

The Dative Alternation in African American English: Researching Syntactic Variation and Change in a Conglomerated Corpus

Tyler Kendall, Northwestern University & NC State University

Gerard Van Herk, Memorial University Newfoundland

Joan Bresnan, Stanford University

AAFL 2009, University of Alberta

Slides: <http://ncslaap.lib.ncsu.edu/pdfs/aacl2009-kvhb.pdf>

Contact: t-kendall@northwestern.edu

The dative alternation

a. *Who gave that wonderful watch to you?*
prepositional (*to-*)dative
theme-NP recipient-PP

b. *Who gave you that wonderful watch?*
double object construction
recipient-NP theme-NP

Bresnan and Hay (2008: 246)

Previous research

- Previous corpus-based studies of the alternation (cf. Gries 2003, 2005, ...)
- Bresnan, Cueni, Nikitina, and Baayen (2007)
 - Examining the Switchboard Corpus (Godfrey et al. 1992) and the Wall Street Journal corpus (Marcus et al. 1993), Bresnan et al. demonstrate that a statistical model can predict the choice of the alternate in unseen data with 94% accuracy
 - Based on factors like *animacy*, *pronominality*, *discourse accessibility*, *givenness*

Inter-variety research on the alternation

- Bresnan and Hay (2008)
 - Compared *give in* American and New Zealand English varieties
- Bresnan and Ford (2009)
 - Compared American and Australian subjects' knowledge of probabilistic grammatical choices through psycholinguistic experiments, found subtle differences
- Mukherjee and Hoffman (2006)
 - Following up Olavarría de Ersson and Shaw (2003)
 - Demonstrated that the prepositional dative is more common in Indian English than British English
 - “Verb complementation has so far been underestimated as an area of the language system in which regional differentiation figures prominently” (Mukherjee and Hoffman (2006: 149))

Beyond macro-regional differences?

- Are the sorts of macro-regional differences found for the alternation evidenced in other sociolinguistic varieties, such as ethnic varieties like African American English?
- How do the probabilistic models developed by Bresnan et al. (2007) extend to ethnic, often regionally-embedded, varieties?
- Do these regionally-embedded ethnic varieties exhibit more or less variation than regionally distinct varieties?

African American English (AAE)

- A central object of study in North American sociolinguistics (cf. Wolfram 1969, Labov 1972a, ...)
- More than five times as many publications on AAE than any other ethnic or regional dialect (Schneider 1996:3)
- AAE rarely studied outside of sociolinguistic and applied/educational perspectives
 - Rarely used to examine theoretical linguistic questions

Labov (1972b)

- Sociolinguistic variables can be considered as one of:
 - **Sociolinguistic Indicators**
 - Correlate with social differences, but carry little or no social meaning (e.g., *caught/cot* vowels in US)
 - **Sociolinguistic Markers**
 - Vary stylistically as well as socially, carry observable meaning (e.g., *-in'* for *-ing*)
 - **Sociolinguistic Stereotypes**
 - Highly salient, often commented upon and/or consciously avoided (e.g., Southern *y'all*; *ain't* for *isn't*)
- Typically, AAE studies focus on markers and stereotypes

Sociolinguistic/AAE datasets

- No available large-scale corpora of AAE
- Individual sociolinguistic collections of AAE data are most often
 - Closed resources
 - Fairly small
 - Not transcribed or not well transcribed
 - Actually contain multiple varieties – e.g., often non-AAE speaking interviewers

Our “conglomerated” data collection

- Sociolinguistic interview recordings (transcribed audio)
 - From the Sociolinguistic Archive and Analysis Project (SLAAP; <http://ncslaap.lib.ncsu.edu/>)
~ 165,000 words of AAE
 - From generous colleagues
~~ 160,000 words of AAE
- Sociolinguistically compiled antebellum ex-slave letters
 - From the Ottawa Repository of Early African American Correspondence (OREAAC)
~ 140,000 words of AAE written between 1834 and 1866 by semi-literate African American immigrants to Liberia

