Collostructional analysis: A short primer

Stefan Hartmann
HHU Düsseldorf

Overview

- Theoretical background: Construction Grammar
- What is collostructional analysis?
- Types of CA and hands-on examples
- Potential, limitations and criticism

Theoretical background: CxG

Heinrich Heine
Universität Düsseldorf

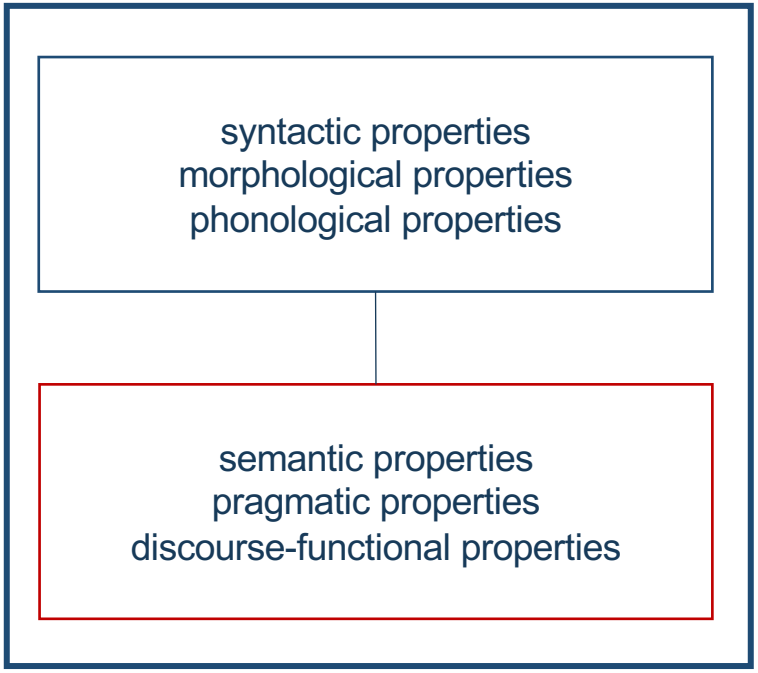
- Construction Grammar (CxG) sees language as a network of **constructions**, i.e. form-meaning pairs
- C is a CONSTRUCTION if and only if C is a form-meaning pair <F_i S_i> such that some aspect of F_i or some aspect of S_i is not strictly predictable from C's component parts or from other previously established constructions. (Goldberg, 1995, 4)

Constructions

Heinrich Heine
Universität Düsseldorf

- Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not stricdy predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency. (Goldberg, 2006, 5)
- constructions are understood to be emergent clusters of lossy memory traces that are aligned within our high- (hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions (Goldberg 2019, 7)

Constructions



(Croft 2001)

Constructions: Lexicon-syntax continuum

Heinrich Heine
Universität Düsseldorf

Morpheme constructions

e.g. *anti-, pre-, ing*

Word constructions

e.g. *grumpy, cat, say, no*

Morphological constructions

z.B. [X-*er*]; [V-*ing*]; [*un*-X]

Syntactic constructions

z.B. [SUBJ V_{TRANS} OBJ]

Filled and partially filled constructional idioms

e.g. *kick the bucket; the* X-*er the* Y-*er*

e.g. Croft 2001, Goldberg 2006, Hilpert 2014

What is collostructional analysis?

Heinrich Heine
Universität Düsseldorf

- family of methods for investigating relationships between constructions

- prototypically, it is used to investigate the relationship between **lexical items** and partially filled **syntactic constructions**

Aims of this tutorial

- find out how collostructional analysis works
- discuss potential and limitations
- hands-on examples in R

R package *collostructions*

- developed by Susanne Flach (Neuchâtel / Zurich)
- available at https://sfla.ch/collostructions/
- (not yet on CRAN)

Author(s)

