381: Topics in Sociolinguistics and CMC

Lecture 5: Style, register, and genre
July 8, 2019
Online written language as a sort of hybrid...at least for English (Tagliamonte and Denis 2008:7).

Ferrara et al. (1991): “a language variety that never existed before”.


Crystal (2001): “a new species of communication”.

Baron and Ling (2003): “the linguistic centaur”.

Baron (2004:413): “a blend of both spoken and written language conventions”.

Tagliamonte and Denis (2008:24): “a vibrant new medium of communication with its own unique style”. 2
Brief recap:

**Style-shifting** is about what individuals do as they move through different situations/contexts/etc.; it is like how formal someone is being, but it is not at all unidimensional.

A **register** is a use of language agreed to be appropriate for a particular purpose/location (e.g. newspaper headlines, personal ads, texting).
style, register, genre

Humor based on a clash of registers:
Let’s start by examining the style-shifting that emerges from the same users but *across registers of written English, both offline and online.*
Excerpt from *xkcd*, ‘Writing Skills’:

WEIRD—
ANOTHER STUDY FOUND THAT KIDS WHO USE SMS ABBREVIATIONS ACTUALLY SCORE HIGHER ON GRAMMAR AND SPELLING TESTS.

WHY ON EARTH IS THAT A SURPRISE?

IMAGINE KIDS SUDDENLY START PLAYING CATCH LITERALLY ALL THE TIME. EVERYWHERE THEY GO, THEY THROW BALLS BACK AND FORTH, TOSS THEM IN THE AIR, AND HURL THEM AT TREES AND SIGNS—NEARLY EVERY WAKING HOUR OF THEIR LIVES.
Tagliamonte (2016)’s study is essentially the variationist equivalent of this comic’s premise.

She investigates different registers associated with different types of CMC as of 2009-10.

She collected examples of students’ writing in four contexts: in formal writing for a class and “in three internet registers: email, instant messaging, and texting on phones” (2016:7).
The features of CMC that are decried by prescriptivists (acronyms, respellings, abbreviations, etc.) are not found in formal writing (2016:13-14).

Intensification (of adjectives by overt adverbs) is found at rates of about 30% in all of the CMC registers, and a rate of about 10% in formal writing (2016:20).

Remember, these are results from the same set of people.
Tagliamonte 2012:20-21: The only intensifier found in formal writing is the old standard very.
"[T]he character of [several variants’] use is systematically patterned according to register. [Email] is the most formal and the most like the written essays. It has the longest turns and the lowest frequency of CMC forms. EM also has the lowest frequency of [the informal variants] intensifier so and future going to" (Tagliamonte 2016:27).

Even though the participants are talking to friends, the effect of register is clear and consistent.

Instant messaging (on a computer) and texting (on phones) are less distinguishable but both are very different from email (2016:27).
No one’s grammar is broken; “these young people are fluidly navigating a complex range of new written registers and are using conventions that are particular to each one” (2016:28).

Young people write essays as they always did, and now they also use casual written language.
Tagliamonte (2016)’s study shows what a subtle palette of registers users have access to.

This should seem pretty intuitive to sociolinguists, but it’s a crucial point that the prescriptivists miss:

People use language in more than one context, and always have. Different registers of language are appropriate for different situations and levels of formality.

There is nothing new here. What CMC introduces is merely seeing more of a range of it in writing.
style, register, genre

That was in 2009-10.

What has changed since? What might we expect to be different if we did a similar study on undergraduate students in 2019-2020?
Now let’s turn our attention to a subtler question: what happens on the micro-level in terms of style-shifting, with platform held constant, and how can we use big corpora to study the effect of different addressees?

Enter Pavalanathan and Eisenstein (2015), a study of style on Twitter.
Quick overview of approaches to style (see also Pavalanathan and Eisenstein 2015:190):

**Attention-to-speech model** (Labov 1966, etc.):

- Minimal pairs
- Reading off word lists
- Reading out of a book
- Careful speech
- Conversational speech

more vernacular
The attention-to-speech model proposes that people use standard variants more in formal contexts because of self-monitoring: paying more attention to how they sound. When people are less self-conscious, their speech is closer to the vernacular.

This is a straightforward approach and there are certainly elements of it in style-shifting, but it assumes that style is one-dimensional (a single line from formal to casual).

It also does not truly capture language as it is used: we do not often read off word lists or minimal pairs.
One of the approaches newer than Labov’s is **audience design** (Bell 1984, but see also Coupland 1980, Baugh 1983, Rickford and McNair-Knox 1994).

This model of style-shifting is related to accommodation and proposes that people style-shift in order to **shape their relationship with their audience**.

Offline, there are so many types of audiences: interlocutors (other speakers), passive audiences, eavesdroppers, people judging a speaker’s linguistic behavior, etc.

**What about online?**
Pavalanathan and Eisenstein (2015:188): “[S]ocial media data have more to offer sociolinguistics than size alone: even though platforms such as Twitter are completely public, they **capture language use in natural contexts with real social stakes**. These platforms play host to a diverse array of interactional situations, from high school gossip to political debate, from career networking to intense music fandom. As such, social media data offer new possibilities for understanding the social nature of language: not only who says what, but also how stylistic variables are perceived by readers and writers and how they are used to achieve communicative goals.”
Pavalanathan and Eisenstein (2015:190):

“This question of ‘which audience’ is particularly salient in the context of publicly readable social media: in principle, any of the millions of users of Twitter could witness the conversation, though authors may be aware of the specific preferences and interests of their own follower networks (Marwick and boyd 2011).”
Pavalanathan and Eisenstein (2015:194) classify tweets as follows and use the categorization as a proxy for size of intended audience:

1. “@-mentions: posts with a person tagged, e.g. @inthelemonlight (smallest audience assumed)

2. “hashtags”: posts with a hashtag e.g. #linguadork (medium audience assumed)

3. “broadcasts”: neither hashtags nor @-mentions (largest audience assumed)

They’re curious about whether size of intended audience correlates with use of **nonstandard features**.
Note: the authors employ the term “variable” in a very loose way that does not match standard variationist practice. That said...

