

SYNTAX

PHONOLOGY
Grammar as system vs. process

- Grammar as system assumes abstract structures, and leaves external substance (meaning, sound) out.
- Grammar as process treats the speaker as verbalizing an experience and physically realizing that verbalization.
- That is, grammar is the process of lingueme replication embedded in a larger evolutionary/complex adaptive system.
Grammar as process

SPEAKER

SYNTAX → PHONOLOGY
Another non-“analogy” with biological evolution

- It is sometimes assumed that the mechanisms of variation and selection in language change must be the same as in biology—random mutation and adaptation respectively.

- But the mechanisms are more likely to be different in each phenomenal domain.

- In fact, language change (and cultural transmission) offer a richer typology of evolutionary mechanisms than biology seems to
Progressively less intentional

**Teleological**
process aims towards change

**Intentional**
intention to change, but for goal other than change

**Nonintentional**
intention to *not* change (i.e. conform), but change happens as unintended result

Rejected by biologists and linguists
Variation: grammatical form and meaning

- **Language as system**: each language form (word, construction) has a specific meaning or meanings associated with it.

- The standard view is based on interpretation: \( \text{form} \rightarrow \text{meaning} \); the interpretation perspective makes grammar look invariant.

- **Language as process** is based on verbalization: \( \text{meaning} \rightarrow \text{form} \); the verbalization perspective reveals how variation in grammar is generated.
Variation is a two-sided process

- Both speaker and hearer are involved
- In “speaker-instigated” variation, the speaker produces variation in the production process, but the hearer must also reinterpret the output as aiming towards a different target
- In “hearer-instigated variation”, the speaker produces a potentially ambivalent utterance, but the hearer must reanalyze the linguistic structure of the utterance
Speaker-instigated variation: sound

The vowel space of Peterson and Barney (1952), illustrating the concept of categories as density distributions in a parametric space. Figure created by Stef Jannedy, and reproduced from Pierrehumbert (2003).

Phonetic realization of sound categories is variable

Sound categories are density distributions over a parametric space

Sound change is a shift in the density distributions across phonetic space

1 [3.3 [.85] A-nd u-h [1.5]] and then he gets down out of the tree,
2 [ 1.6? and [.5]] anyway,
   he comes down with a load of pears,
4 [1.4 [.55 {laugh}] . . a--nd [.35]] he walked down the ladder,
6 [.5] and he walks down the [.75] ladder,
7 [2.25 [.6] tsk [.1] A--nd [.75]] he-- [.35] was going up and down the ladder,
8 and comes down,
9 he comes off of the ladder,
10 and then he'll walk down the ladder,
11 [.4] climbs down the ladder,
12 [.85] /the whole idea/ he picked pears came down the ladder,
15 [2.65 [1.4] A--nd [.8]] he went down the ladder,
17 [1.05? [.35] and the--n . . u--m . . ] going . . down off the ladder,
18 [1.4 [.9] tsk And then . . ] climbing very carefully . . down the [.2] the ladder,
19 And he comes down,
   . . from the ladder,
20 [.6] So he didn't have to go down to the ground,
Speaker-instigated variation: grammar

• Verbalization of accompaniment events:

• Verbalized either by the preposition with or by a verb in a subordinate or coordinate clause (or, once, by a transitive verb)

2,13 [.75] And a man comes along with a goat,

7,20 [.3] A . . man with a goat [.9] tsk comes by,

19,25 [.45] and a man comes by leading a goat.

14,29 [4.1 [1.9] And u--m [1.35]] this man came by,
14,30 [.3] walked by,
14,31 and he's leading a goat,
14,32 . . that didn't want to go with him,
11,31 and this man is pulling . . a goat /h/.
Speaker-instigated variation: grammar

- Verbalization of accompaniment events:
- Verbalized either by the preposition *with* or by a verb in a subordinate or coordinate clause (or, once, by a transitive verb)

<table>
<thead>
<tr>
<th></th>
<th>go/come</th>
<th>with</th>
<th>goat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>go/come</td>
<td>lead</td>
<td>goat</td>
</tr>
<tr>
<td>Man</td>
<td>pull</td>
<td></td>
<td>goat</td>
</tr>
</tbody>
</table>
Speaker-instigated variation: grammar

- Verbalization of accompaniment events:
- Verbalized either by the preposition with or by a verb in a subordinate or coordinate clause (or, once, by a transitive verb)
- Verbs of this semantic domain are a source of comitative prepositions

*Twi* (Heine et al. 1993:212)

he- de né nnípa foro bépow
‘take’ his men ascend mountain
‘He ascends a mountain with his men.’

