# Think 53: Food Talks Dan Jurafsky & Yoshiko Matsumoto

## How to Read a Menu

How to do "Linguistics of the everyday"

Do social science by studying aspects of everyday life

Computational social science if the datasets are big

## Goal for today

Understand the methods that we used for studying menus

With the idea that some of you will want to study something like menus (or other texts) for your final projects

How do menus reflect socio-economic differences?







What linguistic differences are reflected in cheap versus expensive menus?

Menus as reflections of attitudes toward socioeconomic class

Online menus from 6562 restaurants

- 650,000 menu items
- 5,000,000 words
  - allmenus.com, yelp.com

### Datasets

# http://www.allmenus.com/

## Yelp Challenge Dataset

#### The Challenge Dataset:

- 2.2M reviews and 591K tips by 552K users for 77K businesses
- 566K business attributes, e.g., hours, parking availability, ambience.
- Social network of 552K users for a total of 3.5M social edges.
- Aggregated check-ins over time for each of the 77K businesses
- 200,000 pictures from the included businesses

#### Cities:

- U.K.: Edinburgh
- Germany: Karlsruhe
- Canada: Montreal and Waterloo
- U.S.: Pittsburgh, Charlotte, Urbana-Champaign, Phoenix, Las Vegas, Madison

**Get the Data** 

Johnny Rockets	fast food	bos	ŚŚ	bacon lettuce and tomato crisp thick bacon fresh lettuce tomato and mayonnaise on your choice of bread	5.49
Johnny				chicken club sandwich grilled marinated chicken breast with thick bacon fresh lettuce tomato and mayonnaise on triple decker	
Rockets	fast food	bos	\$\$	sourdough	6.59
Johnny Rockets	fast food	bos	\$\$	egg salad sandwich made from scratch just like mom's served with fresh lettuce and mayonnaise on your choice of bread	4.59
Johnny Rockets	fast food	bos	\$\$	chili size served open faced with our exclusive all meat chilli and topped with fresh grated cheddar cheese and chopped onions	6.39
Thai Chef & Sushi Bar	sushi	dc		sweet corn soup delicious creamy corn soup with minced chicken finished with egg	4.25
Thai Chef & Sushi Bar	sushi	dc		chu chee shrimp grilled shrimps topped with chef's spicy chu chee sauce served with steamed vegetables	
	italian	sf	\$\$\$\$		9
Quince	Italiali	51	<b>\$</b> \$\$\$		9
Quince	italian	sf	\$\$\$\$	woodside farm jerusalem artichoke soup homemade pancakes and thyme	8
				composed cheese plates chef s selection of three served with warm cranberry nut bread and house made accompaniments	
L'Impero	italian	nyc	\$\$\$\$	supplement with prix fixe menu	14
Dana John's				the works x large sink your teeth into this pizza loaded with a mouthwatering combination of pepperoni ham spicy italian sausage fresh sliced onions green peppers gourmet baby	
Papa John's Pizza	pizza	nyc	\$	portabella mushrooms and ripe black olives	18.99
Joe's Shanghai	chinese			prawns w exotic sea scallop sauce	20.95

## How to test hypotheses

- Form hypotheses about interesting classes of words – create word lists
- 2. Count to see which restaurants these words are associated with
- 3. Confirm the hypotheses by running statistics
  - Linear regression
  - Logistic regression
    - Ordinal logit regression

### Choice

Robin Lakoff (2006). Identity a la carte: you are what you eat. In De Fina, Schiffrin, and Bamberg (eds), *Discourse and Identity*. Cambridge.