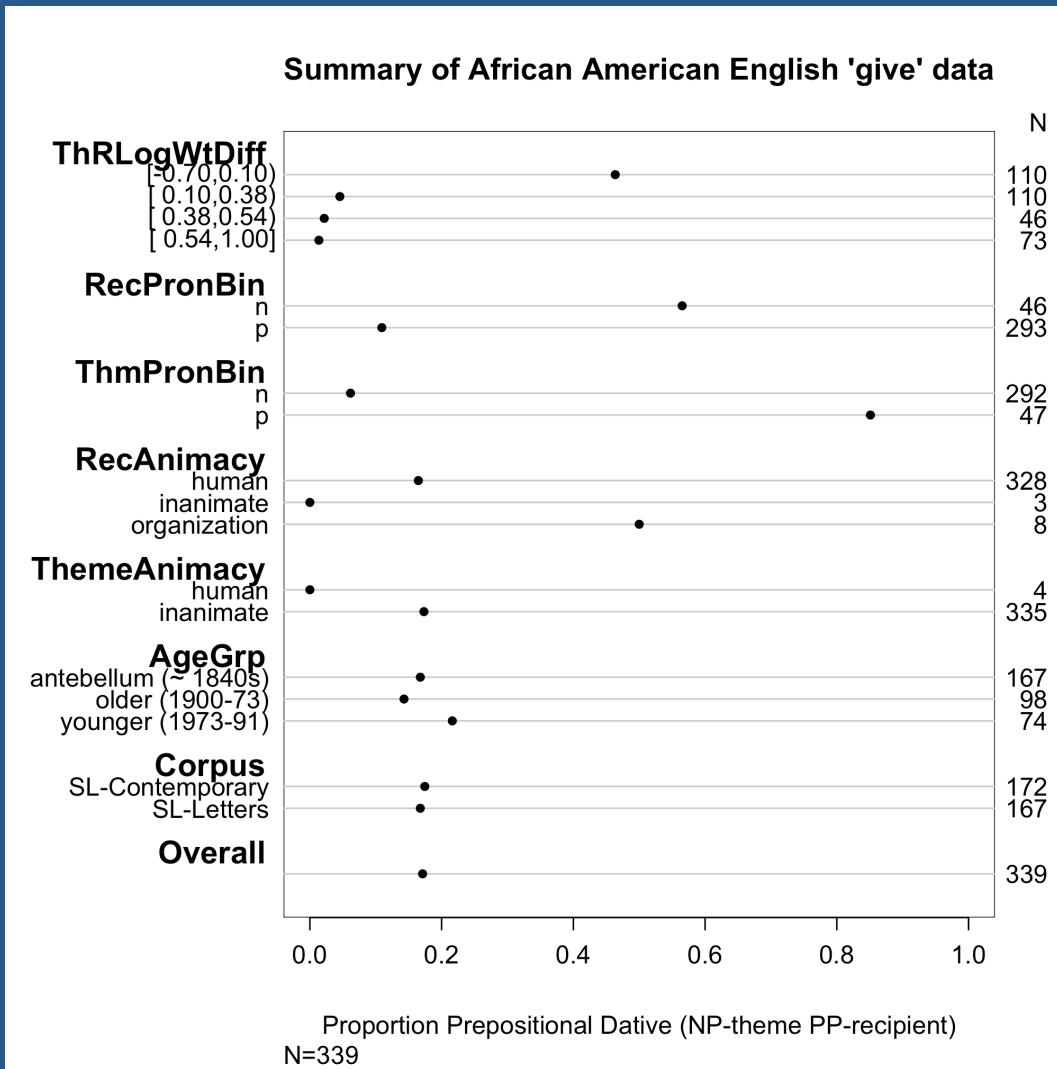
Actual examples

- “So he *gave me* potato chips”
 - Tinky, Contemporary Spoken, Rickford-EPA
- “i will *give you* Speciman of it”
 - DM, Historical Written, OREAAC
- “it has already *Given* and impetus to **Evry Branch of Bisness**”
 - NN, Historical Written, OREAAC
- “She's liable to *send me* a check or something”
 - RS, Contemporary Spoken, SLAAP-Hyde County
- “she *took* a balloon **to that baby**”
 - LM, Contemporary Spoken, SLAAP-Robeson County

