

Collocation

Gede Primahadi Wijaya Rajeg

2024-07-22

Roadmap

1. Collocation: A brief history & basic notion
2. Two methodological perspectives on “collocation”
3. Why is collocation useful?
4. Practice

Collocation: A brief history & basic notion

- First used by J. R. Firth (1957):

“I propose to bring forward as a technical term, meaning by *collocation*, and apply the test of *collocability*.” (Firth 1957: 194, cited from Xiao 2015: 107)
- “Collocations of a given word are statements of the habitual or customary places of that word” (Firth 1962: 12)
 - NOT just a “mere juxtaposition”, but “an order of *mutual expectancy*” (Firth 1962: 12)
- Aspect of the “Neo-Firthian Corpus Linguistics” (McEnery & Hardie 2012, Ch. 6)
 - leading figure: John Sinclair
 - Firth’s idea meets corpus linguistics
 - Two central approaches: **collocation** & **discourse**
 - meaning by collocation:

* an important aspect of word meaning does not exist in isolation but subsists in the features of association that word participates in with other words and/or structures it frequently co-occurs (McEnery & Hardie 2012: 123)

- broad definition:

“co-occurrence patterns observed in corpus data” (McEnery & Hardie 2012: 123)

- but how is this “co-occurrence” identified?
- there are two broad methodological/operational perspectives

Roadmap

~~1. Collocation: A brief history & basic notion~~

2. Two methodological perspectives on “collocation”
3. Why is collocation useful?
4. Practice

Two methodological perspectives on “collocation”

“collocation via concordance”

- manual scanning of concordance lines
- linguist’s intuition + linguistic skill
- e.g., Sinclair’s classic study on *naked eye*.
- my experience:
 - focus on syntactically relevant words of the node
 - I use it to study metaphoric construction and conceptual metaphor for emotion

Two methodological perspectives on “collocation”

“collocation via concordance”

- manual scanning of concordance lines
- linguist’s intuition + linguistic skill
- e.g., Sinclair’s classic study on *naked eye*.
- my experience:
 - focus on syntactically relevant words of the node
 - I use it to study metaphoric construction and conceptual metaphor for emotion

“collocation via significance”

- statistical test on the co-occurrence frequency
- commonly available in modern corpus tools
- mainly ignore syntactic relation
 - relying on linear span of words around the node (cf. Hunston 2002: 69)
 - but cf. SkE’s *Word Sketch*
- there is also syntactically-oriented collocation (cf. Stefanowitsch & Gries 2009)
 - e.g., *Collostructional Analysis* (Stefanowitsch 2013)
 - * manual inspection is still used
 - or SkE’s Word Sketch
 - * in a click of a button

See McEnery and Hardie (2012: 123ff for details)

Roadmap

- ~~1. Collocation: A brief history & basic notion~~
- ~~2. Two methodological perspectives on “collocation”~~
3. Why is collocation useful?
4. Practice

Why is collocation useful?

- to detect different senses of a word based on its collocate types (Hunston 2002)
 - For instance, the different sense of *deep* (<https://ske.li/16i>)
- to reveal phraseology of a given node word (Hunston 2002: 77–78)
 - We could use an n-gram tool, another kind of co-occurrence concept (McEnery & Hardie 2012: 123)
 - For example, the phraseology for the word *shoulder* (<https://ske.li/16l>)
- to unpack variation in the usage patterns of near-synonyms (Xiao 2015)
 - For instance, distinct co-occurrence patterns of *happiness* and *joy* (<https://ske.li/16j>) (cf. Stefanowitsch 2004)
- to obtain the profile of semantic field of a word (Hunston 2002)
 - Example 1: the semantic field of the EXPERIENCER ([fig](#)) noun-collocates of *angry* vs. *mad* (Suari, Rajeg & Sudipa 2024) (cf. the repository for data and code [here](#))