# Cheap restaurants focus on diner choice

Expensive restaurants don't

# Expensive restaurant menu:

Robin Lakoff (2006). Identity a la carte: you are what you eat. In De Fina, Schiffrin, and Bamberg (eds), *Discourse and Identity*. Cambridge.

caviar

(Cru) golden eye snapper amberjack herring

abalone

trout roe

sea urchin

cardoon

brassicas

toffee

wood pigeon

winter citrus

black walnut

ice cream

tea

## Cheap restaurant menu



# Measuring more subtle aspects of choice

- Baby lamp chops grilled to your liking
- Marinated flak steak with eggs your way
- Quiche with your choice of either house salad or a cup of soup
- Biscuits and gravy with eggs anyway you like'em

# Measuring more subtle aspects of choice

 choice, choose, any, add, or, specify, substitutions, specifications, options, pick, your way, your own, your liking, your style, your favorite, you like, you want, you request, way you, you may, select your, select from, you select, select one, select any, select or, select a, select up, select two

 chef's selection, chef's selections, chef's daily selection, chef's daily preparation, chef's choice, chef's selected, chef's assortment, tasting menu, omakase, kaiseki, courses, prix fixe

# More subtle aspects of choice: who is the menu talking about?

#### Cheap restaurants:

"you"

"your way/style"

"you like/want"

### **Expensive:**

"chefs choice"

"chef's selection"



# Even more subtle differences Let's look at adjectives

# There were 1065 adjectives in the datasets

#### 165 sensory adjectives

 Chunky, crispy, crunchy, doughy, fluffy, rich, smolky, tangy, zesty

#### 176 participial adjectives

Grilled, mixed, sliced, steamed, baked, smoked,

#### 55 extreme positive sentiment words

 Amazing, awesome, exciting, extraordinary, incredible, legendary, perfect

#### 15 delicious words

 Delicious, delectable, scrumptious, luscious, tasty, gourmet, savory

## Finding adjectives

Aloo Paratha: Mashed potatoes flavored with tangy spices and herbs

Paneer do Pyaza Roll: Homemade cottage cheese and grilled onions

Masala Aloo Roll: Pan fried spiced potatoes and herbs

# Finding adjectives automatically: Part-of-Speech tagging

https://nlp.stanford.edu/software/tagger.shtml

```
Aloo_NNP Paratha_NNP:_: Mashed_JJ potatoes_NNS flavored_VBN with_IN tangy_JJ spices_NNS and_CC herbs_NNS

Paneer_NN do_VBP Pyaza_NNP Roll_NNP:_: Homemade_JJ cottage_NN cheese_NN and_CC grilled_JJ onions_NNS

Masala_NNP Aloo_NNP Roll_NNP:_: Pan_NN fried_VBD spiced_JJ potatoes_NNS and_CC herbs_NNS
```

JJ = Adjective NNP = Proper Noun NN = Noun VBD = Preterite VBN = Past Participle

### Cheap restaurants use vague "tasty" words

### Delicious, gourmet, flavorful,

The delicious taste
delicious outdoor grill flavor
gourmet veggie special
flavorful entrées
flavorful ancho-chile

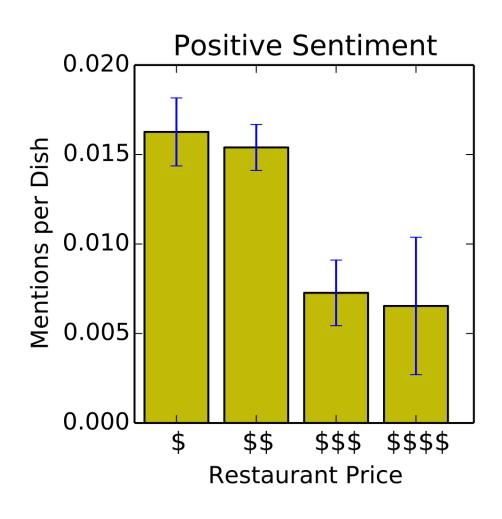


For every vague adjective the average price of a dish drops 2%

### Cheap restaurants use vague "positive" words

### Amazing, great, fantastic

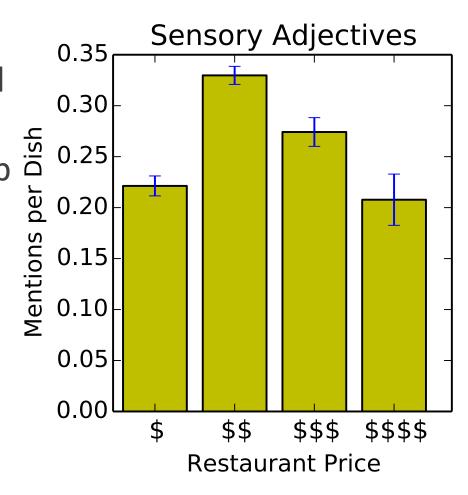
Our amazing signature subs
Amazing Spicy Chicken
Casserole
Our fantastic house made
dressing
unbelievably great tasting