Data extraction and coding

- Materials started in a number of formats
 - Spoken data were converted to plain text
 - SLAAP data are stored in a relational database, but SLAAP software exports to plain text (cf. Kendall later today)
 - Other spoken data in a range of formats
 - Word documents saved to plain text and then cleaned up in Emacs
 - Transcriber transcripts and Praat TextGrids converted to plain text via tools at <http://ncslaap.lib.ncsu.edu/tools/>
 - Then a Perl script searched for regular expression “`^g[ai]v\w*`” using a look-up table to skip non-African American speakers
 - All matches were reviewed/pruned by hand and then coded by hand
 - OREAAC materials retain highly non-standard spelling so all tokens were manually extracted and coded

AAE data summary



- SL-Contemporary, spoken:
 - N = 172 (30 NP PP, 17.4%)
- SL-Letters, historical/ written:
 - N = 167 (28 NP PP, 16.8%)
- Total:
 - N = 339 (58 NP PP, 17.1%)
- Nb,
 - About 5 tokens for every 10,000 words in our overall corpus of spoken language
 - About 12 tokens for every 10,000 words of our letters corpus

How do these data compare to SAE?

- How do these data relate to the macro-regional “standard” American English?
 - The findings of Bresnan et al. (2007)
- Extracted all *give* tokens from their dataset:
 - Switchboard
 - N = 1,263 (180 NP PP, 14.3%)
 - Wall Street Journal
 - N = 403 (76 NP PP, 18.9%)

Bresnan et al. + AAE logistic regression model

Model Coefficients

| Factor | Log-odds | <i>p</i> |
|--|----------|----------|
| Intercept | -0.4798 | 0.0037 |
| Recipient = Pronoun (RecPronBin="p") | -3.1269 | <0.0001 |
| Theme = Pronoun (ThmPronBin="p") | 4.8766 | <0.0001 |
| Theme - Recipient Log Weight Difference (ThRLogWtDiff) | -0.9969 | <0.0001 |
| Modality = Written | 0.1607 | non-sig |
| ThRLogWtDiff * Modality = Written | -1.3106 | 0.0001 |