(Croft, “The origins of grammaticalization in the verbalization of experience”, *Linguistics*, 2010)
## Speaker-instigated variation: grammar

Variation in the verbalization of PUT in the Pear Stories

<table>
<thead>
<tr>
<th>put</th>
<th>Other verb</th>
<th>Other verbs used</th>
<th>Object being PUT</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>C5</td>
<td>15 –</td>
<td></td>
<td>singular</td>
<td>hand</td>
</tr>
<tr>
<td>A5</td>
<td>8 2</td>
<td>drop, stuff</td>
<td>distributive</td>
<td>hand</td>
</tr>
<tr>
<td>E5</td>
<td>8 4</td>
<td>load, throw, toss, pour</td>
<td>plural</td>
<td>hand</td>
</tr>
<tr>
<td>A7</td>
<td>9 13</td>
<td>empty, dump, tumble, drop, place, deposit</td>
<td>plural</td>
<td>apron</td>
</tr>
<tr>
<td>G3</td>
<td>1 6</td>
<td>deposit, dump, empty, unload</td>
<td>plural</td>
<td>apron</td>
</tr>
</tbody>
</table>
Frequency distribution of verbalizations

Horizontal axis is a one-dimensional conceptual space

Grammatical change is a shift in the frequency distribution across conceptual space

Horizontal axis is a one-dimensional conceptual space
## Verbalization and grammatical change

### Pear Stories verbs and etymology of PUT in Indo-European

<table>
<thead>
<tr>
<th>Pear Stories</th>
<th>Indo-European PUT verb</th>
<th>Source/related verb in older language</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>throw, toss</em></td>
<td>Modern Greek <em>vazo</em></td>
<td>Ancient Greek <em>bállō</em> ‘throw’, occasionally ‘put’</td>
</tr>
<tr>
<td></td>
<td>French <em>mettre</em>, Italian <em>mettere</em>, etc.</td>
<td>Latin <em>mittere</em> ‘let go, throw’, Late Latin ‘put’</td>
</tr>
<tr>
<td></td>
<td>Modern Irish <em>cuirim</em></td>
<td>Old Irish <em>cuirim</em> ‘throw, put’</td>
</tr>
<tr>
<td><em>stuff</em></td>
<td>English <em>put</em></td>
<td>Old English <em>potian</em> ‘thrust, push’</td>
</tr>
<tr>
<td><em>place</em></td>
<td>Dutch <em>plaatsen</em></td>
<td>Dutch <em>plaats</em> ‘place [n.]’</td>
</tr>
</tbody>
</table>

(Croft, “The origins of grammaticalization in the verbalization of experience”, *Linguistics*, 2010)
Implications for grammatical organization

• There is no simple one-to-one mapping between form and meaning

• Instead, the mapping is a frequency distribution of forms for each function

• The functions are arranged in a conceptual space, ideally so that each form has a unimodal frequency distribution of occurrence over functions
Variation is an inherent and ubiquitous aspect of the verbalization process.

Variation in verbalization is of the same type as grammatical change.

Grammatical change is semantically gradual, in that subtle differences in the experience being verbalized lead to small differences in the frequency distribution of variant forms across conceptual space.

Grammatical change is a change in the frequency distribution of forms across conceptual space.
### Hearer-instigated sound change

#### Table 22.2  Probabilities of identification of initial consonants as /p/, /t/, /k/ in the columns of the stimuli in the rows

<table>
<thead>
<tr>
<th>Heard →</th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>/pi/</td>
<td>.46</td>
<td>.38</td>
<td></td>
</tr>
<tr>
<td>/pa/</td>
<td>.83</td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>/pu/</td>
<td>.68</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>/ti/</td>
<td>.03</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>/ta/</td>
<td>.15</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td>/tu/</td>
<td>.10</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td>/ki/</td>
<td>.15</td>
<td>.47</td>
<td></td>
</tr>
<tr>
<td>/ka/</td>
<td>.11</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>/ku/</td>
<td>.24</td>
<td>.18</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Values on the diagonal (with borders) represent correct judgments; those off the diagonal are misperceptions. The average rate of misperception is .173. Confusions that occurred at much higher rate than this are given in italics.

*Source: Winitz et al. (1972)*

Ohala, "Phonetics and historical phonology", *Handbook of Historical Linguistics, 2003*
## Hearer-instigated sound change

<table>
<thead>
<tr>
<th>Sound change</th>
<th>Language</th>
<th>Example</th>
<th>Origin, root</th>
</tr>
</thead>
<tbody>
<tr>
<td>k &gt; t, ŋ, ž, s/ __ i, j</td>
<td>English</td>
<td><em>chicken</em> [ˈtʃɪkən]</td>
<td><em>cocc</em> + diminutive</td>
</tr>
<tr>
<td>k &gt; t, ŋ, ž, s/ __ i, j</td>
<td>English</td>
<td><em>church</em> [tʃəʧ]</td>
<td>kirke</td>
</tr>
<tr>
<td>k &gt; t, ŋ, ž, s/ __ i, j</td>
<td>French</td>
<td><em>racine</em> [ʁasɛn]</td>
<td>Gallo-Roman</td>
</tr>
<tr>
<td>k &gt; p / __ u, w</td>
<td>Classical Greek</td>
<td><em>hippos</em> ‘horse’</td>
<td>PIE <em>ekwos</em></td>
</tr>
<tr>
<td>k &gt; p / __ u, w</td>
<td>West Teke</td>
<td><em>pfuma</em> ‘chief’</td>
<td>PB *-kumu</td>
</tr>
<tr>
<td>p &gt; t / __ i, j</td>
<td>E. Bohemian Czech</td>
<td><em>tet</em> ‘five’</td>
<td>pʲt</td>
</tr>
<tr>
<td>p &gt; t / __ i, j</td>
<td>Genoese Italian</td>
<td><em>tʃena</em> ‘full’</td>
<td>pjeno</td>
</tr>
<tr>
<td>p &gt; t / __ i, j</td>
<td>Zulu</td>
<td><em>tʃʰa</em> ‘new’</td>
<td>PB *pia</td>
</tr>
</tbody>
</table>
Why does variation in verbalization and interpretation happen?