### Middle priced restaurants use sensory adjectives

#### zesty, rich, golden brown, crispy, creamy

Crispy white-meat tenders served with a creamy Creole sauce rich, creamy spinach artichoke dip Creamy, homemade fettuccine alfredo zesty chili pepper cream sauce



# How do we know this is true? Control variables

name category city restdollars numitems provenance traditional generous participle positive delicious choice adj sensoryadj length wordlen

- 1 Tiffin Etc. indian philadelphia 2.0 24 0.0833 0.4167 0.0000 1.2500 0.0000 0.0000 0.0000 2.0417 0.5000 12.4583 6.5806
- 2 City Pie pizza nyc 1.0 47 0.0000 0.0000 0.0000 0.1915 0.0426 0.0000 0.2128 1.2766 0.2553 8.8085 6.2053
- 3 Wayo sushi bars sf 2.0 57 0.0000 0.0000 0.0000 0.0526 0.0000 0.0000 0.0000 0.3509 0.0877 3.5439 5.7928

## Linear regression

After controlling for the city, the total number of words, the average wordlength, the number of items on the menu...,

do cheap restaurants have more "delicious" words than expensive restaurants?

# Thinking about linear regression

Could I predict the number of dollar signs from the number of "delicious words"?

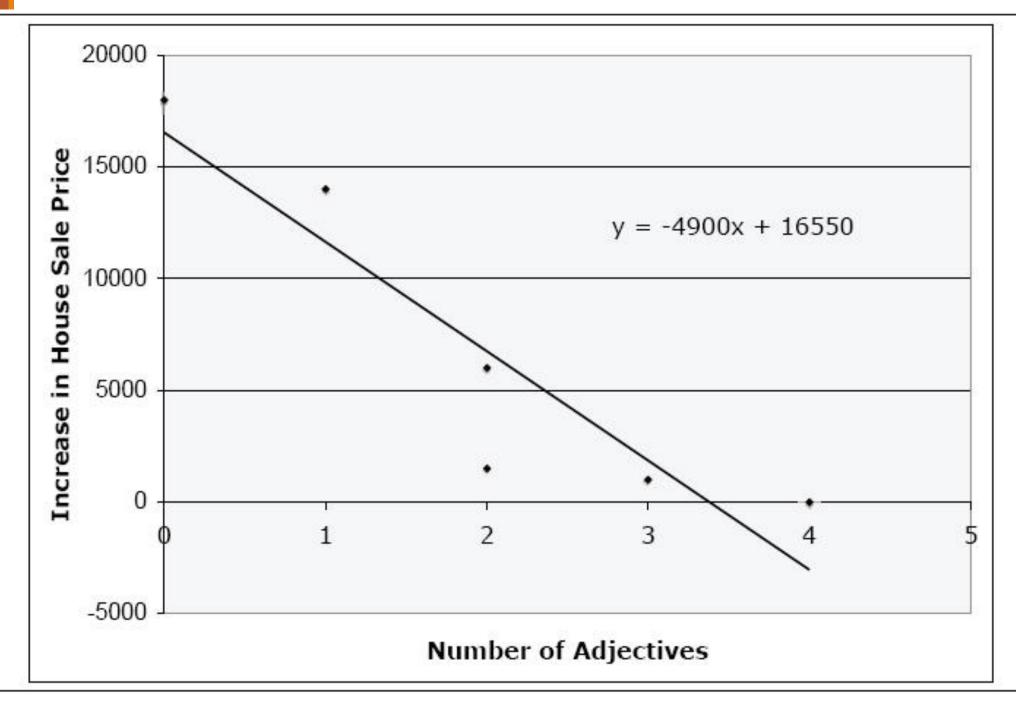
# Linear Regression Predicting Housing Prices

