

The Geographic Distribution of Linguistic Variation

John Nerbonne & Wilbert Heeringa University of Groningen, Humanities Computing {j.nerbonne,w.j.heeringa@rug.nl}@rug.nl

International Conference on Linguistic Evidence
 University of Tbingen, Feb. 2, 2006
 http://www.dialectometry.net/





This Talk

- Review of pronunciation measures as foundation for variation studies.
- General Question: How is variation distributed geographically?
- Gravity hypothesis, sociolinguistic objections
- Sublinear relation of linguistic vs. geographic distance





Pronunciation Difference

- Levenshtein distance calculates the (least) cost of changing one string into another
- Example: *afternoon* is pronounced as ['æəftə,nʉn] in the dialect of Savannah and as ['æftər'nun] in the dialect of Lancaster.

æəftənʉn	delete ə	1
æftənun	insert r	1
æftərnun	subst. u/u	1
æftərnun		
		3

- All operations cost *one* unit (initially), further refinements possible
- Large amounts of data compensate for rough treatment of pronunciation similarity.





Dialect distances

- Many sequence operations map [æ∂ft∂nʉn] → [æft∂rnun].
 —LEVENSHTEIN algorithm guarantees that the *least* cost mapping is found.
- \bullet Distance between two dialect varieties is equal to the average of w Levenshtein distances.
 - o automatically weights differences involving more frequent sounds more heavily
- All distances between n dialects are arranged in a $n \times n$ matrix for further analysis
- Different further analyses possible, e.g. multi-dimensional scaling.





A New Perspective (ZDL, to appear)



Foundation for dialect continuum via MDS on distance table.





Background Assumptions

- We can measure pronunciation difference consistently and validly
- Validation wrt established dialect divisions, also wrt dialect speakers' judgments of "difference"
- Studies over Dutch, Am. English, German, Bulgarian, Sardinian, Norwegian, Louisiana Creole, ...
- Via numerical distance measure an analytical foundation for "dialect continua".





Geography and Linguistics

- Part of larger assembly of questions on geography and culture
- How has geography influenced the spread of culture?
- What does the pattern of culture reveal about cultural dynamics?





Trudgill's Linguistic Cohesion via Gravity

$$F = G\frac{m_1m_2}{r^2} = G\frac{p_1p_2}{r^2}$$

- ${\cal F}~$ is the attractive force,
- m_1, m_2 the populations (p_1, p_2) of the two settlements,
- $r\,$ the distance between them, and
- ${\boldsymbol{G}}\;$ won't be speculated on

Idea: social contact promotes linguistic accommodation and linguistic similarity.





Celestial Gravity







Motivating Linguistic Cohesion via Gravity

Chance of social contact should be

- proportional to the product of settlement size and
- (if travel is random) inversely proportional to squared distance

Notate bene: we measure linguistic dissimilarity, which we postulate stands in inverse relation to the attractive force of social contact.





Predictions of Linguistic Cohesion via Gravity

$$F = G \frac{p_1 p_2}{r^2} = 1/D$$
$$D \propto 1/G \frac{r^2}{p_1 p_2}$$

$$D \propto r^2, ~~~ D \propto 1/p_1 p_2$$

- F is ling. attraction, which should produce similarity
- $D\,$ is ling. dissimilarity
- p_1, p_2 the populations of the two settlements, and
- $\boldsymbol{r}~$ the distance between them





Gravity Studies

- Trudgill examined changes in progress in East Anglia, Norway
- Callary (1975) noted /a/ changes in Am. mid-west followed degree of urbanization.
- Bailey et al. (1993) noted changes primarily in the direction predicted by gravity in Oklahoma, but also counterexamples. Inchoative *fixin' to* spread from rural to urban areas.
- Boberg (2000) criticizes gravity for ignoring political border (U.S./Canada), shows effect of border.
- \bullet Horvath (2001) see little confirmation of gravity in predicting /l/ vocalization in Australia and New Zealand
- Boberg and Horvath criticize the insensitivity of "Gravity" to (social) factors other than geography.





Evidence

- Other studies have observed diffusion of concrete changes
- We focus rather on the the residues of many changes
- Compare the reasoning linking the (flat) form of planetary systems to theories of their origins—inferring dynamics from resulting (synchronic) patterns













Look at Data

Linguistic Distance vs. Geographic Distance



Geographic Distance (m) Gravity predicts a positive quadratic effect!





Quadratic?

Linguistic Distance vs. Geographic Distance



Geographic Distance (m) Optimal positive quadratic line as predicted by gravity hypothesis

Shape? Zero? $(r^2 = 0.5)$







Linguistic Distance vs. Geographic Distance



Geographic Distance (m) Gravity predicts a positive quadratic effect!

Shape? Zero? $(r^2 = 0.58)$





Interpreting Results

Trudgill's gravity model predicts:

- Attraction is relatively stronger over short distances
- Therefore linguistic distances should be relatively smaller over these short geographic distances

Observations:

- Linguistic distance increases positively with geographic distance, *but*
- Effect is proportionately **greater** over short distances rather than proportionately smaller





Speculation on Cultural Dynamics

Not **attraction**, as Trudgill postulates, but rather **repulsion/fission/differentiation** is the fundamental cultural dynamic.

It is natural to see this grow relatively weaker over long distances.

In spite of enormous linguistic pressures toward accommodation.





Range of Populations



Ramaer (1931) Geschiedkundige atlas van Nederland; Het koninkrijk der Nederlanden 1815-1931





Further Results

• Very weak, and surprisingly, also **positive correlation** of ling. distance with population size

-likewise suggests fundamentally repellent force

• Van Gemert (2002) and Gooskens (2004) attempt improvement using 19th cent. travel time instead of geography

-no improvement in (flat) Netherlands (van Gemert), massive improvement in rugged Norway (Gooskens)





Least cost travel routes



Correlation travel cost, as-the-crow-flies distance (r = 0.92)





Conclusions & Speculations

Conclusions

- Geography explains 58% of aggregate variation in sample
- Linguistic variation increases sublinearly wrt geography, not quadratically, as "Gravity" predicts

Speculations

- Social factors will explain much less than 58% of aggregate variation (needs empirical study!)
- The fundamental dynamic in linguistic variation is differentiation!





Further Questions

- How does varietal distance compare to other indicators of cultural affinity?
 —schooling, dress, church, recreation, architecture, ...
- How does varietal distance compare to Indicators of genetic relationship? —genetic distance, patronymic distance
- Are there better (secondary) predictors of varietal distance?
 —waterways, trade connections, marriage patterns, pilgrim routes, …