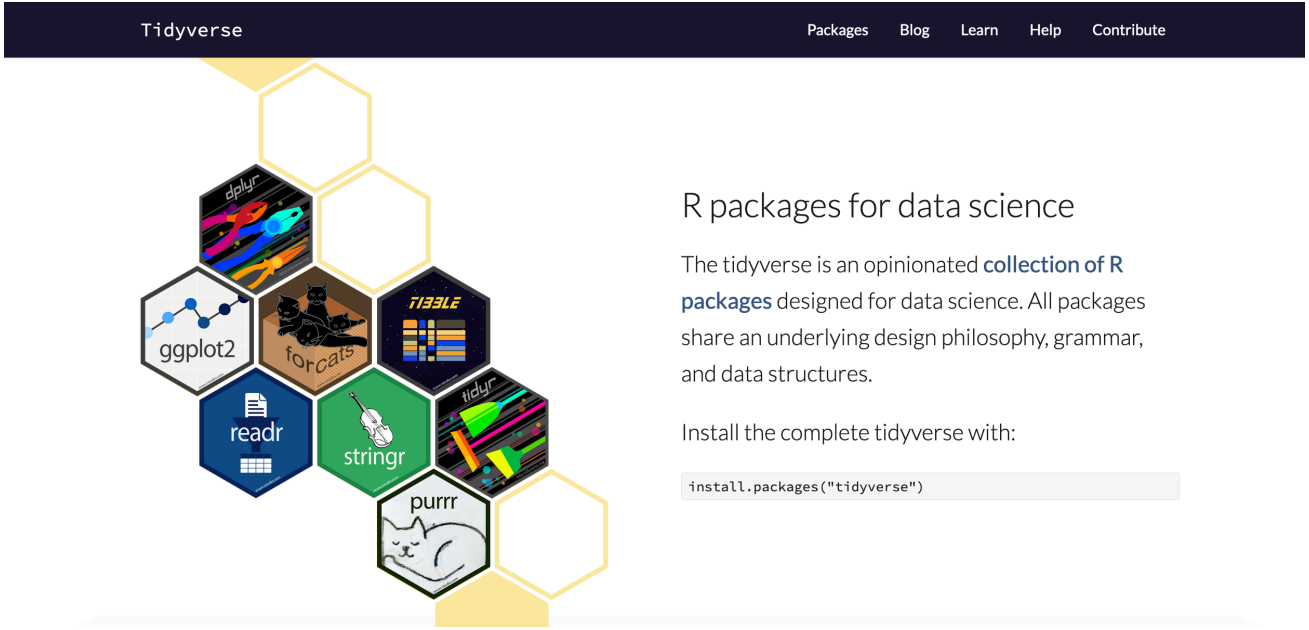
Susanne Flach, susanne.flach@unine.ch

Thanks to Anatol Stefanowitsch, Berit Johannsen, Kirsten Middeke and Volodymyr Dekalo for suggestions, debugging, and constructive complaining, and to Stefan Hartmann, who never complained, but provided valuable feedback when asked how the package could be improved.

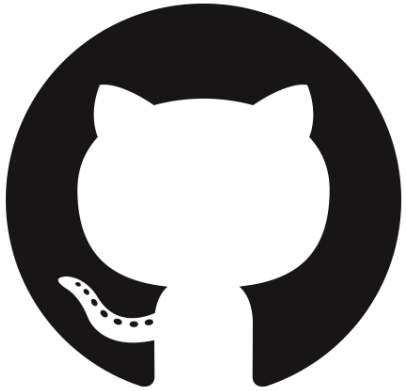
Other R packages used in this tutorial

Heinrich Heine
Universität Düsseldorf

- *tidyverse* family of packages (apologies to base R purists ☺)



Materials



https://github.com/empirical-linguistics/collostructions-tutorial

(folder "data")

Examples: Snowclones (Ungerer & Hartmann in prep.)

Heinrich Heine
Universität Düsseldorf

- extravagant formulaic patterns, e.g. *mother of all X, X is the new Y*
- lend themselves well to collostructional analysis:
  - patterns with 1 or 2 open slots
  - CA can give clues to semantic constraints on their productivity

Example: Snowclones

- examples: *mother of all X, X is the new Y*
- database: DECOW (Schäfer & Bildhauer 2012)

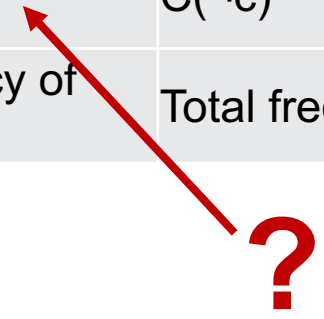
Corpus	**No. of tokens**	**Hits for [*the mother of all* X]**	**Hits for [X *is the new* Y]**
ENCOW	16.8bn	4,127	3,848