Example from Freakonomics (Levitt and Dubner 2005)

Vague Adjectives: Fantastic, cute, or charming

# of Vague Adjectives	Amount House Sold Over Asking Price
4	0
3	\$1000
2	\$1500
2	\$6000
1	\$14000
0	\$18000

### Linear Regression



## Muliple Linear Regression

#### Predicting values:

 $price = w_0 + w_1 * Num\_Adjectives + w_2 * Mortgage\_Rate + w_3 * Num\_Unsold\_Houses$ 

In general:

$$price = w_0 + \sum_{i=1}^{N} w_i \times f_i$$

Let's pretend an extra "intercept feature" f0 with value 1

### Multiple Linear Regression

$$y = \sum_{i=0}^{N} w_i \times f_i$$

# Linear regression predicting \$ from language, using "R"

```
lm(formula
restdollars
city + category + logwordlen +
loglength + lognumitems +
adjectives + largeportions+
choice+ provenance+ traditional)
```

	Estimate	StdError	z_value	Pr(> z )	
(Intercept)	-2.856242	0.329042	-8.68	<2e-16	***
citychicago	-0.043169	0.033967	-1.271	0.203809	
cityla	0.061897	0.042524	1.456	0.145562	
citynyc	0.166924	0.02487	6.712	2.08E-11	***
cityphiladelphia	0.018239	0.031218	0.584	0.559081	
citysf	-0.047399	0.027739	-1.709	0.087544	
citywashington	0.102401	0.033002	3.103	0.001925	**

categoryamerican_(traditional)	-0.226379	0.036237	-6.247	4.44E-10	***
categoryasian_fusion	0.155454	0.084581	1.838	0.066119	
categorybakeries	-0.53602	0.080458	-6.662	2.92E-11	***
categorybarbeque	0.122788	0.095623	1.284	0.199156	
categorybars	-0.210265	0.041823	-5.028	5.10E-07	***
categorychinese	-0.266647	0.040417	-6.597	4.52E-11	***
categorycoffee_&_tea	-0.573856	0.054413	-10.546	<2e-16	***
categorydiners	-0.290413	0.061352	-4.734	2.25E-06	***
categoryethiopian	-0.143439	0.122629	-1.17	0.242165	
categoryfast_food	-0.549233	0.047937	-11.457	<2e-16	***
categoryfrench	0.331963	0.046377	7.158	9.10E-13	***
categorygreek	-0.287144	0.134892	-2.129	0.033317	*
categoryindian	-0.319241	0.049826	-6.407	1.59E-10	***
categoryitalian	0.003328	0.034616	0.096	0.923404	
categorykorean	-0.070347	0.081356	-0.865	0.387249	
categorylatin_american	-0.332607	0.050576	-6.576	5.20E-11	***

categorymexican	-0.489798	0.038614	-12.684	<2e-16	***
categorymiddle_eastern	-0.426998	0.056383	-7.573	4.15E-14	***
categoryother	-0.203001	0.046395	-4.375	1.23E-05	***
categoryotherasian	-0.372029	0.057198	-6.504	8.40E-11	***
categoryothereuropean	-0.289007	0.097082	-2.977	0.002922	**
categorypizza	-0.64579	0.034175	-18.897	<2e-16	***
categorysandwiches	-0.85918	0.044656	-19.24	<2e-16	***
categoryseafood	0.301452	0.061283	4.919	8.91E-07	***
categorysoul_food	-0.036079	0.154407	-0.234	0.815257	
categorysouthern	-0.1077	0.153362	-0.702	0.482544	
categoryspanish	0.110111	0.078953	1.395	0.163175	
categorysteakhouses	0.928564	0.065865	14.098	<2e-16	***
categorysushi_bars	0.205773	0.038764	5.308	1.14E-07	***
categorythai	-0.193253	0.04799	-4.027	5.71E-05	***

