Collocation: Practice

Gede Primahadi Wijaya Rajeg

2024-07-20

Materials

- source files for all materials:
  - https://github.com/complexico/dipscorling2024
- pdf version as a handout here
- How to cite these materials:

> Rajeg, Gede Primahadi Wijaya. 2024. Materials for the *Diponegoro Summer Course in Corpus Linguistics* (*DipSCORLING 2024*) (22 - 27 July 2024). R Quarto. Zenodo. https://doi.org/10.5281/zenodo.12793922. (22 July, 2024).

Collocation via concordance

1. Generate 50 random concordance-lines with the word-form *endangered* (you may want to use the `ADVANCED` interface with CQL so that you do not get the form with capital letter to exclude proper name like *Endangered Language Archive*)
   - try to identify the syntactically relevant collocates of *endangered*
   - pay attention to the part-of-speech of *endangered* (it can be a verb in simple past or past participle form and as a participial adjective)
     - pay attention to the relevant syntactic relation of *endangered* in a given part-of-speech to identify the collocates
   - what kind of entity gets *endangered*?
   - what is the proportion of verbal vs. adjectival usage of *endangered*?

Phraseology 1

You will use the `ADVANCED` tab of the N-GRAMS feature

Tasks

1. say you are interested in multi-word expression (from three to four words) that revolves around the word *shiny* (case **insensitive**) (my output: https://ske.li/158)
    - identify the co-occurrence of *shiny* with another word within a nominal coordination construction
        - this is the qualitative aspect of corpus analysis
        - this is a syntactically-oriented analysis of co-occurrence data from corpus
2. find multi-word expression containing three words
    - the expression is ended with words containing the suffix *-ly*
3. find multi-word expression containing the word *talk*
    - you want *talk* to initiate (i.e., appear in expression-initial) the expression
4. how would you find multi-word expression with the following criteria?
    - a three-word sequence
    - containing the coordinating conjunction *and*
    - only in the following three-gram pattern: [`ANY.WORD` *and* `ANY.WORD`]
        - my answer after you did yours: https://ske.li/16m (check the criteria of my query)
    - Hint: Sketch Engine does not have a ready-made feature to handle this query in the N-grams, but:
        - you could make use of the regular expression feature on the output, OR
        - I want you to think about another possible workaround on this issue and let's discuss
    - Take away message:
        - a feature in our tool may not always provide an *explicit*, *direct* way to do thing
            * we need to find a workaround given this issue
        - any one doing corpus linguistics *must* know regular expression, in my opinion

Phraseology 2: Semantic field

You are interested in studying the semantic landscape of lexical verbs that express certain action towards body parts in the constructional pattern [`LEX-VERB pronoun in the BODY-PART-NOUN`] (as in "*poke X in the eye*" (Langacker 2008: 20)).

The point of this practice relates to the topic of:

a. the profile of semantic field of collocates of a (class of) word (cf. the lecture slide) (Hunston 2002)

b. the role of collocation to find phraseology of a word (Hunston 2002)

c. corpus query

Here are the list of body-part noun lemmas that you will include in the search queries:

> *face*, *body*, *eye*, *neck*, *head*, *chest*, *stomach*, *belly*, *leg*, *foot*, *hip*, *buttock*, *ass*, *cheek*, *arm*.

You can add yours too.

Task

- How would you translate the aforementioned theoretical inquiry into operational query in Sketch Engine?
- What corpus tool of Sketch Engine would you use?
  - in your query, attempt to include the body-part noun at once/simultaneously in one go
  - HINT: you can solve this inquiry in ONE search
- How many tokens do you get?
- Can you directly get the type frequency of the pattern expressing the meaning 'exerting action/force towards somebody's body part'?
- How would you go about processing the output of your query so that you could answer your inquiry?
  - the semantic range of the lexical verb slot in the pattern
  - whether every verb can co-occur with every body-part noun
- LET'S DISCUSS YOUR ANSWERS and ANY ISSUES
  - My own solution to this inquiry: https://ske.li/16n

Meaning via collocation

You will use the WORD SKETCH feature

Task

- We will look at the lemma LEAK (Hunston 2002: 76)
  - make an initial prediction about *what is it that leaks* to check in the output
- In the output focus on the syntactic relation of the lemma as **verb** *LEAK*.
  - is your initial prediction confirmed?
  - how many senses could you postulate for the use of the verbal lemma *LEAK*?

Meaning via collocation across text-topic

You will use the WORD SKETCH feature

Task

- Determine separately what gets *viral* in:
  - *culture & entertainment* VS. *science* text topic
  - How would you translate that instruction into query?
  - What could you discuss from the output of the two text topics regarding what gets *viral*?

Distinct co-occurrence patterns of near-synonyms

You will use the WORD SKETCH DIFFERENCE feature.

Tasks

The adjectives chosen here are taken from Stefanowitsch's (2003: 2) lecture note.

- Contrast *costly* vs. *expensive*
  - Do these adjectives apply to (i.e., can modify) the same nouns?
  - What are the nouns that tend to be costly but not expensive?
    - could you abstract away from the specific word form to a coherently semantic grouping of the collocates?
- Contrast *earn* vs. *gain*
  - what could you earn vs. gain?
    - could you abstract away from the specific word form to a coherently semantic grouping of the collocates?
  - **how/in what way** could you earn vs. gain something?
    - which collocate type would you check to answer this question?

References

Hunston, Susan. 2002. *Corpora in applied linguistics*. 1st edn. Cambridge ; New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139524773.

Langacker, Ronald W. 2008. *Cognitive grammar: A basic introduction*. Oxford: Oxford University Press.

Stefanowitsch, Anatol. 2003. Synonymy problems. Corpus Linguistics lecture notes. (20 November, 2013).