

“Toto, I’ve a feeli[n] we’re no[r] in Kansas anymore”

Phonological variation in real and imaginary worlds



David Bowie
University of Alaska Anchorage
david.bowie@alaska.edu
Twitter @sociolx

Hriana Bowie
Polaris K-12 School
hrianamaie@gmail.com



To what extent does phonological variation occur in even scripted speech?
What sorts of effects can issues such as setting have?

ING and T/D in *The Wizard of Oz*

A number of studies have looked at the sociolinguistics of language use in movies during the past few years. Most of these have looked at discourse-level phenomena (e.g., gender-directed compliments and insults, amount of speech allocated to characters by gender and age), but we wanted to see whether there is anything to find at the level of phonological variation. To that end, we conducted a pilot study of ING and T/D variation in the 1939 movie *The Wizard of Oz*, in part because the same actors play multiple roles in the film.

The speakers

We tracked all of the actors who appear in both the Kansas and Oz settings in *The Wizard of Oz*. This means that we looked at these actors in these roles (in order of appearance):

Actor	Kansas	Oz
▪ Judy Garland	Dorothy Gayle (in both settings)	
▪ Bert Lahr	Zeke	Cowardly Lion
▪ Ray Bolger	Hunk	Scarecrow
▪ Jack Haley	Hickory	Tin Man
▪ Margaret Hamilton	Almira Gulch	Wicked Witch of the West
▪ Frank Morgan	Professor Marvel	Gatekeeper Carriage Driver Guard Wizard of Oz

The variables

We tracked ING and T/D for all actors appearing in both the Kansas and Oz settings. ING is the word-final sequence consistently spelled <ing> and variably produced with, most commonly, a [ŋ] (the citation form) or [n]. T/D involves the retention or deletion of [t] or [d] at the end of word-final consonant clusters, and so involves simplification. This study did not investigate deletion of word-final [t] or [d] in postvocalic contexts, but some was observed in the process of acquiring data, which points toward one possible future expansion of this line of inquiry.

These are both widely studied linguistic variables, and so we tracked the usual independent variables for each: For ING, grammatical category; and for T/D, preceding sound, following sound, and grammatical category. In addition, for each token, the speaker in terms of both the actor and the role were noted, as well as the setting.

The usual practices for excluding tokens of these variables from analysis were followed (e.g., tokens of T/D preceding [t] or [d] were not used), and in addition all tokens of both ING and T/D that occurred in songs were excluded, since they were purposefully manipulated at times to create rhymes, and the one instance of T/D that occurred during the tornado sequence was not included, since it was unclear whether it should be placed in the Kansas or Oz setting.

In all, 128 tokens of ING and 287 of T/D were collected. Binomial regressions for ING and T/D were run using the `glm` function in R. For those factor groups with significant variables, odds ratios were calculated to gauge the strength of the effect.

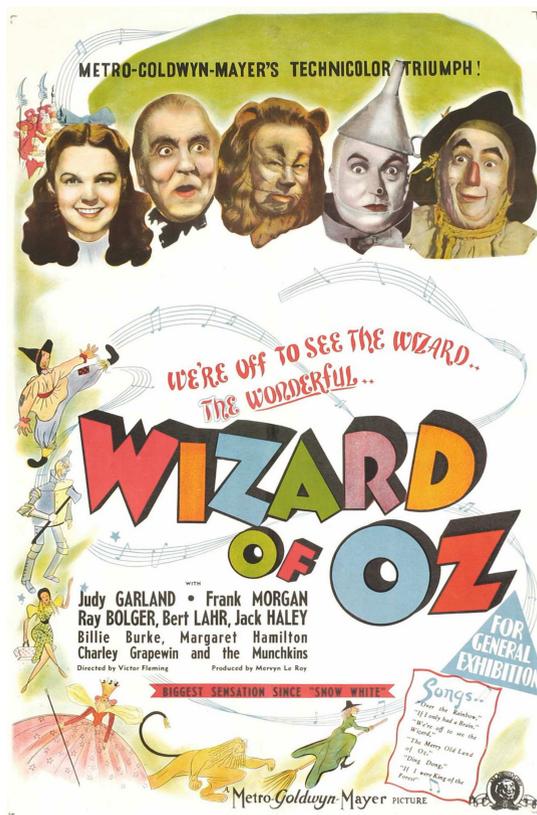
Some thoughts on movies as data sources

Movies are, of course, generally not created with the intent of being used for linguistic analysis. As such, they have the benefits and drawbacks common to any similarly situated data source—that is, their content is less predictable and, often, less comprehensive than data from sociolinguistic interviews, but they sidestep the problem of interviewers possibly changing the linguistic situation simply through their presence.