Model: $C = 0.960$, Somers' $D_{xy} = 0.919$, Nagelkerke $R^2 = 0.690$

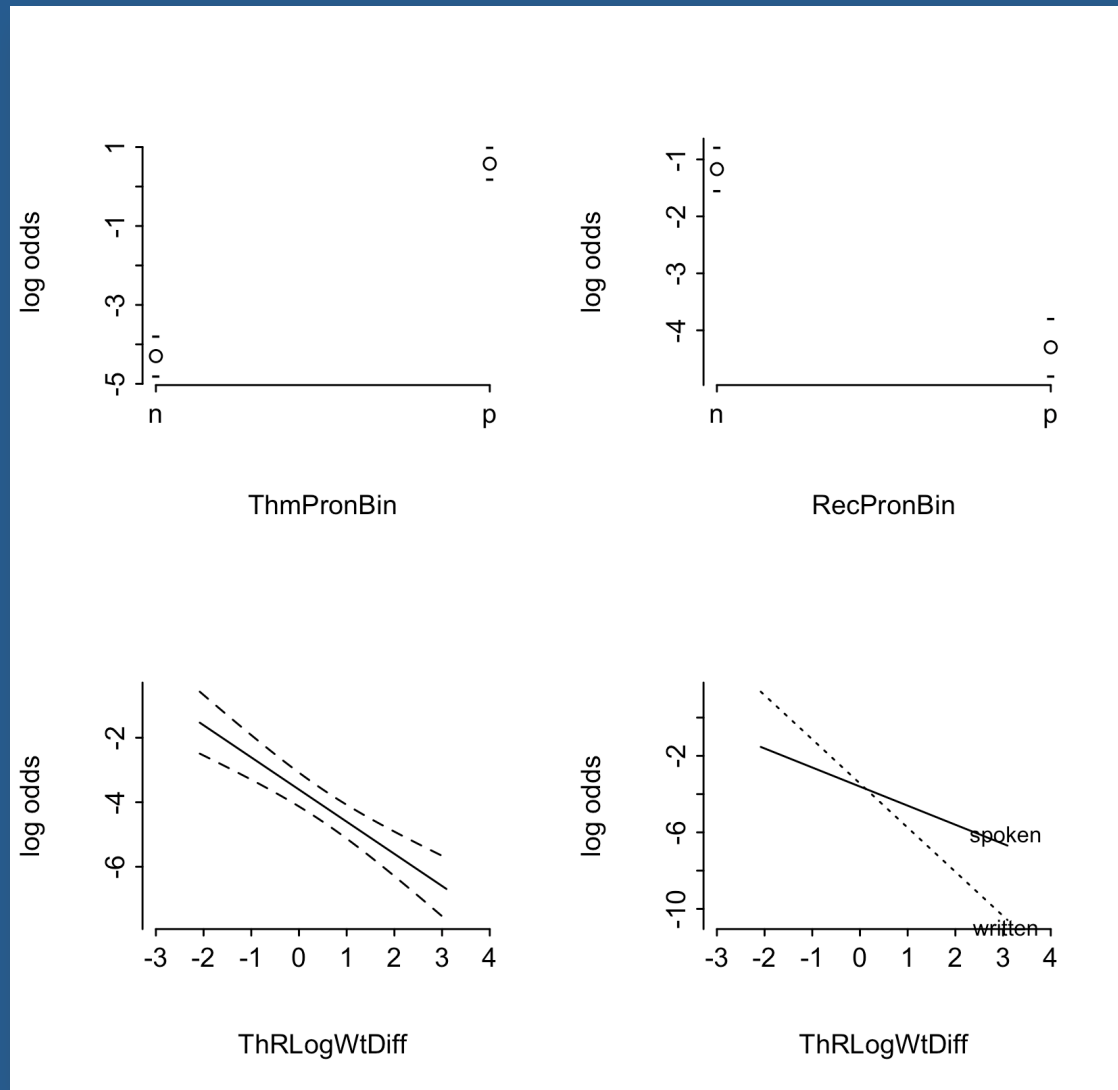
Bresnan et al. + AAE

logistic regression model

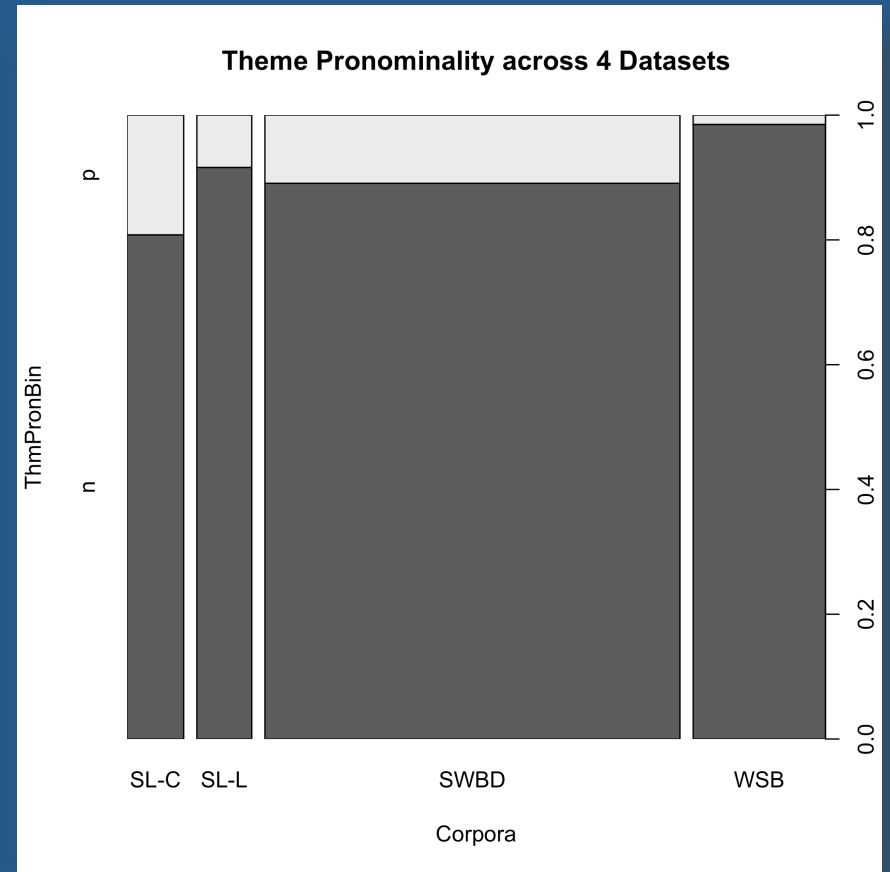
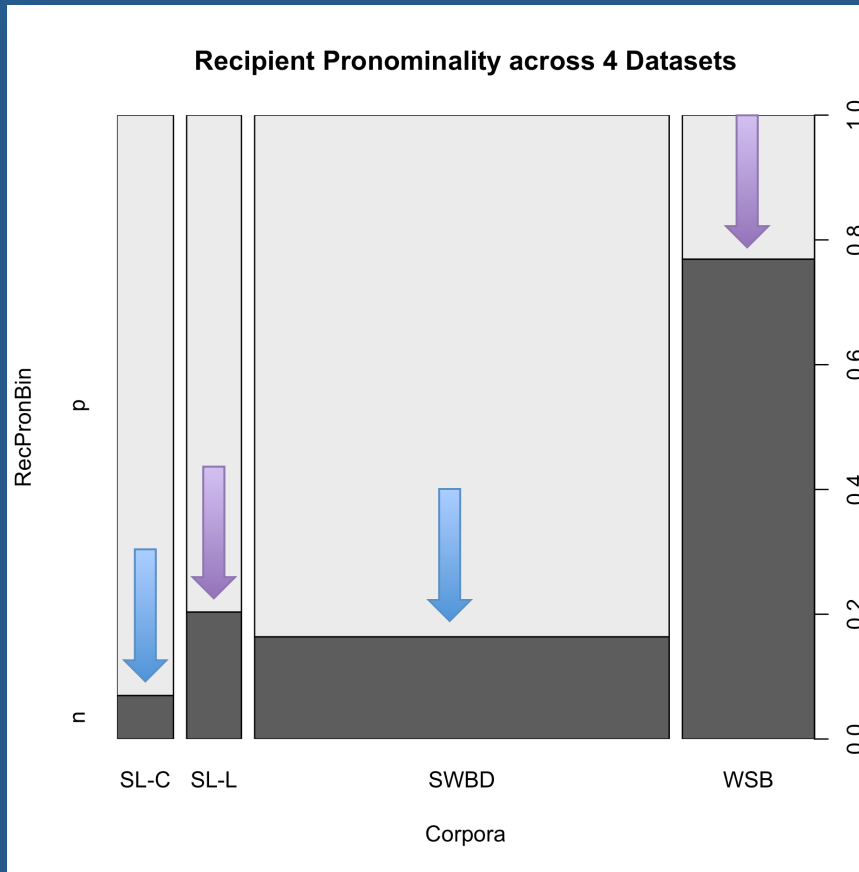
Wald Statistics

| Factor | χ^2 | d.f. | <i>p</i> |
|---|----------|------|----------|
| Recipient Pronominality (RecPronBin) | 110.06 | 1 | <0.0001 |
| Theme Pronominality (ThmPronBin) | 237.67 | 1 | <0.0001 |
| Theme - Recipient Log Weight Difference (ThRLogWtDiff) (Factor+Higher Order Factors) | 104.00 | 2 | <0.0001 |
| All Interactions | 16.19 | 1 | 0.0001 |
| Modality (Factor+Higher Order Factors) | 16.20 | 2 | 0.0003 |
| All Interactions | 16.19 | 1 | 0.0001 |
| ThRLogWtDiff * Modality (Factor+Higher Order Factors) | 16.19 | 1 | 0.0001 |
| Total | 366.42 | 5 | <0.0001 |

Model main effects & interaction



Differences between AAE data and SAE data



In closing

- We do not see here any evidence that the alternation in African American English is substantially different than it is in the macro-regional standard of US English
 - E.g., corpus and language variety do not obtain significance in statistical models
- The dative alternation does not seem to be a *sociolinguistic indicator* (Labov 1972b) at least for AAE
 - If it does become socially meaningful, we will then expect the dative alternation to show different patterns within AAE and the macro-regional AE

Thank you!