logwordlen	2.314061	0.160053	14.458	<2e-16	***
loglength	0.274683	0.016534	16.613	<2e-16	***
lognumitems	-0.082184	0.010498	-7.828	5.74E-15	***
adjectives	-0.228804	0.052673	-4.344	1.42E-05	***
plenty	-0.931954	0.305653	-3.049	0.002305	**
consumer choice	-0.595383	0.06144	-9.691	<2e-16	***
natural authenticity	2.308947	0.159477	14.478	<2e-16	***
traditional authenticity	-0.45395	0.134103	-3.385	0.000716	***

## Why Expensive Menus are short

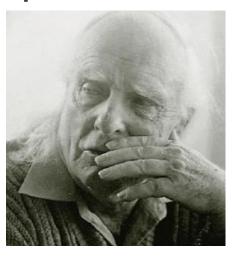
Why say food is "fresh"?

Because Grice

H. Paul Grice, Philosopher of Language

Known for his "Cooperative Principles"

4 Maxims of language



## Grice's "Cooperative Principle"

Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchanged in which you are engaged."

### Grice's Maxims

#### I. Quantity

- A. Be informative
- B. Don't give more information than is required

#### **II. Quality**

- A. Don't lie or mislead
- B. Don't make statements unless there is adequate evidence

#### **III. Relation**

A. Be relevant

#### IV. Manner

- A. Avoid obscurity
- **B.** Avoid ambiguity
- C. Be succinct
- D. Be orderly

### These Maxims sound obvious

But they are useful in two ways:

- 1) Sometimes what seems to violate a Maxim actually doesn't, and the listener has to figure something out
- 2) Sometimes they are violated ("flouted"). When that happens, listeners search for an explanation for the violation

Both of these let us *imply* things without saying them And these explain how speakers can mislead without technically lying

### Maxim of Relevance

Make your contribution relevant

Relevance:

A: Where's Bill?

B: There's a yellow VW outside Sue's house.

Flouting Relevance

A: How did you like my lecture?

B: Nice weather we're having!

### Relevance and ketchup

https://www.youtube.com/watch?v=1xbYD7 9 g98

## Maxim of Quantity

Make your contribution neither more nor less than is required.

a. Make your contribution as informative as is required for the current purposes of the exchange.

b. Do not make your contribution more informative than required

Flouting Quantity:

[Letter of Recommendation]:

"Dear Madam, Mr. X's handwriting is excellent, and his attendance at class has been regular.

Yours Sincerely"

### Maxim of Quantity

Make your contribution neither more nor less than is required.

a. Make your contribution as informative as is required for the current purposes of the exchange.

b. Do not make your contribution more informative than required

**Violating Quantity:** 

A: How did Harry fare in court the other day?

B: Oh he got a fine

If it turned out Harry also got a life sentence, then B is guilty of misleading A, even though what he says is technically true

#### Maxim of Manner

#### Be perspicuous

- a. Avoid obscurity of expression
- b. Avoid ambiguity
- c. Be brief (avoid unnecessary prolixity)
- d. Be orderly

#### Flouting Manner:

- a. "Miss X sang `Home on the Range'
- b. "Miss X produced a series of sounds that corresponded closely with the score of 'Home on the Range'.

## Maxim of Quality

Try to make your contribution one that is true

- a. Do not say what you believe to be false
- b. Do not say that for which you lack adequate evidence

Flouting Quality:

Bouncer: Are you 21?

Student: I'm 35.

Bouncer: Oh, right, and

I'm 99.

## Why Expensive Menus are short

Why say food is "fresh"?