Two types of nonstandard features:

1. The **geographically** distinctive nonstandard: culled from **geotagged** tweets per large metropolitan area in the U.S., removing proper nouns, then finding the 30 most common strings per each city.

Mix of **words, initialisms, abbreviations, emoticons, punctuation**, etc.
2. The more universally nonstandard on Twitter (in the English-speaking U.S., but not specific to a smaller geographical area). Generated by finding the 1,000 most frequent strings and removing the ones that are standard words, or symbols, or hashtags, or proper nouns, or non-English (Pavalanathan and Eisenstein 2015:198).
So now there are two sets of features:

a) nonstandard and **localized**;

b) nonstandard and **not** localized.

Data: Starting with 114 million geotagged tweets from 2009 to 2012, ignoring retweets, URLs, and accounts with large numbers of followers. Controlled for same number of messages from each of more than 2 millions users (Pavalanathan and Eisenstein 2015:199).

Location of user: scraped from **3 or more distinct people mentioning of the place name in conjunction with the username, and no one mentioning a different place name alongside them**. Intentionally stringent!
Note the very clear hypotheses! This is excellent practice and is to be encouraged.

**Hypothesis 1:** “nonstandard lexical variables are used more often in messages that target a limited audience” (2015:202).

**Hypothesis 2:** “geographically specific linguistic variables are used more frequently in messages targeted at individuals from the same geographical area” (2015:204).
Results:

Hypothesis 1: **Yes**: nonstandard features “are used more frequently in interactional contexts with more specifically directed audiences” (2015:203).

Hypothesis 2: **Yes**: “nonstandard lexical [forms] are especially likely to be used in messages that mention individuals from the same metropolitan area as the sender” (2015:204).

Overall: “as the audience becomes smaller and more local, nonstandard [features] are increasingly likely to be used; as the audience becomes larger and less directly connected to the author, the frequency of the [nonstandard forms] decreases” (2015:205).

Universal nonstandard features are also used more between people from the same community. How do we account for this? Do they have local indexical value of ‘youngness’ or ‘informality’? (2015:206).

What do you think?
“[N]onstandard [features] are most likely to be used in precisely those settings that are most similar to face-to-face communication: conversational messages aimed at narrow, local audiences” (Pavalanathan and Eisenstein 2015:206).

How does this go hand-in-hand with the findings of Tagliamonte (2016) from 2009-2010?
Now let’s shine a light on the matter of genre.
Relevant quotation from graphic novelist Raina Telgemeier’s website (in the FAQ section):

“What genre are your books?

Please note that graphic novels are a format and a medium, but not a genre. There are many genres WITHIN graphic novels! *Smile* and *Sisters* are considered autobiography and memoir, and sometimes non-fiction. *Drama* and the BSC graphic novels are considered realistic fiction and slice of life. *Ghosts* is realistic fiction with a supernatural flavor.”
Likewise, we could say that **YouTube videos or streams** are a format/medium, but not a genre.

One YouTuber’s videos might span several genres, or be difficult to classify.

**What kinds of genres might we find among them these days?**
style, register, genre

Personal blogs, diaries, travel videos, family updates
Instructional/educational videos
Comedy
Music videos
Animation
Sports
Let’s Play, recreational gaming, game reviews
Teaching/lectures
Food/cooking
Culture/entertainment
Political commentary/news
Visual art

(partially adapted from Burgess and Green 2018)
Some might be culturally mediated.

*Mukbang* is a Korean genre of someone *eating* on camera (see Choe 2019 for a conversation-analysis perspective) while onlookers watch.
Frobenius (2011): on the conventions surrounding the “opening sequences of video blogs”.

Let’s look at some examples!

Note: apologies for assigning an article with a totally needless ableist slur – editor should have caught it, but we should have noticed too.
As we saw in Lecture 3, most bloggers do not know their subscribers: the relationship tends to be parasocial (Horton and Wohl 1956; Frobenius 2011:815) the way it is for a celebrity or a fictional character.
Given that, how does an experienced YouTuber build connections to their audience?

This paper is a look at a sub-question of that: since any “assumed audience presumably plays an important role in vlogs, especially if there is a greeting and terms of address” (Frobenius 2011:817), what do openings of YouTube videos look like, and why?

What kind of thing are they? Do they resemble e.g. news broadcasting on TV? Answering-machine messages?
“Since viewers can also decide to stop watching the vlog at any moment, one principal function might be to raise the viewers’ interest and induce them to watch the whole vlog” (Frobenius 2011:817). Do you agree? Why or why not?
“What is interesting with regard to openings in general is that some vloggers try to conventionalize them and turn them into routines. Others completely deny a need for an introductory phase, so that they do not establish personalized routine formulae” (2011:825).

Can you think of examples of each?

Is this still as true as it was of the time of writing?

What would each strategy convey to you?
Next class:
Thursday, July 11

Preview the Lecture 6 readings and read one or more of them. Post two discussion questions on Orbund. Homework 3 is due on Monday.