- The hearer cannot read the speaker’s mind.
- Each experience being communicated is unique and different; interlocutors must rely on common ground.
- Since each scene can have alternative construals, the hearer cannot be certain of the construal intended by the speaker.
- Speaker and hearer’s interpretation of the utterance is determined by their previous use of and exposure to the words and constructions used, but those differ for speaker and hearer.

Croft, “Toward a social cognitive linguistics”, 2009
Form-function reanalysis

• The indeterminacy in speaker construal of experience and the hearer’s interpretation of the speaker’s verbalization leads to form-function reanalysis

• In form-function reanalysis, the hearer reanalyzes the mapping between form and meaning in constructions (Croft, Explaining language change, chapter 5)
Hearer-instigated grammatical change

- **Hyperanalysis**: the hearer “overanalyzes”, analyzing out a semantic feature of a morpheme

  Locative agreement in Chichewa (Bresnan & Kanerva 1989:3):

  Mw-akhala a-nyani m-mi-tengo
  CL18-sit CL2-baboon CL18-CL4-tree

  ‘In the trees are sitting baboons.’

  > Impersonal in Zulu (Doke 1930:296):

  ku-khona izinga emazweni onke
  CL17-be.present dogs countries.LOC all

  ‘There are dogs in all countries.’
Hearer-instigated grammatical change

- **Hypoanalysis** (exaptation): the hearer “underanalyzes”, attributing a contextual semantic feature to a morpheme.

- Simple present > subjunctive: simple present is replaced in main clauses; subjunctive meaning often found in subordinate clauses is attributive to simple present form (Bybee et al. 1994:230-36; Haspelmath 1998).

*Modern Armenian (Fairbanks and Stevick 1958:118):*

```
p‘aymanón vor ušadrutyámb varèk mekenèn
condition that carefully drive.2SG car
```

‘On condition that you drive the car carefully.’

gains subjunctive meaning
Hearer-instigated grammatical change

- **Metanalysis** (invited inference): the hearer switches a contextual and an inherent semantic feature for a morpheme.

- Negative cycle: negative emphatic $\rightarrow$ negative marker:

  - **Old French:** `blet n’i poet pas creistre ‘wheat cannot grow there’`
  - **Modern French:** `je ne dis pas ça ‘I didn’t say that’`
  - **Spoken French:** `je dis pas ça ‘I didn’t say that’`

  The original negator disappears (hyperanalysis).
Hearer-instigated grammatical change

- **Cryptanalysis**: the hearer introduces a morpheme to express a meaning that is otherwise not transparent

- Reinforcement of irregular plural by a regular plural

<table>
<thead>
<tr>
<th>Chagatay</th>
<th>1ST</th>
<th>men</th>
<th>biz, bizlär</th>
</tr>
</thead>
<tbody>
<tr>
<td>2ND</td>
<td>sen</td>
<td>siz, sizlär</td>
<td></td>
</tr>
</tbody>
</table>
Mechanisms and models for propagation (selection) of a variant in a speech community
Two mechanisms of language change

• Labov cites Sturtevant: ‘the association of particular forms of speaking with the social traits of opposing social groups’ (Labov 2001:24).

• This weighting of variants has been attributed to various social measures:
  ✦ prestige (Labov 1972:290, etc.)
  ✦ covert prestige (Labov 1966:108; Trudgill 1972)
  ✦ acts of identity (LePage and Tabouret-Keller 1985)
  ✦ speaker choices in creole formation (Mufwene 2008:4)
Two mechanisms of language change

• Alternatively Labov cites Bloomfield: ‘changes in interlocutor frequencies which are in turn the result of changes in social preferences and attitudes’ (Labov 2001:24).

• Interlocutor frequencies are a reflection of social network structure:
  ✦ social network theories (Milroy & Milroy 1985)
  ✦ determinism in new-dialect formation (Trudgill 2004)
Two mechanisms of language change

- An evolutionary framework for language change provides a model of the Sturtevant-inspired mechanism and the Bloomfield-inspired mechanism.
- It also provides a model of two other mechanisms which have been appealed to in sociolinguistics and historical linguistics.
- The evolutionary framework allows us to formally model theories of the mechanisms underlying language change, and evaluate their plausibility.
Selection mechanisms

• **Replicator selection**—the Sturtevant theory—is in operation when the replicator (variant)—weights are asymmetric (unequal)

• **Interactor selection** alone is operating when replicator weights are equal

• When interactions are symmetric, then interactor (speaker) selection is *neutral*—the Bloomfield theory (social network structure)
Replicator selection

• Sturtevant-inspired mechanism

  “Linguistic change [through] the association of particular forms of speaking with … social groups”
  (Labov 2001:24)
Interactor selection

• Bloomfield-inspired mechanism
  – “Language change may simply reflect changes in interlocutor frequencies”

(Labov 2001:191)
Selection mechanisms

- But interactions may be asymmetric, in which case interactor selection is *weighted*

- Weighted interactor selection is a plausible model for accommodation, as a more local, more transitory effect (between interlocutors on a particular occasion)

- Weighted interactor selection is also a plausible model for “adopter group” theories of propagation—the Rogers theory, as a more global, more lasting effect (attitude toward a particular individual/group of individuals)
Rogers’ model of the diffusion of innovations

- **Innovators**: those who are most receptive to innovations and will transmit them to other members of the community.