Maxim of Quantity

Don't say more than you need

Maxim of Relation

Be relevant

There must be a purpose to say "fresh"

## Why Expensive Menus are short

#### Why say food is "fresh"?

- Maxims of Quantity and Relevance:
  - The hearer needs to know it's fresh
  - I am trying to communicate the freshness of my food

#### Why would you need to know?

- Because you don't already know if it's fresh
- i.e. you aren't sure if it's fresh

Expensive restaurants want you to **assume** food is fresh

#### Grice in action: real on menus

\$

- chocolate chip pancakes served with real whipped cream
- home made meatloaf served with real mashed potatoes
- lettuce, tomato, Russian dressing and real bacon bits

\$\$

- California roll: real crab and avocado
- blueberry whole grain pancakes with real maple syrup

"real" not on expensive menus!

## Menus as a window onto history

- New York Public Library's Buttolph collection
- 45,000 menus dating from the 1840s to the present,
- More than half collected by Miss Frank E. Buttolph (1850-1924)

http://menus.nypl.org/menus/decade/1850s

## When does "real" first became common on menus?

From the New York Public Library's Buttolph collection:

real butter 1960s
real whipped cream 1970s
real bacon 1990s

Why then?

1950: Margarine tax repealed and sales of margarine skyrocket

1960s: First popular artificial whipped cream introduced

1970: First popular artificial bacon introduced

The word "real" on menus is a marker of the rise of "artificial" food!

## Remember from last week Words promising more food

hearty, enormous, plenty, loads, lots, hefty, gigantic, generous, generously, largest, heaping, refills, bottomless, unlimited, huge, big, bigger, biggest, ginormous, mega, largest, and more, with more, tons of, king sized, texas sized

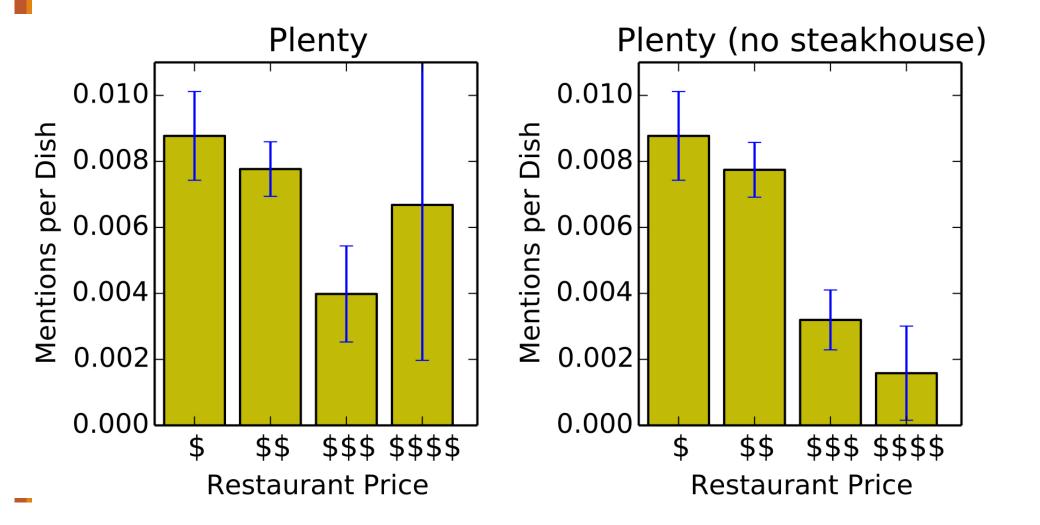
### Plenty

Baked Meat Lasagna with our *hearty* meat sauce Tuna Supreme: A *generous* scoop of tuna



### Plenty in steakhouses

**Big** Shoulders: and you thought our petite porterhouse was **big**! this is a **full forty ounces**! that's eight ounces short of three **pounds**! are you up to the challenge?



