Empirical methods in linguistics [06]

Corpus linguistics 2

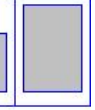
Prof. Dr. Martin Schäfer
Department of English Studies, Universität Leipzig

HA for today: *sneak* revisited

- At `www.english-corpora.org`, choose a corpus that can provide insight into the long-term diachrony (100+ years) of the verb *to sneak* and its past-tense variants in American English.
- Come up with a query that yields a list of the usage frequencies of the verb with all its inflected word forms.
- You should also be able to show the usage frequencies across time, e.g. more of a table with columns for decades (depending on your choice of corpus). What does the table reveal about the diachronic development of the forms sneaked versus snuck? Do you find this unexpected?
- Try different ways to visualise numerical information, e.g., just for either *sneaked* or *snuck*, generate a graph plotting the frequency of *sneaked* diachronically. Use the **display option** to change raw frequency to nomalized frequency, try the CHART display.

sneak again

The changing usage frequency of one of the two forms during the past 200 years of written American English:

SECTION	ALL	1820	1830	1840	1850	1860	1870	1880	1890	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
FREQ	572	0	0	0	0	0	0	1	0	1	4	1	6	10	8	27	45	44	94	130	201
WORDS (M)	405	7.0	13.7	15.8	16.5	16.9	18.8	20.1	20.4	22.0	23.1	25.7	27.7	27.4	28.7	29.1	28.8	29.9	33.1	34.8	35.5
PER MIL	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.05	0.17	0.04	0.22	0.36	0.28	0.93	1.56	1.47	2.84	3.73	5.67
SEE ALL YEARS AT ONCE																					

More than words

- Combinations of specific words = collocations
- Combinations containing specific words and open slots = patterns/constructions
  - morphological
  - syntactic

Collocations

- aim: detect meaningful word combinations (= exclude high frequency items like *the* etc.)
- terminology: node (word) + its collocate(s)
- vs. idioms: *kick the bucket*

Collocations

- aim: detect meaningful word combinations (= exclude high frequency items like *the* etc.)
- terminology: node (word) + its collocate(s)
- vs. idioms: *kick the bucket*

commit + *suicide, crime(s), offence*

Collocations

collocate	node	syn/sem description
a	smoker	(adjective: someone who smokes a lot)
to	ill	(verb: to become ill)
to	regret	(adverb: to regret sth. very much)
to	a party	(verb: to have/host a party)

Combinatorial patterns: wordforms/constructions

- *delicious* again :)
- *as fit as a fiddle*

Combinatorial patterns: Syntax

Syntactic constructions with word-/POS-specific slots
How could you examine

- the type and token frequency of the future progressive in American English?
- whether temporal *just* occurs with past tense in British English? Also inspect a random sample after your query to get a better feel for the result.
- the distribution of the standard passive construction in different text types (e.g. in British English)?

Next week and more

- Read Plag (2018), chapter 4, Affixation **prepare for this session with the help of the reading guideline!**
- Next week: discussion of Plag's chapter
- In two weeks:
  1. The specifics of your projects (which affix, which corpus etc.)
  2. Distribution of slots (in class).

References I