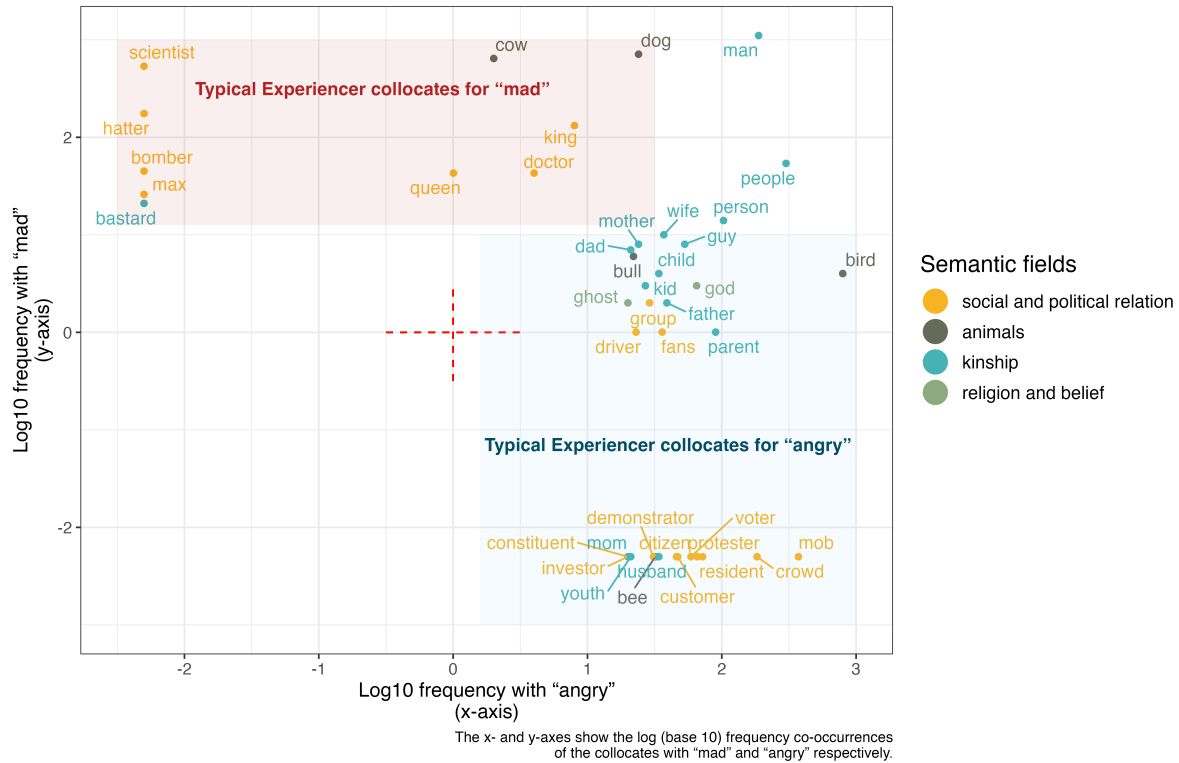


Figure 1: The semantic field of EXPERIENCER collocates for *angry* and *mad* in COCA in the pattern [*angry/mad* NOUN].

* the semantic categorisation for these collocates refers to the cross-linguistic semantic catalogue called the *Concepticon* (List, Cysouw & Forkel 2016)

– Example 2: the semantic field profile of *bribe* and *bribery* (see Hunston 2002: 78)

Roadmap

1. Collocation: A brief history & basic notion
2. Two methodological perspectives on “collocation”
3. Why is collocation useful?
4. Practice

End of Collocation

- 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).

References

- Firth, John Rupert. 1962. *Studies in linguistic analysis: Special volume of the philological society*. Oxford: Blackwell. <http://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf>.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. 1st edn. Cambridge ; New York: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524773>.
- List, Johann-Mattis, Michael Cysouw & Robert Forkel. 2016. Concepticon: A resource for the linking of concept lists. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, et al. (eds.), 2393–2400. Paris, France: European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2016/summaries/127.html>.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.
- Stefanowitsch, Anatol. 2004. HAPPINESS in English and German: A metaphorical-pattern analysis. In Michel Achard & Suzanne Kemmer (eds.), *Language, Culture, and Mind*, 137–149. Stanford, CA: CSLI.
- Stefanowitsch, Anatol. 2013. Collostructional Analysis. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford handbook of Construction Grammar*, 290–306. Oxford: Oxford University Press.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2009. Corpora and Grammar. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics: An international handbook*, vol. 2, 933–951. Berlin: Mouton de Gruyter.
- Suari, Ida Ayu Saskara Tranggana, Gede Primahadi Wijaya Rajeg & I. Nengah Sudipa. 2024. Exploring Grammatical and Semantic Profiles of Angry and Mad: A Corpus-Based Study. *EJI (English Journal of Indragiri) : Studies in Education, Literature, and Linguistics* 8(2). 272–296. <https://ejournal-fkip.unisi.ac.id/eji/article/view/2749>.

Xiao, Richard. 2015. Collocation. In Douglas Biber & Randi Reppen (eds.), *The Cambridge handbook of English corpus linguistics* (Cambridge handbooks in language and linguistics), 106–124. Cambridge: Cambridge Univ. Press.