Simple collexeme analysis

	Word l_i of Class L	Other Words of Class L	Total
Construction c of Class C	Freq. of L(l_i) in C(c)	Frequency of L(¬l_i) in C(c)	Total frequency of C(c)
Other Constructions of class C	Frequency of L(l_i) in C(¬c)	Frequency of L(¬l_i) in C(¬c)	Total frequency of C(¬c)
Total	Total Total frequency of L(l_i)	Total frequency of L(¬l_i)	Total frequency of C

Simple collexeme analysis

	Word l_i of Class L	Other Words of Class L	Total
Construction c of Class C	*mother of all hangovers*	*mother of all* [¬hangover]	Total frequency of C(c)
Other Constructions of class C	[¬mother of all] *hangover*	[¬mother of all] [¬hangover]	Total frequency of C(¬c)
Total	Total Total frequency of L(l_i)	Total frequency of L(¬l_i)	Total frequency of C



Simple collexeme analysis

	Word l_i of Class L	Other Words of Class L	Total
Construction c of Class C	*mother of all hangovers*	*mother of all* [¬hangover]	Total frequency of C(c)
Other Constructions of class C	[¬mother of all] *hangover*	[¬mother of all] [¬hangover]	Total frequency of C(¬c)
Total	Total Total frequency of L(l_i)	Total frequency of L(¬l_i)	Total frequency of C

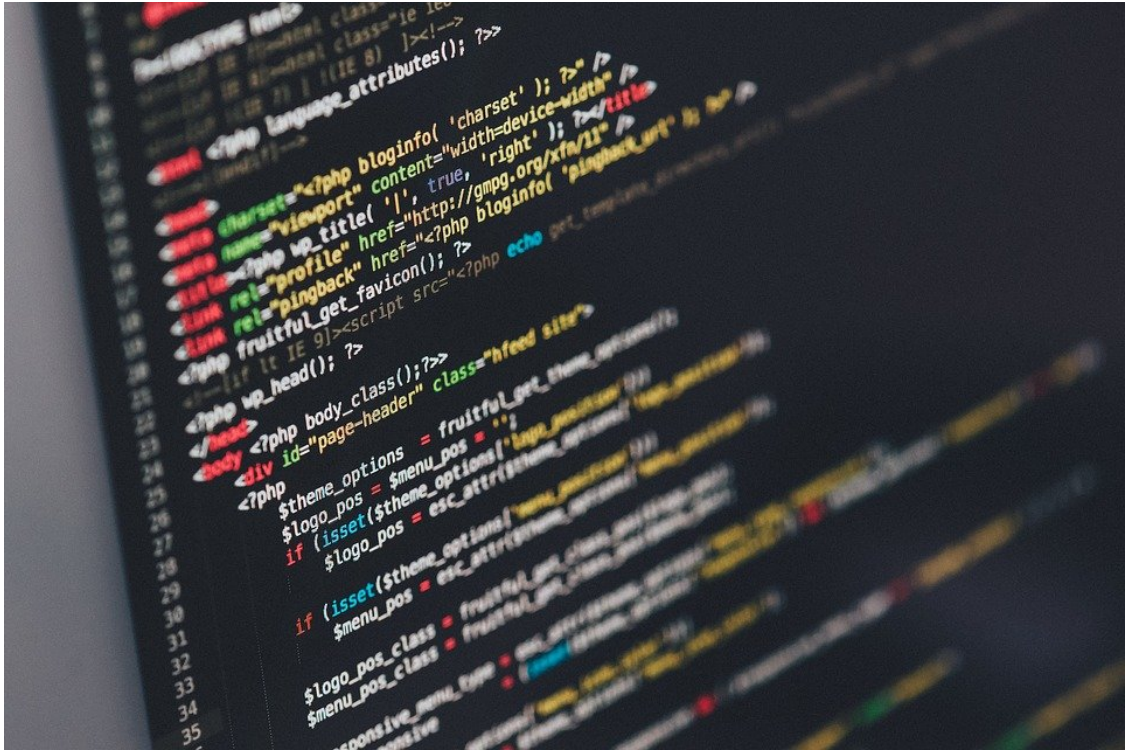
association measure

Association measures

Heinrich Heine
Universität Düsseldorf

- most widely-used AMs in CA (at the moment):
  - p-value of Fisher-Yates Exact Test (for smaller samples)
  - Lok-Likelihood ratio G^2 (for larger samples)
- all measures from Evert (2005) implemented in Flach's (2017) R package

