

Language and its Applications

LT5903



Jixing Li

Lecture 11: Sociolinguistics

Lecture plan

- Computational Linguistics review
- Language varieties
- Variation at different linguistic levels
- Factors influencing variation
 - regional and geographic
 - social
- Short break (15 mins)
- Group discussion on HW11

Computational Linguistics review

computational linguistics (CL): employs computational methods to understand **properties of human language**.

natural language processing (NLP): aims to develop methods for solving **practical problems** involving language.

NLP tasks: information extraction, automatic speech recognition, machine translation, sentiment analysis, question answering, and summarization.

Topics:

tokenization: segmenting words/sentences

POS tagging: assign a part-of-speech tag to each word in a sentence

CFG and parsing: recognize and parse syntactic structures of sentences:
bottom-up/top-down parsing

word embeddings: represent word meaning as vectors: words with related meanings have higher cosine similarities

Sociolinguistics

sociolinguistics: the study of the relationship between language varieties and social structure as well as the interrelationships among different language varieties.

language variety: any form of language characterized by systematic features.

dialect: any variety of a language spoken by a group of people that is characterized by systematic differences from other varieties of the same language in terms of **structural** or **lexical** features.

Standard dialect

Linguistically speaking, no one dialect or language is better, more correct, more systematic, or more logical than any other

standard dialect: the variety used by political leaders, the media, and speakers from higher socioeconomic classes. It is also generally the variety taught in schools and to non-native speakers in language classes.

prestige: Socially speaking, a standard dialect is a dialect of prestige and power.

*Kim and **I** went to the store.
This is a matter between Kim and **I**.
Give the books to Kim and **I**.*

→ violations of a prescriptive rule may be perceived as standard if they are used by members of the prestige group.

Standard dialect

Standard

I like *myself*

You like *yourself*

He likes *himself*

She likes *herself*

We like *ourselves*

You like *yourselves*

They like *themselves*

Nonstandard

I like *myself*

You like *yourself*

He likes *hissself*

She likes *herself*

We like *ourselves*

You like *yourselves*

They like *theirselves*

→ The non-standard variety is not any less systematic or less sensible than the standard variety

Phonetic and phonological variation

Phonetical variation

phonetics: how sounds are physically produced, the acoustics of sound waves, the perception of sounds

/t,d,n,s,z/: New York City dialect produces these sounds as **dental**, while other American English dialects produce these sounds as **alveolar**.

phonetic variation in different **styles** or **registers** of speech:

little: [lɪrl] or [lɪtəl]

Phonological variation

phonology: the system of sounds in a language
caught v.s. cot

/kɔt/ and /kat/ → minimal pairs

/kat/ and /kat/ → homophones in some dialects

Morphological variation

morphemes: the smallest unit of sound-meaning pairs in a language

-ish: foolish, childish, reddish

-self/-selves

Standard American English (SAE): *tree-like, about now | himself, themselves*

dialects: *tree-ish, now-ish | hisself, theirselves*

past tense:

dive → dived / dove

drag → dragged / drug

Syntactic variation

syntax: the way words are put together to form phrases and sentences

done v.s have

modal combinations

SAE: *She has already told you. | might be able to, might perhaps*

Southern dialects: *She done already told you. | might could, might would*

need + verb

SAE: *The corps need to be watered.*

Midwestern dialects: *The corps need watered.*

Lexical variation

differences in the words people use to mean the same thing, or
differences in what the same word means

soda: Northeast, West

pop: Midwest, Northwest

coke: South

gentleman: formal

guy/dude: informal

to be pissed

American English: to be mad

British English: to be drunk

Regional and geographic factors

Geography plays a large role in determining a dialect: People who live close to each other will have considerably more influence on each other's dialects.



Regional dialect boundaries often coincide with natural boundaries such as rivers, mountains, or swamps.

New England

area: western New York (except New York City, which is a distinct speech island), eastern Massachusetts, Connecticut, Vermont, New Hampshire, Rhode Island, and Maine

phonetic: /r/ before a consonant or at word final is not pronounced
park your car → [p^hɑ:k jə k^hɑ:]

phonological: *caught* and *cot* are homophones /kat/

syntactic: *so do I* → *so don't I*

lexical: *water fountain* → *bubbler*
soda



The South

area: Texas, Louisiana, Arkansas, Mississippi, Alabama, Georgia, Tennessee, Kentucky, Virginia, south-ern Ohio, Illinois, Indiana, and Missouri, and most of Oklahoma, West Virginia, and North and South Carolina. (Note that it excludes Florida, which is considered a separate speech island.)

phonetic: /ɛ/ → /ɛɪ/
led [lɛɪd] *net* [nɛɪt]

phonological: *pin/pen merger:* /ɛ/ → [ɪ] / _ [n]

syntactic: *double modal*
I might could help you clean your house tomorrow.

lexical: *coke*

Socioeconomic variation

socioeconomic status is correlated with language variety

rhotic pronunciation → prestige

non-rhotic pronunciation → lower socioeconomic status

/ɹ/ at the end of a syllable is usually dropped

Percentage of [ɹ] in *floor*

	Casual	Careful
Saks	63	64
Macy's	44	61
S. Klein	8	18

Age variation

Some relatively recent changes in English:

hopefully as a sentential adverb:
Hopefully it won't rain tomorrow.

like as quotative:
He was like, "Well, I don't know either."

Gender variation

In western cultures, women tend to use more prestige variants than men.

In Norwich, England, women are more likely to use the standard form *running* than the non-standard form *runnin'*

In Sydney, Australia, boys are more likely to use non-standard syntactic forms such as double negation: *they don't say nothing*.

Hypotheses:

1. Women have traditionally been considered inferior to men in terms of social status, so they may make more of an effort to be perceived as more prestigious.
2. Women are the primary caretakers for children and they want to expose their children to prestige dialects

Ethnic variation

Ethnic variations often represent an ethnic group's heritage and culture.

African American English (AAE):

phonological:

- diphthongs → monophthongs / _# or _C[+voice]
now [na:] *side* [sa:d]
- word-final consonant cluster reduction
best day [bɛs deɪ]

syntactic:

- multiple negation: *It doesn't make no sense.*
- copula absence: *John going to the store.*
- habitual *be*: *The coffee be cold (The coffee is always cold).*

To do

Do HW11

Read:

This lecture: **File** Ch10