Measuring the Diffusion of Linguistic Change
Wave Theory and Gravity Models

John Nerbonne
CLCG, Rijksuniversiteit Groningen

Cultural and Linguistic Diversity
Centre for Evolution of Cultural Diversity
Dec. 9-11, 2008
Overview

1. Linguistic Diffusion
2. Measuring Linguistic Distance
3. Gravity Seen Dialectometrically
4. Simulation
5. Results
6. Discussion
1872, Johannes Schmidt
- Demonstrated limitations of tree model
  - Any two branches of IE appear to share innovations
- Proposed WAVE model of diffusion
- (Some) changes propagate as waves through distinct branches of language families
1933, Leonard Bloomfield

Assumed that speech habits are constantly modified in all verbal interactions

Suggested therefore attention to the DENSITY of communication
1974, Peter Trudgill

Proposed Gravity Model, taken from social geography, as model of diffusion

- Diffusion mediated by density, à la Bloomfield
- Chance of diffusion weakens with distance (quadratically), and is strengthened by population size

Presently, the dominant model of linguistic diffusion

Studied exclusively on the basis of ongoing linguistic change
Cascades of diffusion

- Gravity models differ from wave models
- Changes flow from one larger population center to another due to effect of population size.
  - Changes may therefore pass by intermediate, smaller settlements
- aka CASCADE model
Reception of Gravity Model

- Positive reactions:
  - Callary (1975): /æ/ raised in Am. Midwest cities
    (interesting attention to transportation paths)

- Criticism that model is lacking:
  - Boberg (2000): national borders
  - Horvath & Horvath (2001): particular places may differ in influence
  - Britain (2002): calls for more sophisticated view of geography
  - Labov (2007): diffusion through adults vs. transmission through children

- No competing model, no *measure* of influence, however.
Dialectometry

- Collect several linguistic variables, measured categorically
  - lexical: hood vs. bonnet; dragonfly vs. sewing needle
  - phonological: /i/ vs. /ai/ (either); /ai/ vs. /a/ (night)
  - morphological: he was vs. were; has mowed vs. mown
  - syntactic: needs (to be) washed
- Count difference as zero where variants are the same; and count difference as 1 where they’re not
- Sum (mean) characterizes differences between varieties (dialects)
Lexical Distance vs. Geography: Séguy (1971)

\[ y = 36\sqrt{\log(x + 1)} \]
Refined Pronunciation Distance Measure

- Kessler, Heeringa, Nerbonne, Kondrak et al.
- Refine EDIT DISTANCE aka LEVENSHTEIN DISTANCE to measure pronunciation difference
- Counts cost of transforming one string into another

<table>
<thead>
<tr>
<th>Operation</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>æəftənun</td>
<td></td>
</tr>
<tr>
<td>æftənun delete œ</td>
<td>d(œ,[])=0.3</td>
</tr>
<tr>
<td>æftərnun insert r</td>
<td>d([],r)=0.2</td>
</tr>
<tr>
<td>æftərnun replace [ʊ] with u</td>
<td>d([ʊ],[u])=0.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>0.6</strong></td>
</tr>
</tbody>
</table>
Edit Distance

- Likewise produces an ALIGNMENT of strings
- Increase sensitivity of measures enormously
- Cronbach’s $\alpha > 0.8$ from $n \geq 30$
- Validated with respect to dialect speakers’ judgments of “similar” vs. “strange” ($r \approx 0.7$)

\[ \ddot{a} \ f \ t \ \ddot{e} \ \emptyset \ n \ u \ n \]
\[ \ddot{a} \ \emptyset \ f \ t \ \ddot{e} \ \ddot{r} \ n \ u \ n \]
Distance Analysis, MDS
Explaining Aggregate Distances

- (Bootstrap) clustering indicates groups with (some) explanatory significance

- Social class only indirectly present in linguistic atlases, with limited (measurable) significance.
  - Older atlases not designed to detect social influences.

- Gravity model?
Nerbonne & Heeringa’s “Dynamics of Differentiation”

- Low Saxon dialect area examined wrt “Gravity” predictions:
  \[
  L(x, y) \propto \text{geo-dist}(x, y)^2
  \]
  \[
  L(x, y) \propto 1/\text{pop}(x) \times \text{pop}(y)
  \]

- Result:
  - Clear support for inverse distance
  - No support for inverse-square relation
  - No support for influence of population size

- Linguistic distance is a sublinear function of geography
Heeringa et al (2007)

- Larger area, larger selection of large settlements
- Result:
  - Clear support for inverse distance
  - No support for inverse-square relation
  - Some influence of population size (6% of explained variance)
- Linguistic distance is a sublinear function of geography
Some Assumptions

- We examine situations of **NORMAL diffusion**
  - No recent conquests with linguistic consequences
  - No recent migrations with linguistic consequences

- *All* (remaining) variation is the result of diffusion and spontaneous innovation.