- We are extremely grateful to the following researchers for generously sharing their data with us for this project:
 - John Rickford, Shana Poplack, Valerie Fridland, Natalie Schilling, Kirk Hazen, Erik Thomas, Walt Wolfram, Christine Mallinson
- This material is based in part upon work supported by the National Science Foundation under Grant Number IIS-0624345 to Stanford University for the research project “The Dynamics of Probabilistic Grammar” (PI Joan Bresnan)
- Data for this project also come from a number of other projects supported by the NSF as well as the William C. Friday Endowment at North Carolina State University and the Sociolinguistics Lab at the University of Ottawa
- We also thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for support
- Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other funding agencies

References

- Bresnan, Joan and Marilyn Ford. 2009 ms. Predicting syntax: Processing dative constructions in American and Australian varieties of English.
- Bresnan, Joan and Jennifer Hay. 2008. Gradient grammar: An effect of animacy on the syntax of *give* in New Zealand and American English. *Lingua* 118: 245-59.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the dative alternation. In G. Bouma, I. Kraemer, and J. Zwarts (eds.), *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science: Amsterdam. 69-94.
- Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
- Godfrey, J., Holliman, E., and McDaniel, J. 1992. Telephone speech corpus for research and development. *Proceedings of ICASSP-92*. 517-20.
- Green, Lisa J. 2002. *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Gries, Stefan Th. 2003. Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1: 1-27.
- Gries, Stefan Th. 2005. Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34.4: 365-99.
- Grimm, Scott and Joan Bresnan. 2009. Spatiotemporal variation in the dative alternation: A study of four corpora of British and American English. Third International Conference on Grammar and Corpora. Mannheim, Germany.
- Harrell, Frank Jr. 2008. Design: Design package. R package version 2.1-2. [Software: <http://biostat.mc.vanderbilt.edu/s/Design>]
- Hoffman, Sebastian. 2007. Processing Internet-derived text: Creating a corpus of Usenet messages. *Literary and Linguistic Computing* 22.2: 151-65.
- Jurafsky, Dan. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In R. Bod, J. Hay, and S. Jannedy (eds.), *Probabilistic Linguistics*. Cambridge, MA: MIT Press. 39-95.
- Kendall, Tyler. 2007. The Sociolinguistic Archive and Analysis Project: Empowering the sociolinguistic archive. *Penn Working Papers in Linguistics* 13.2: 15-26.
- Kendall, Tyler. 2008. On the history and future of sociolinguistic data. *Language and Linguistics Compass* 2.2: 332-51.
- Labov, William. 1972a. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania.
- Labov, William. 1972b. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania.
- Marcus, Mitchell. Santorini, Beatrice, and Marcinkiewicz, Mary Ann. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19.2: 313-30.
- Mukherjee, Joybrato and Sebastian Hoffman. 2006. Describing verb-complementation profiles of New Englishes: A pilot study of Indian English. *English World-Wide* 27.2: 147-73.
- Olavarría de Ersson, Eugenia and Philip Shaw. 2003. Verb complementation patterns in Indian Standard English. *English World-Wide* 24.2: 137-61.
- R Development Core Team. 2009. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. [Software: <http://www.R-project.org/>]
- Schneider, Edgar. 1996. *Focus on the USA*. Philadelphia: John Benjamins.
- Van Herk, Gerard and Shana Poplack. 2003. Rewriting the past: Bare verbs in the Ottawa Repository of Early African American Correspondence. *Journal of Pidgin and Creole Languages* 18.2: 231-66.
- Van Herk, Gerard and James Walker. 2005. *S* marks the spot? Regional variation and early African American correspondence. *Language Variation and Change* 17.2: 113-31.
- Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In W. P. Lehman and Y. Malkiel (eds.), *Directions for Historical Linguistics*. Austin: University of Texas. 95-188.
- Wolfram, Walt. 1969. *A Sociolinguistic Description of Detroit Negro Speech*. Washington, D.C.: Center for Applied Linguistics.