- **Early adopters**: opinion leaders in the community who are respected by the rest of the community.

- **Early majority**: those who willingly adopt an innovation, but only after early adopters have done so.

- **Late majority**: those who wait until a majority of members of the community have adopted an innovation before adopting it themselves.

- **Laggards**: those who resist innovations, and ‘possess almost no opinion leadership’.
Adopter group models in theories of language change

- Labov (2001: 356-60): two-way division into leaders and followers, following Katz and Lazarsfeld (1955)
- Milroy & Milroy (1985): cite Rogers (1962), but focus analysis on leaders and early adopters
- Sankoff & Blondeau (2007): describe early adopters, late adopters, a late majority-like group, and laggards
- Nevalainen et al. (2011): progressive, in-between and conservative for each of a series of changes
Neutral evolution (genetic drift)

- Finally, no selection may be operating at all
- Yet propagation can happen by random processes in a finite population
- This is called *neutral evolution* or genetic drift (Trudgill 2004, Kauhanen 2017)
- Genetic drift $\neq$ Sapir’s concept of drift!!
## Summary of selection mechanisms

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>Variants (replicators)</th>
<th>Speakers (interactors)</th>
<th>Interaction frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicator selection</td>
<td>Asymmetric</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weighted interactor selection</td>
<td>Symmetric</td>
<td>Asymmetric</td>
<td></td>
</tr>
<tr>
<td>Neutral interactor selection</td>
<td>Symmetric</td>
<td>Symmetric</td>
<td>Unequal</td>
</tr>
<tr>
<td>Neutral evolution</td>
<td>Symmetric</td>
<td>Symmetric</td>
<td>Equal</td>
</tr>
</tbody>
</table>
Illustration of mechanisms of selection

- *Replicator selection:* I wear a hoodie because hoodies are cool (or make me look cool); it has nothing in particular to do with you
Illustration of mechanisms of selection

- **Weighted interactor selection**: I wear a hoodie because you’re cool, and you (usually) wear a hoodie; it has nothing in particular to do with hoodies per se.
• **Neutral interactor selection**: I wear a hoodie because I hang around with you a lot, and you (usually) wear a hoodie; it has nothing in particular to do with either you or hoodies.
Illustration of mechanisms of selection

- *Neutral evolution*: I wear a hoodie because I just feel like it right now; it has nothing in particular to do with anything at all
Illustration of mechanisms of selection

- **The conundrum**: In all these scenarios, I am observed wearing a hoodie; on the face of it, we can’t tell why
Empirical study of mechanisms

- Replicator selection, weighted interactor selection: survey attitudes towards social groups, linguistic variants
  - most attitudes appear to be subconscious
  - self-report of conscious attitudes don’t match behavior

- Neutral interactor selection: survey network interactions
  - not possible in practical terms for real networks (though perhaps with online networks)
  - also impractical across time
Modeling the mechanisms

• Allows us to overcome practical constraints in data collection

• Model is only as good as what you put in it
  ✴ start simple, to isolate the mechanisms
  ✴ only put in what you believe is real
  ✴ test all values on all parameters

• Modeling must be combined with empirical data (don’t model in a vacuum)

• Recognize the limits of any model
Modeling reality

• A priori, there is no obvious way to choose between these alternative mechanisms of selection

• In real societies, all factors may be operating, and they are too large and too complex to collect data directly

• However, we can model speaker behavior and observe what happens with the models when confronted with empirically observed patterns and rates of language change

• The results at least narrow the range of possible mechanisms to account for these patterns of change
The emergence of New Zealand English
New Zealand populated by waves of immigrants from Britain and Ireland from the mid 19c.

- **STAGE I** – Immigrants’ dialects largely unchanged - but were brought into contact with each other.
- **STAGE II** – Individuals from next generation used (different) mixtures of the variants available to them.
- **STAGE III** – Individuals from following generation focused on majority variant, yielding rapid homogenization of dialect across entire community.

Community size: ~100,000 speakers
Fixation timescale: ~50-60 years

(Gordon et al 2004; Trudgill 2004)
Bloomfield’s principle of density

“The reason for this intense local differentiation is evidently to be sought in the principle of density. Every speaker is constantly adapting his speech-habits to those of his interlocutors; he gives up the forms he has been using, adopts new ones, and perhaps oftenest of all, changes the frequencies of speech-forms without entirely abandoning any old ones or accepting any that are really new to him. The inhabitants of a settlement, village, or town, however, talk much more to each other than to persons who live elsewhere. When any innovation in the way of speaking spreads over a district, the limit of this spread is sure to be along some lines of weakness in the network of oral communication…”

(Bloomfield, Language [1933], p. 476; see also Labov 2001:19-20)
Trudgill’s deterministic theory

- New-dialect formation is “deterministic”, that is, no social selection mechanisms operate
- Speakers replicate what they hear around them in language use
- The majority variant that they hear ultimately wins out – no social value effects

In other words: neutral evolution, neutral interactor selection only

(Trudgill, *New-Dialect Formation* [2004])
0. Speaker $i$ has a store of variant frequencies ('grammar') $x_{iv}$ for variant $v$

