Studying Language Change with data from the Penn-Corpora

Cqp@fu for beginners

Logging on…

(1) Log onto the structeng-server by typing in your FU-username and password into putty (Win)/the terminal (Mac). (Host name: `login.fu-berlin.de`)

(2) Enter the cqp room by typing `cqp`.

(3) Find out which corpora are available by typing `show`.

(4) Open the Penn Parsed Corpus of Early English Correspondence by typing `PPCEEC`.

First queries…

(5) Type `"going"`. What happens?

(6) Type `"goinge?"`. What happens?

(7) Type `"go[iy]nge?"`. What happens?

(8) Type `show +pos`. Then repeat your query. (Hit ↟ to re-use your last commands.) What do you see now?

(9) Type `show -pos`.

Multi-word queries

(10) Type `[word="go[iy]nge?" & pos="N.*"]`. Compare the results to those of `[word="go[iy]nge?" & pos="V.*"]`. What is the difference?

(11) Type `show +pos`. Then type `"is" "going"`. Which pos-tags are *is* and *going* tagged with? Use these pos-tags to construct a query that also finds other inflected forms of *to be* followed by *going* (and spelling variants).

(12) Type `show -pos` (if you want).

Sorting, randomizing, counting

(13) Type `[word="go[iy]nge?" & pos="V.*"][word="to"][pos="N.*"]`. How many hits do you get?

(14) Repeat the query, but replace the third node by `[pos="VB"]`. How many hits do you get now?

(15) Type `sort Last randomize`. What happens?

(16) Type `reduce Last to 20`. Then type `cat Last`. What happens now?

(17) Type `count Last by word`. What happens?

(18) Type `count Last by word match[1]` and `count Last by word matchend`. What happens?

Restricting

(19) Repeat your last query (`[word="go[iy]nge?" & pos="V.*"] [word="to"] [pos="VB"]`). How many hits?

(20) Repeat the query, adding `::match.letter_period="E2"`. How many hits? How many hits for E3? (You can also try the query on the PPCEME, replacing `"letter"` by `"text"`.)

(21) Type `info` to learn what you've just done.

Inspiration for research projects

Asleep, awake, afloat – a-adjectives

(1) Query the OED (www.oed.com; campus license!) for adjectives with the prefix *a-* (*asleep, alive, afloat* etc. – only those where *a-* is a proper prefix in the sense that the rest of the word constitutes a meaningful word in English). Combine some of these adjectives in a cqp-query that might look like this: `[word="asleep|alive|afloat|…"]`. How are your adjectives distributed? Establish the part of speech of the slot before and after your node (`count Last by pos match[-1]/match[+1]`).

(2) Return to the OED and look up the etymology of some of your adjectives. Can you correlate the original function of the prefix with modern-day restrictions on the adjectives' use? (Further reading: Boyd & Goldberg 2011)

(3) Implications: What does this have to do with the prototype-nature of lexical categories?

You and me used to be together… Subject and object pronouns

(1) English has no case system any more, but in the pronoun system, some case distinctions have survived. English distinguishes subject pronouns (*I, we*) from object pronouns (*me, us*). *I* is used for subjects and *me* is used for objects. However, the phrases *you and I* and *you and me* seem to be somewhat interchangeable. *You and me* in subject function is not historical and therefore considered bad style (or just plain WRONG), as is *you and I* in object function.

  (a) Find out how often *you and me* and *you and I* occur in subject function (i.e. pre-verbally) and in object function (i.e. post verbally or following a preposition) in the PPCMBE/the BNC/the COCA-S or COHA-S.

  (b) Visualize your results.

  (c) Interpret what you find. (A sideways glance at French *moi* or Swedish *dom* may be instructive, if you happen to be familiar with either.)

  (d) Any thoughts on what's right or wrong, and why?

The emergence of the *about to* construction

(1) Find out what the etymology of *about* is (ask the OED). While you're at it, check spelling variants.

(2) Query BE + *about* in the Penn Corpora. Which senses does *about* have?

(3) Focus on constructions in which *about* is followed by a verb form (*to*-infinitive or gerund). What do these mean?

(4) How might *about to* have acquired its present meaning?

The emergence of *do*-support

(1) In which sentence types does modern English require the auxiliary *do*? Consult the BNC to determine whether this is a law or only a tendency.

(2) *Do*-support was not obligatory in EmodE. Query the Penn-Corpora to establish the relative frequencies of do-periphrasis and simple constructions for each sentence type.

(3) In the PPECCE, do women use it more than men?

The emergence of the *going to* future construction

(1) Query the Corpus of Early English Correspondence (PPCEEC) for the construction *going to* (remember to include spelling variants – use the OED to find them).

(2) Repeat the query four times, limited to texts matching the `text_period` M3, E1, E2, E3. Take a random sample of 10 instances of *going to* for each subperiod (`sort Last randomize`, then (!) `reduce Last to 10`). Which of the sentences express futurity, which (also) express something else?

	passage	motion	either/ both	futurity
M3 S1	...			
M3 S2				
...				

(3) What is the relative (!) frequency with which *going to* is followed by infinitives, as opposed to NPs, in each subperiod? (Type `count Last by pos matchend[+1]` to find out.) Visualize!

(4) Return to your non-reduced results (repeat query if you haven't saved them). Which are the most frequent infinitives immediately following *going to*? Can you make out any changes between the subperiods?

(5) When did *going to go* become possible in English? (Ask appropriate corpora.) What is the significance of this?