Hands-on example



Covarying collexeme analysis

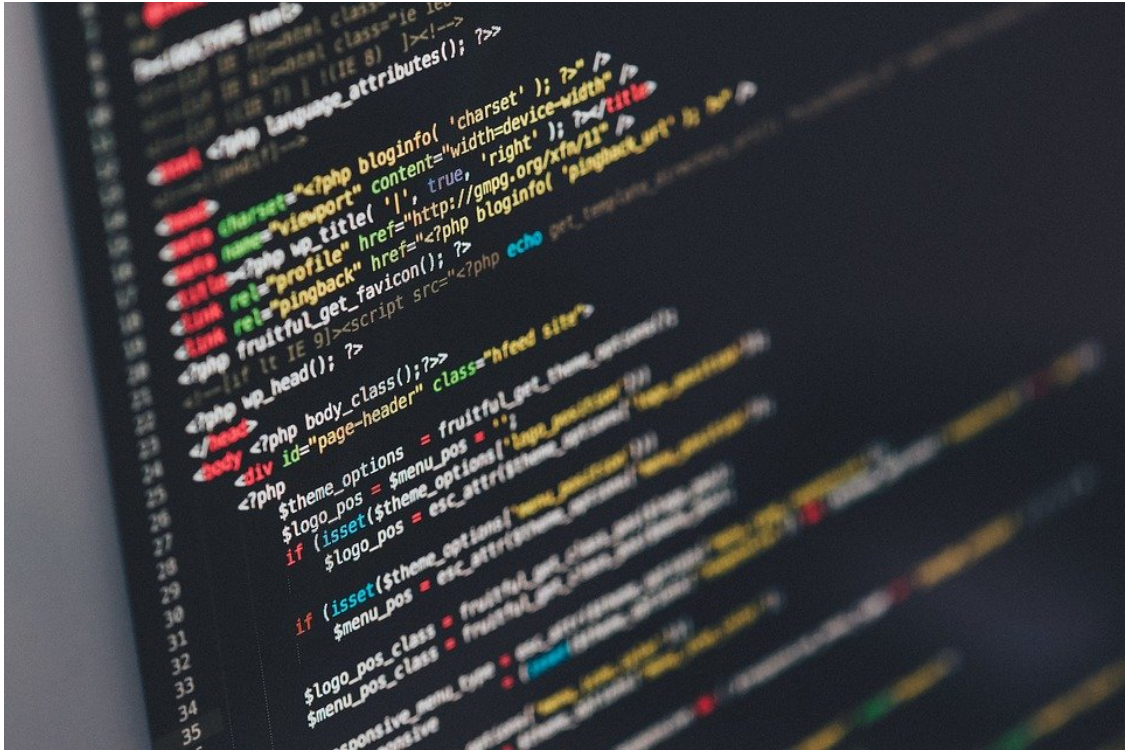
Heinrich Heine Universität Düsseldorf

	Word l_1 in slot s_1 of construction C	Other words in slot s_1 of construction C	Total
Construction c of Class C	Freq. of $s_1(l_1)$ and $s_2(l_2)$ in C	Freq. of $s_1(\neg l_1)$ and $s_2(l_2)$ in C	Total frequency of $s_2(l_2)$ in C
Other Constructions of class C	Freq. of $s_1(l_1)$ and $s_2(\neg l_2)$ in C	Freq. of $s_1(\neg l_1)$ and $s_2(\neg l_2)$ in C	Total frequency of $s_1(l_1)$ in C
Total	Total frequency of $s_1(l_1)$ in C	Total frequency of $s_1(\neg l_1)$ in C	Total frequency of C

Covarying collexeme analysis

	Word l_1 in slot s_1 of construction C	Other words in slot s_1 of construction C	Total
Construction c of Class C	pink is the new black	¬pink is the new black	Total frequency of $s_2(l_2)$ in C
Other Constructions of class C	pink is the new ¬black	¬pink is the new ¬black	Total frequency of $s_1(l_1)$ in C
Total	Total frequency of $s_1(l_1)$ in C	Total frequency of $s_1(\neg l_1)$ in C	Total frequency of C