1. In each timestep pick speakers $i$ and $j$ to interact with probability $G_{ij}$

2. Speakers produce a sample of tokens, speaker $i$ producing variant $v$ with probability $x_{iv}$

3. Speakers add sample frequencies to the store, speaker $i$ giving weight $H_{ij}$ to speaker $j$’s utterances.

4. Renormalize the store and return to step 1.

variationist, usage-based, model

neutral interactor selection

neutral evolution

weighted interactor selection
Evidence for determinism: the majority rule

<table>
<thead>
<tr>
<th>Retained variant</th>
<th>Initial freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-retention</td>
<td>75%</td>
</tr>
<tr>
<td>Weak vowel schwa</td>
<td>32%</td>
</tr>
<tr>
<td>Short front TRAP vowel</td>
<td>60%</td>
</tr>
<tr>
<td>Diphthong shift</td>
<td>75%</td>
</tr>
<tr>
<td>Rounded LOT vowel</td>
<td>53%</td>
</tr>
<tr>
<td>DANCE vowel</td>
<td>52%</td>
</tr>
<tr>
<td>Fronted and lowered STRUT</td>
<td>34%</td>
</tr>
</tbody>
</table>

Data from Trudgill, New-dialect formation (2004)

Social factors associated with a variant’s origin (SE England vs. elsewhere) seem to be suppressed

Neutral interactor selection: fixation probability = initial token frequency
## Evidence for determinism: the majority rule

<table>
<thead>
<tr>
<th>Retained variant</th>
<th>Initial freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-retention</td>
<td>75%</td>
</tr>
<tr>
<td>Weak vowel schwa</td>
<td>32%</td>
</tr>
<tr>
<td>Short front TRAP vowel</td>
<td>60%</td>
</tr>
<tr>
<td>Diphthong shift</td>
<td>75%</td>
</tr>
<tr>
<td>Rounded LOT vowel</td>
<td>53%</td>
</tr>
<tr>
<td>DANCE vowel</td>
<td>52%</td>
</tr>
<tr>
<td>Fronted and lowered STRUT</td>
<td>34%</td>
</tr>
</tbody>
</table>


★ Most likely outcome: majority variants fix

★ Observed outcome: relative likelihood 10%

★ Model supports the dialect mixture hypothesis
Testing Trudgill’s hypothesis

• Trudgill’s hypothesis is that only neutral evolution and neutral interactor selection play a role in the emergence of New Zealand English.

• One way to test this hypothesis is to test the likelihood that New Zealand English would emerge in the time period observed (~50-60 years).

• A well-known result is that neutral evolution and neutral interactor selection slow down as the population size ($N$) increases.

(Baxter et al. 2009, Language Variation and Change)
Testing Trudgill’s hypothesis

- The Utterance Selection model allows us to determine, for given network structure etc. the number of interactions $l_{\text{fix}}$ needed until fixation of a variant.

- We equate this to a single speaker’s lifetime (60 years) and ask how long speakers’ memory time windows (for remembering/forgetting) must be for this to be achieved.

- Since Trudgill’s model uses neutral interactor selection only, we set all $H_{ij}$ equal (proportional to a new parameter $h$).

- It turns out that our results will hold for all network structures (i.e. are independent of values for $G_{ij}$).

(Baxter et al. 2009, Language Variation and Change)
Testing Trudgill’s hypothesis

• For any network...

Seek upper bound on the shortest window in the model, for a wide range of interaction strengths (actually, $hT$)

$$t_{\text{mem}} \leq \frac{-\ln \epsilon}{\sqrt{2NT^* f(hT)\omega(x)}}$$

(derived from Baxter et al. 2006:15, equation (69))

Seek upper bound on the shortest window in the model, for a wide range of interaction strengths (actually, $hT$)
Giving him the best chance

\[ t_{\text{mem}} \leq \frac{-\ln \epsilon}{\sqrt{2NT^* f(hT)\omega(x)}} \]

- \( N \): Population in New Zealand went from 100,000 (1864) to 600,000 (1890), so we take the low value

- \( T^* \): estimated by rate of production of variants in the ONZE data (www.soundarchives.co.nz), so we take the smallest value, approximately 1.3 million tokens (actually, we use a range of values from 10^5-10^8)
Giving him the best chance

- \( x \): we use the largest value among the ONZE variables, namely 75% for H retention
- \( \epsilon \): we assume a token’s decay to 1% counts as forgetting that token (with exponential decay rate)
- \( t_{\text{mem}} \): we assume an infant’s shortest time window for remembering of 2 days (Rovee-Collier 1995)
Results

• For the generous values of $N$ (population), $x$ (variant frequency) and $\varepsilon$ (forgetting threshold) that we adopted, we find that $t_{\text{mem}}$ is well below the two-day (48 hour) limit for our range of values of $T^*$ – note that this is a log-log scale:
More realistic models

- New generations of speakers decrease the fluidity of the collective grammar and increase fixation times
- Incorporating observed immigration of new speakers delays the onset of the consensus process and increases fixation times
- Declining receptiveness with age puts a brake on the dynamics and increases fixation times
- Replicator selection can speed things up, however
Final Speculations

• Probable major factors for New Zealand English:
  ★ Immigration largely ceased around 1890, hence the speech community was more stable
  ★ Infrastructure was built in the 1870s-1880s, and schooling became universal around the same time

• *But we have already incorporated these factors (no immigration, fully connected social network)*

• It is possible that replicator selection (social valuation of variants) was due to the crystallization of internal structure in New Zealand society (stratification, European gender roles, etc.)
S-curves and the mechanisms of propagation in language change
The prevalence of S-curves

- S-curves are widely accepted by sociohistorical linguists as the normal course of propagation of an incoming variant for a linguistic variable.