Movies, though, present some particular challenges. To begin with, they are scripted, which means that what researchers are faced with will be different from naturalistic speech in a number of ways; one concrete example from *The Wizard of Oz* is that there were no filled pauses after any instances of T/D in the entire movie, which seems unlikely to happen in real life. In addition, though there was no dialect coaching or direction for this movie, at the very least Bert Lahr’s portrayal of the Cowardly Lion and Frank Morgan’s of the Carriage Driver involved some conscious dialect manipulation—but exactly what they were aiming for, and whether other actors had similar aims is unknowable. That said, we find that there are still coherent sociolinguistic patterns in the data from this movie, and would expect to find coherent patterns in other movies.

In the *Wizard of Oz*, as I’m sure you all know, there is a trio of farmhands in Kansas that become the Scarecrow, Tin Man, and Cowardly Lion in Dorothy’s dream. Why do three guys from Kansas talk like they’re from Brooklyn?

—CookingWithGas, Straight Dope Message Board, 1 March 2006



What comes next?

This is just an initial stab at things—we wanted to see if there’s the possibility of anything worth looking into. Since it looks like this might be something worth pursuing, we’ve developed a few ideas for next steps:

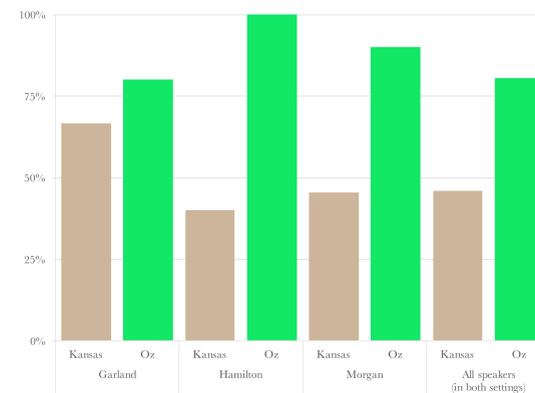
- Looking at a wider variety of movies in which the same actors appear in realistic and fantastic settings (e.g., *The Chronicles of Narnia* movies, *Enchanted*, the *Harry Potter* movies, *Labyrinth*, *MirrorMask*, *Peter Pan*)
- Tracking the same actors (e.g., Judy Garland, Frank Morgan) in different movies to determine whether the differences seen here are actually attributable to setting, or may really be due to different roles
- Following actors who reprise the same role in different modes (e.g., movie, television, stage production)
- Expanding the T/D data so that issues such as flapping and post-vocalic deletion (and possibly word-medial NT simplification) can be included in the analysis
- Developing a hierarchical model so that possible confounders (e.g., characterization and setting) can be dealt with more elegantly

And, of course, anything you can think of. That’s a large part of why we’re doing a poster, after all—to find out what ideas everyone else can give us!

Patterns of ING

Binomial regression selected *only* setting as significant. Given previous study of this variable we expected grammatical category to be selected as significant, but it wasn’t; it is unclear whether the insignificance of this variable was the result of the language used being scripted, or simply because the vast majority (~80%) of ING tokens occurred in progressive verbs. Either way, the result drives home that we aren’t dealing with truly naturalistic language here, and so truly naturalistic patterns shouldn’t be expected.

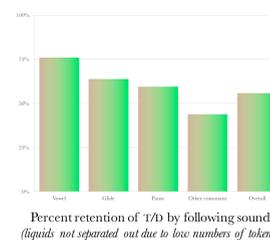
Across all actors appearing in both the Kansas and Oz settings, the citation form (i.e., [ŋ] rather than [n]) was more preferred in Oz, and this was a large-sized effect (OR=4.88).



Percent citation forms (i.e., [ŋ]) for ING, progressive verbs only, produced in Kansas and Oz settings for actors with at least 5 tokens in each, plus all actors appearing in both settings (not statistically significant for Garland, significant for all others)

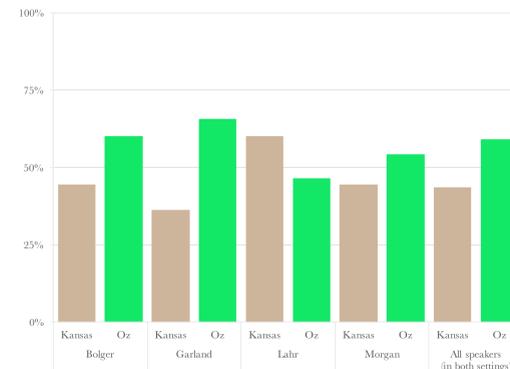
Patterns of T/D

Binomial regression selected both following sound and setting as significant, but no other factors.



Effect sizes for all categories of following sound with enough tokens to test were weak ($0.3 < OR < 3.00$). Also, as a further reminder that we aren’t dealing with fully naturalistic speech here, neither preceding sound nor grammatical category were selected as significant, and as for the distribution of following sounds, there were no filled pauses following any tokens of T/D.

Across all actors appearing in both the Kansas and Oz settings, the citation form (i.e., retaining [t] or [d]) was more preferred in Oz, and this was a medium-sized effect (OR=2.12).



Percent retention for T/D produced in Kansas and Oz settings for actors with at least 5 tokens in each, plus all actors appearing in both settings (not statistically significant for Lahr or Morgan, significant for all others)