### Factor that increases price of a dish:

#### Use rare words

tonnarelli, bastilla, persillade

#### Use long words:

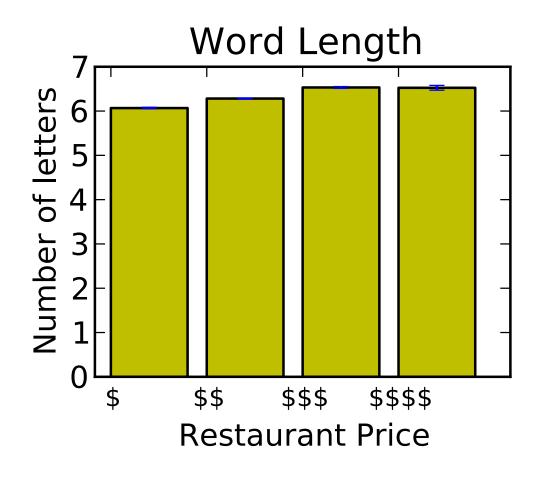
 decaffeinated, accompaniments, complements, exquisitely

#### cheap restaurants:

 decaf not decaffeinated, sides not accompaniments.

Silverstein, Michael. 2003. "Indexical Order and the Dialectics of Sociolinguistic Life." Language and Communication 23: 193–229.

## Factors predicting higher prices Each additional average letter = 18 cents



Digression:

Rare words are also longer words

"Accompaniments" vs

"Sides"

https://books.google.com/ngrams/graph? content=accompaniments,sides&year\_star t=1800&year\_end=2000&corpus=15&smo othing=3&share=&direct\_url=t1%3B,acco mpaniments%3B,c0%3B.t1%3B,sides%3B,c 0

# Word length is inversely correlated with frequency

Short words are frequent Frequent words are short

## This fact was discovered in the 8<sup>th</sup> century

Sibawayh: Founder of Arabic grammar and Arabic linguistics

 Author of the Kitāb, "The Book", a 900-page treatise on Arabic grammar (+morphology/phonology/phonetics)

~752 CE, born, probably Shiraz

- Ethnic Persian, non-native speaker of Arabic
- ~762 CE, arrives in Basra to study religious law
  - Makes a serious grammatical error in public
  - Is shamed in public
  - Decides to study linguistics instead

## Sibawayh and word frequency

A. A. Al-Nassir. 1993. *Sibawayh the Phonologist*. Kegan Paul. M. G. Carter. 2004. *Sibawayhi*. Oxford: I. B. Tauris

"Frequency of usage" (katra al-isti'māl) is a major cause of language changing over time

- Some words get shorter over time
- Because they "occurs more frequently"

Rediscovered by European and American linguists

1100 years later in the late 19<sup>th</sup> century

## 100 words overrepresented on \$\$\$ menus

28 of the 100 words were from high-status foreign cuisines:

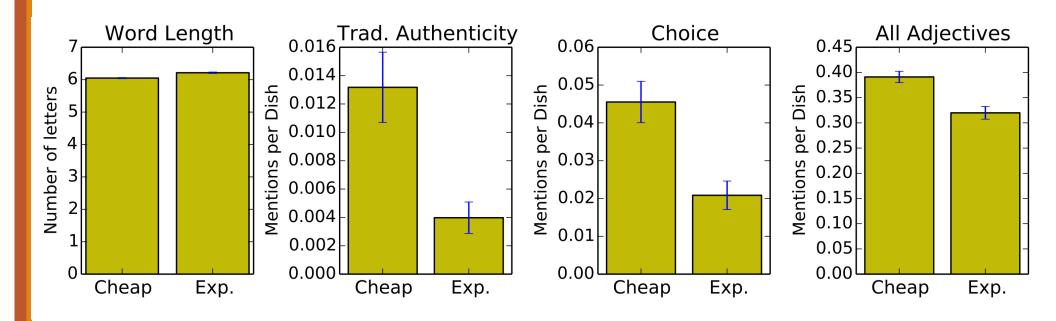
French (les, de, le, fois, gras, crème, mousse, tarte, pommes)

**Italian** (e, con, risotto, pancetta, burrata, polenta, parmigiano)

Japanese (tempura, uni, wagyu, sushi, yuzu, sashimi, miso, shabu, kobe, ponzu, wasabi, ninja, and sake)

## This has been true for over 100 years

New York Public Library's Buttolph Collection 2858 menus from 1892-1921



One difference: trickle-down in status (Veblen, Simmel)

gourmet: only fancy restaurants
 "braised lamb gourmet" Waldorf Astoria
 "tomato gourmet salad" Hotel Manhattan

#### 2010:

1910:

• gourmet: only cheap restaurants
"gourmet large pepperoni" delivery pizza

## There's more in menus: How healthy food is framed

Bradley P. Turnwald, Dan Jurafsky, Alana Conner, and Alia J. Crum. 2017. Reading Between the Menu Lines: Are Restaurants' Descriptions of "Healthy" Foods Unappealing? *Health Psychology*.