- Which of the four mechanisms (neutral evolution, neutral interactor selection, weighted interactor selection, replicator selection) produce an S-curve trajectory of the propagation of a novel variant?

(Blythe and Croft 2012, Language)
S-curves in language change

Brazilian Portuguese future

Proportion vs Century

- Synth
- Haver
- Pres
- Ir

S-curves in language change

The prevalence of S-curves

- S-curves are widely accepted by sociohistorical linguists as the normal course of propagation of an incoming variant for a linguistic variable.

- Which of the four mechanisms (neutral evolution, neutral interactor selection, weighted interactor selection, replicator selection) produces an S-curve trajectory of the propagation of a novel variant?

(Blythe and Croft 2012, Language)
No selection (neutral evolution)

Successful trajectories exhibit large fluctuations

Most innovations fail
Neutral interactor selection

Exactly the same!
Directionality by boosting?

• It has been sometimes suggested that an incoming variant may be boosted, either to get a low-frequency variant across the 50% threshold to become the majority variant, or to cause a high-frequency variant to go to completion of a change.

• Frequency-based boosting represents a nonlinear function based on the input and the frequency of the variant in the speaker’s store (memory, in a usage-based model).
The threshold problem

Overproduction of majority variant exacerbates the threshold problem

Underproduction of majority variant leads to stable coexistence
Weighted interactor selection
(Labov-type leaders/followers model)

Neutral fluctuations
Rapid early rise to a characteristic frequency
Weighted interactor selection
(Rogers-type adopters categories)

Innovativeness distribution very highly skewed towards laggards

Bell-shaped distribution of innovativeness

Uniform distribution of innovativeness
How plausible are S-curves via weighted interactor selection?

- S-curves are possible via weighted interactor selection, only if the relative size of adopter groups increases approximately exponentially from leaders/early adopters through to laggards.

- What evidence is there for an exponential-like pattern?
Adopter categories: exponential vs. actual

Montreal (r) - Sankoff and Blondeau (2007)
Adopter categories: exponential vs. actual

![Graph showing the adoption categories of exponential vs. actual in Early Modern English - Nevalainen et al. (2011). The graph displays the progression of adoption over time, categorized as exponential, progressive, in-between, and conservative.]
How plausible are S-curves via weighted interactor selection?

• It is known that more generally, younger females are the leaders/early adopters, and older males are the laggards, for many types of language change.

• So community demographics would have to be highly skewed towards older people and towards males—which is far from the typical demographic distribution.
Actual age/gender pyramids

The Gambia, 1993

United States, 2000
Age/gender pyramid for exponential adopter groups (younger females leading, older males lagging)
Weighted interactor selection and S-curves

- S-curves are possible with weighted interactor selection alone, but only under very specific population structures.

- But the empirically observed patterns of leaders and laggards in individual changes, and of general age/gender demographics, indicates that the special circumstances are very far from the observed demographics.

- Another problem: Nevalainen et al. (2011) and Tagliamonte & Waters (2010 NWAV) show that different individuals are leaders or followers for different changes.
Replicator selection
(Sturtevant-inspired mechanism)

![Graph showing innovation frequency over time.

- Top graph: All speakers prefer innovation.
- Bottom graph: Some speakers prefer innovation.]
Replicator selection
(Sturtevant-inspired mechanism)

- Replicator selection produces an S-curve quite robustly
- Individuals may have different weights in favor of the incoming variant, and an overall S-curve will be produced
- Some individuals may also disfavor the incoming variant, as long as the average weight across the population is positive toward the incoming variant
Responses: Kauhanen

• Kauhanen argues that a continuous generation model plus a “socialization algorithm” that favors connecting new speakers to highly connected speakers leads to an S curve

• But this “socialization algorithm” slips in weighted interactor selection

• Kauhanen measures only monotonicity of an increase, which includes r- and J-curves, not just S-curves (measuring “S-ness” would be useful though)

• We have no reason to think our result would be different under a continuous generation model

Responses: Newberry et al.

- Newberry et al. argue that the linguistic data is binned (lumped together over a time interval), and binning a neutral trajectory that fixes will result in an S-curve.

- But this is only true of some neutral trajectories reaching fixation. Others will still display fluctuations; the latter are not observed. Stable variation, socially or functionally conditioned, doesn’t fluctuate.

- Nevertheless, their FIT algorithm may be useful for measuring the amount of noise in an S-curve, which may be due to other processes.

Newberry et al., “Detecting evolutionary forces in language change”, *Nature*, 2017
How do replicators gain weight?
Where does differential weighting of variants come from?