- We may operationalize the chance of significant social contact via geographic distance.
Generality of Séguy’s law

- We examine several other studies of linguistic distance as a function of geography.
- Is aggregate linguistic distance a sublinear function of geography?
Gabon Bantu

Area: Bantu
Data: 53 sites, 160 words
Source: Van der Veen, Lyon
Note: Late settlement
Bulgaria

Area: Bulgaria
Data: 482 sites, 54 words
Source: Stoykov’s atlas
Note: Long Turkish domination
Germany

Area: Germany
Data: 186 sites, 201 words
Source: Kleiner Deutscher Lautatlas (Göschel)
Area: Eastern Seaboard, US
Data: 357 sites, 145 words
Source: Mid & South Atlantic, LAMSAS (Lowman)
Note: Settlement in last 400 yr.
The Netherlands

Area: The Netherlands
Data: 424 sites, 562 words
Source: Goeman-Taeldeman-van Reenen Atlas
Norway

Area: Norway
Data: North Wind & Sun
Source: www.ling.hf.ntnu.no/nos
Summary of Empirical Results

- **SÉGUY’S LAW**: Aggregate linguistic distance grows as a sublinear function of geography.
  - Like population geneticists’ “isolation by distance” (Wright, 1943; Malécot, 1955)
- Disturbances such as late US settlement, late settling of Bantus, Turkish domination of Bulgaria do not disturb the overall trend.
- Linguists have proposed a quadratic attractive force associated with distance, but they’ve focused not on aggregates, but on *individual features*.
  - Is the sublinear aggregate distribution compatible with the quadratic influence postulated by gravity models?
Simulation of Geographic Influence

- Two means of study the relation of distribution of individual variation to distribution of aggregate variation.
  - Empirical: examine a large number of the components of aggregate curves
    - Choice of variables nontrivial
  - Via simulation: examine effect of different strengths of attraction on aggregate difference
Simulation of Geographic Influence

Strategy

- Represent each variety by a 100-dim. bit vector (100 variables), various distances $d$ from reference.
- Initialize randomly with noise (area is initially heterogeneous).
- Give each site $d \cdot c$ (alternatively, $d^2 \cdot c$) chances to change one variable to be unlike the reference site.
  - The further the site, the more chances to “escape” the influence of the reference site.
- Sum the bit vector to obtain aggregate distance.
Initialisation

declare Site record:
   geo-dist  Numeric;             %% distance from reference pt.
   item      array 100 of binary; %% 100 linguistic items
   ling-dist Numeric              %% will be aggr. sum of item vector
end declare

%% initialize 8000 sites
for d = 1:8000 of site do
   site[d].geo-dist <= d/20;     %% sites at reg. intervals
   for i = 1:100 do
      site[d].item[i] <= 0        %% initially homogeneous
   end-for;

%% add noise in the form of random differences
for n = 1:100 do                 %%% add 100 random chances at diff.
   select i random(1:100);      %% random ling. item
   generate p random(0:1);      %% chance of change: 0 <= p <= 1
   if 0.5 < p
      then site[d].item[i] <= 1 %% record difference wrt ref. site
   else site[d].item[i] <= 0   %% record identity wrt ref. site
   end-for;
end-for;                        %% end initialization
%% simulate diffusion
for each site d do:
   chance <= d / 20 ;
   for change = 1:change do
      select i random(1:100);
      generate p random(0:1);
      if 0.5 < p
         then site[d].item[i] <= 1
      else site[d].item[i] <= 0
   end-for;
   site[d].ling-dist <= sum(site[d].item, 1:100)
end-for;

%% alternative: chance <= d**2/20
%% further sites can change more
%% random ling. item
%% chance of change: 0 <= p <= 1
%% record difference wrt ref. site
%% record identity wrt ref. site
%% aggregate distance is sum
Linear vs. Quadratic Models

- The chance of change is a linear function of distance.
- The chance of change is a square of distance.
Local regression

chance of change is square of distance

- Geographic distance
- Linguistic difference

0 10 20 30 40
20 30 40 50 60
chance of change is square of distance
Conclusions wrt Simulation

- Both linear and quadratic forces result in sublinear aggregate distances.
- Only the linear force is compatible with the curve form we typically find in aggregate distributions.
- Séguy’s curve is incompatible with the inverse quadratic force postulated in the gravity model.
Discussion

- Linguists see geography as a convenient operationalization of the chance of social contact, the operative element in diffusion.
- We study diffusion via the distribution of variation rather than via tracing single ongoing changes (innovation wrt linguistic practice).
- This enables us to measure the significance of geography, which accounts for 16% – 37% of aggregate linguistic variation.
- The sublinear distribution of aggregate variation is incompatible with the inverse quadratic effect postulated in the gravity models of diffusion.
Next Steps

- Empirical examination of the distributions of component individual variables
- Use regression technique to study other putative determinants of diffusion, e.g. population size, social class, sex, ...
- Compare historical explanations embodied in phylogenetic view
Bantu and Bulgaria

![Bantu graph]

![Bulgaria graph]
Germany and LAMSAS

Germany

LAMSAS / Lowman
The Netherlands and Norway