As obesity rates continue to climb, blame has fallen on the high-calorie meals at chain restaurants

Many chain restaurants have responded by adding "healthy choices" sections

But how do they describe the food on these healthy menus?

#### The data

Menus from 26 casual-dining chain restaurants with healthy submenus • e.g., Applebee's, Olive Garden, IHOP, Red Lobster

Every word on healthy and standard menu items coded for presence in 22 lexicons.

Which kinds of words occurred less often in the Healthy menus?

Restaurant	2015 Sales Rank	<b>Total Sales (billions \$)</b>
Applebee's	10	4.6
Panera Bread	11	4.3
Olive Garden	16	3.8
Chili's	17	3.6
<b>Buffalo Wild Wings</b>	19	3.2
IHOP	23	2.9
Denny's	25	2.5
Outback Steakhouse	26	2.5
Red Lobster	28	2.4
Cracker Barrel	30	2.1
Texas Roadhouse	33	1.9
Cheesecake Factory	35	1.8
TGI Friday's	36	1.8
Longhorn Steakhouse	39	1.5
Red Robin	41	1.4
Ruby Tuesday's	45	1.2
Waffle House	46	1.1

What does "overused" mean?

likelihood ratio: does "fun" occur more (by %) in menu A or B?

% of words in menu A that are "fun"

% of words in menu B that are "fun"

## Words under-used on "Healthy" menu

Theme	Example words
Exciting	crazy, spellbinding, action, adventure, blasts,
Fun	fun, dippable, bites, skewered, stacker, tanglers
Traditional	countryside, housemade, classic, traditional, famous,
Regional	Philly, California, Maine, New Orleans, Hawaiian
Texture	crispy, creamy, crunchy, flaky, gooey, velvety
Provocative	dangerous, dirty, naked, temptation, sinful, envy
Spicy hot	buffalo, mesquite, chipotle, fiery, firecracker, burnin'
Artisan	artisan, handcrafted, hand-prepared, premium,
Taste	sweet, sour, salty, tangy, flavorful, delicious
Indulgent	bliss, indulge, richest, succulent, decadent

## Words over-used on "Healthy" menu

Theme	Example words
Foreign	Asian, Italian, French, Tuscan, Thai, Mexican
Fresh	fresh, freshly
Simple	simple, dry, mild, plain, mildly, simply
Macronutrients	carb, fiber, fat, grain, protein, whole wheat
Thinness	light, lighter, enlightened, skinnylicious, lighten, lites
Deprivation	fat free, low-fat, sugar-free, no sugar added, low carb, reduced-fat
Nutritious	wholesome, nutritional, fit, healthy

#### Conclusion

- Major chain restaurants describe their healthiest choices as less appealing
- But we know people prioritize taste and indulgence over health
- Important for society: We need to improve how menus present healthy dishes to make them as appealing as standard items

## How to read a menu: Look for the words that aren't there

Luxury advertising is "modest", no mention of "real" or "fresh" or "delicious" or "generous portions"

Menus boasting of "real" foods are a tool to find the rise of "fake" food

(even though advertisers never use the word "fake")

Healthy menus avoid words indicating "fun", "spicy", "exciting", "tasty", "indulgent"

## How to read a menu: What's on expensive menus

Long words

Fancy words

High-status foreign words

# Summary: The linguistics of everyday life

Everyday objects like menus are a window onto culture.

Many such artifacts are covered in words!

Even very simple linguistic and statistical techniques can be useful!

Menus make excellent final projects!!