- Model: given a differential weighting of speaker productions (WIS), always transfer a small increment of that weight to the replicator (the variant—RS), depending on that speaker’s (perceived) frequency of use of that replicator

(Unpublished research by R. Blythe, A. McKane, G. Baxter, K. Smith, A. Jones, J. Renton, et al.)
0. Speaker $i$ has a store of variant frequencies (‘grammar’) $x_{iv}$ for variant $v$

1. In each timestep pick speakers $i$ and $j$ to interact with probability $G_{ij}$

2. Speakers produce a sample of tokens, speaker $i$ producing variant $v$ with probability $(1 + s_{iv})x_{iv}$

3. Speakers add sample frequencies to the store, speaker $i$ giving weight $H_{ij}$ to speaker $j$’s utterances.

4. Renormalize the store and return to step 1.
Extension to the Utterance Selection Model

2. Speakers produce a sample of tokens, speaker $i$ producing variant $v$ with probability $x_{iv}$

2. Speakers produce a sample of tokens, speaker $i$ producing variant $v$ with probability $(1+s_{iv})x_{iv}$

$s_{iv}$ is a score (replicator selection weight) that agent $i$ gives to variant $v$

The score is increased if variant $v$ is used more often by interlocutors $j$ that speaker $i$ has an affinity towards

$$s_{iv} = \sum_j w_{ij} y_{ijv}$$

$w_{ij}$ is affinity: simplest thing $w_{ij} \propto H_{ij}$ (weighted interactor selection)

$y_{ijv}$ is frequency that $i$ believes $j$ uses $v$: simplest thing $y_{ijv} = x_{iv}$ (actual frequency)
Simplest model of structured society

In-group affinity: $w_{in}$

Out-group affinity: $w_{out} < w_{in}$

Over-accommodation

An initial fluctuation can be amplified into a sustained difference, but no more than half the time
Portion of Facebook
Leskovec & Krevl (2014)

Convergence within well-connected social groups

Divergence between sparsely connected groups

Divergence guaranteed

Model with over-accommodation only

Fragile to addition of ties between distant groups
Discussion

- Inclusion of a replicator selection weight (score) allows small spontaneous differences between groups to be amplified and maintained.

- Specific features of social network structures seem to promote diversity between groups: strong, dense local ties, but a sparse and tree-like network of connections at the group level.

- Unclear whether this mechanism allows for low-frequency innovations to propagate: innovation will typically start out with a low replicator selection weight in this picture.
Combining the mechanisms to model subtler patterns in language change
Apparent time in language change

Canadian furniture terms

Proportion

Age

over 80  70-79  60-69  50-59  40-49  30-39  20-29  14-19

0.0%  10.0%  20.0%  30.0%  40.0%  50.0%  60.0%  70.0%  80.0%  90.0%

couch  chesterfield  sofa  davenport  settee

(Chambers, Journal of English Linguistics, 1995)
Modeling apparent time

- Speakers greatly slow down in their ability to change after adolescence; we model this with an exponential decay in the responsiveness parameter $\lambda$ as speakers “age”

- Exponential hazard function used to model death of speakers; new speakers (“children”) added each time a speaker dies

- The S-curve behavior (trajectory of change) is modeled by a uniform replicator weighting $b$, selected to produce a change over 80-150 years

- Network effects ($G$ [neutral], $H$ [weighted]) uniform

(Baxter and Croft, *Language Variation and Change*, 2016)
0. Speaker $i$ has a store of variant frequencies (‘grammar’) $x_{iv}$ for variant $v$

1. In each timestep pick speakers $i$ and $j$ to interact with probability $G_{ij}$

2. Speakers produce a sample of tokens, speaker $i$ producing variant $v$ with probability $(1 + b_{iv})x_{iv}$

3. Speakers add sample frequencies to the store, speaker $i$ giving weight $H_{ij}$ to speaker $j$’s utterances.

4. Renormalize the store and return to step 1.
Apparent time and real life

Circles = 2009 US mortality statistics (Arias 2014); solid curve = age distribution used in simulations
Modeling apparent time

Time progression, with trace for each cohort. Black line = population mean.
Modeling apparent time

Apparent time curve, taken at time indicated by red dotted line in previous figure.
The adolescent peak

Adolescent peak (more accurately, pre-adolescent trough)

Toronto quotatives

The adolescent peak

Adolescent peak (more accurately, pre-adolescent trough)

Figure 2. (ch) lenition in Panama (from Cedergren 1988:54, figure 6).
Labov’s account

• Young children are initially exposed to caregivers, who are young adults and therefore not as advanced in the change as adolescents

• Young children therefore use the incoming variant at a rate similar to their caregivers, and shift to a more advanced rate as they grow older and interact more with their older peers

• This account combines replicator selection with neutral interactor selection (dynamic network structure; $G_{ij}$ in the model)
Modeling the adolescent peak

• New speaker (“child”) has one speaker (“parent” or “caregiver”) with which she converses; parent is at age that grammar typically ceases to change

• As new speakers age, new connections to other speakers are added, with $G_{ij}$ rebalanced so each speaker is interacted with roughly equally often

• When new speaker (“child”) reaches the “parent” age cohort, the new connection may be to another new speaker (“child”)

(Baxter and Croft, Language Variation and Change, 2016)
Modeling the adolescent peak

Time progression, with trace for each cohort. Black line = population mean.
Modeling the adolescent peak

Apparent time curve, taken at time indicated by red dotted line in previous figure.
But the adolescent peak appears in almost any network structure
Two patterns of individual change

- Little individual change, or gradual change

Autocorrelation for 'a gente'
R = .6439  p < .05


(Zilles, Language Variation and Change, 2005)
Two patterns of individual change

• (Near-) categorical jump from old to new variant

Figure 3. Individual percentages of $[R]/([R] + [r])$ for the 32 panel speakers for 1971 and 1984. Trajectories plotted for all speakers who showed a significant difference between the two years.

(Sankoff & Blondeau, Language, 2007)
Modeling the two patterns

• The two patterns can be attributed to the relationship between the replicator (variant) weighting \( b \) [replicator selection] and the interlocutor’s strength of accommodation \( H \) [weighted interactor selection]

• If strength of accommodation is high relative to variant weighting, then the first pattern (gradual change, and more uniform behavior) is found

• If replicator weighting is high relative to strength of accommodation, then the second pattern (more categorical behavior, and sudden shift) is found

(Baxter and Croft, *Language Variation and Change*, 2016)
Modeling the two patterns

- $b = 0.01, H = 0.022$: higher strength of accommodation
Modeling the two patterns

• \( b = .01, H = .01 \): lower strength of accommodation
The balance between accommodation and variant weight

Boundary value (uniform distribution of speakers from 0% to 100% use)
Language change: child-based and usage-based approaches
GAS, the usage-based model, and the child-based model

- GAS as instantiated above represents the usage-based model, specifically the Utterance Selection Model (USM): language change takes place in the process of language use.

- There is another model in historical linguistics, the child-based model: language change takes place as each new generation learns language “imperfectly”.

- Both models have long lineages in linguistics.
GAS, the usage-based model, and the child-based model

- The child-based model represents the following instantiation in GAS:
  - **Replicator**: mental grammar
  - **Interactor**: human being (biological organism)
  - **Environment**: biological environment of human species

- The child-based model is also a two-step model:
  - **Generation of variation**: imperfect learning of mental grammar in child language acquisition
  - **Selection**: biological selection (reproduction)
Empirical problems with the child-based model

• “Errors” (child-instigated changes) do not match the processes observed in language change—but variation in language use does (Ohala 1989, Croft 2010)

• Children are very good at learning the language around them—the “no negative evidence problem” in language acquisition (Bowerman 1987, inter alia)

• It’s not clear how grammars are “replicated”

• Children are rarely emulated in language change; they are usually emulators (cf. the adolescent peak)

• Language itself is variable, and that variability can change in adults (see the preceding study)

(Croft 2000:44-53; Bybee 2010:114-19; and references cited therein)
Data: the article cycles

- The overwhelming majority of definite articles begin as a demonstrative, develop definite functions, diverge in form from the demonstrative, and attach to the noun

- Likewise, indefinite articles begin as the numeral ‘one’, develop indefinite functions, diverge in form from ‘one’, and attach to the noun

- These are unidirectional and nonbranching grammaticalization cycles

(Unpublished research by R. Blythe and W. Croft)
Feature 37A: Definite Articles

This feature is described in the text of chapter 37. "Definite Articles" by Matthew S. Dryer

You may combine this feature with another one. Start typing the feature name or number in the field below.

Values

- Definite word distinct from demonstrative: 216
- Demonstrative word used as definite article: 69
- Definite affix: 92
- No definite, but indefinite article: 45
- No definite or indefinite article: 198

Legend - Icon size - Show/hide Labels
Fitting data to a model

• The typological distribution of definite and indefinite articles into stages, combined with a survey of 59 changes in articles in 52 languages with written records, allows us to estimate likelihoods of transitions from one stage to the next (assuming the current typological distribution is stationary, which is plausible with such an unstable typological trait)

• We can then compare child-based and usage-based models

(Unpublished research by R. Blythe and W. Croft)
The model

• Origin-fixation model: trace process from innovation to fixation (conventionalization), in such a way that the fraction of time of a language in state $i$ is proportional to the fraction of languages in state $i$

• Use the Utterance Selection Model

• Use metrics to measure plausibility of different models to fit the empirical data, that is, range of times between changes and range of population sizes of language communities
Results and discussion

- Child-based model (speaker’s grammar fixed after child language acquisition): low plausibility for the empirical data modeled

- Usage-based model (incremental change across the lifetime): much better plausibility, but depends also on assumption of short memory time for usage exemplars

- Plausibility improves in both models in heterogeneous networks with small effective population sizes
Socially structured variation

• Countless sociolinguistic studies reveal socially structured variation in language

• But how does this variation lead to change? What social behaviors lead to change rather than stable variation?

The use of the non-standard [n] variant of the variable (ing) in Norwich (Trudgill 1974)
Conclusions

• Frequency effects probably due to neutral evolution

• In neutral change, time to fixation is proportional to population size, but language change doesn’t seem to be population size dependent

• S-curves very likely the product of replicator selection

• Adolescent peak not very sensitive to social network structure

• Variation in adoption across the lifetime involves both weighted interactor selection and replicator selection
Conclusions

• Incremental transfer of valuation from speaker to variant appears to model divergence of language in speech communities given an empirically plausible social network (groups with strong ties linked by weak ties, but in a sparse, tree-like structure)

• The child-based model of innovation/propagation is unlikely to fit empirical patterns of change, that is, change which is at best weakly correlated to population size, with differences in time in a language state, that results in typologically uneven distribution of language types