Hands-on example



Distinctive collexeme analysis

Heinrich Heine
Universität Düsseldorf

	Word l_i of Class L	**Other Words of Class L**
Construction c_1 of Class C	Freq. of L(l_i) in C(c_1)	Frequency of L($\neg l_i$) in C(c_1)
Construction c_2 of class C	Frequency of L(l_i) in C(c_2)	Frequency of L($\neg l_i$) in C($\neg c_2$)
Total	Total Total frequency of L(l_i) in C(c_1, c_2)	Total frequency of L($\neg l_i$) in C(c_1, c_2)

Distinctive collexeme analysis

Heinrich Heine
Universität Düsseldorf

	Word l_i of Class L	**Other Words of Class L**
Construction c_1 of Class C	*start to* despair	*start to*¬despair
Construction c_2 of class C	*begin to* despair	*begin to* ¬despair
Total	Total Total frequency of L(l_i) in C(c_1, c_2)	Total frequency of L(¬l_i) in C(c_1, c_2)

Distinctive collexeme analysis

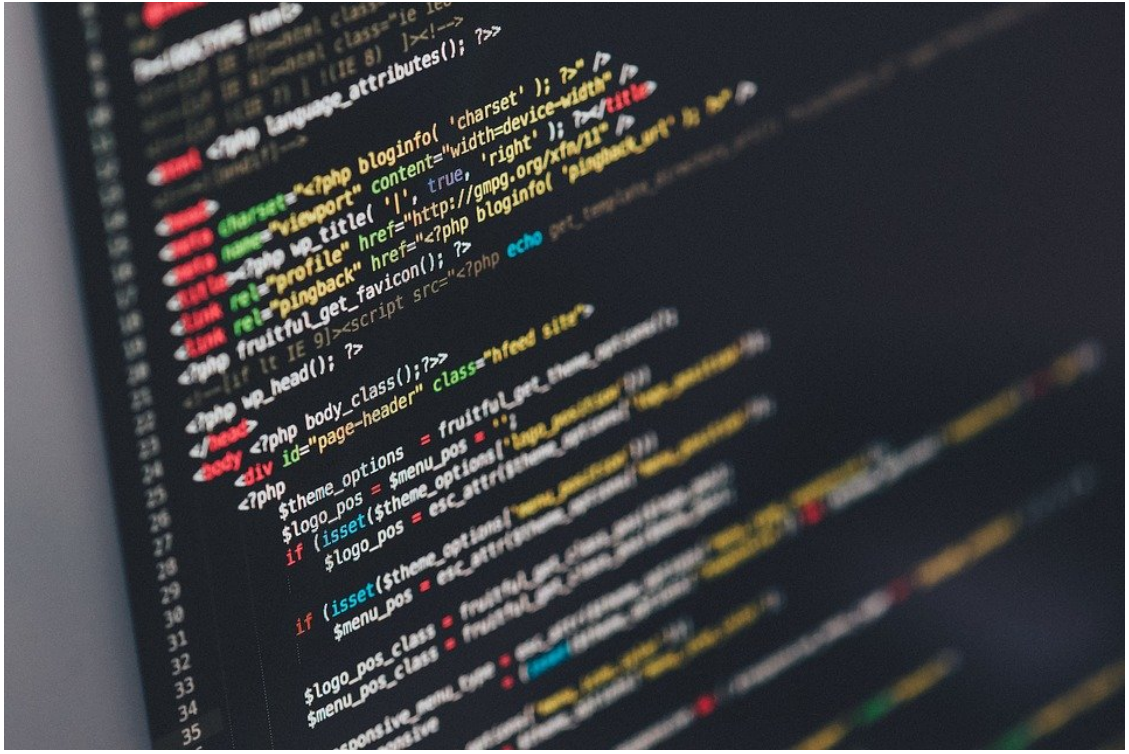
- Example (from the R package):

start to V

vs.

begin to V

Hands-on example



Criticisms

- "Language is never ever ever random" (Kilgarriff 2005)
  - Schmid & Küchenhoff (2013): corpus data are not randomly sampled – as a result, phenomena collected in a corpus cannot be independent observations
  - (but: less relevant if CA is seen as an exploratory, rather than hypothesis-testing, method)
- Filling the fourth cell
  - "the decision concerns the definition of the nature and size of the construction serving as observational unit (Schmid & Küchenhoff 2013: 544)

Alternatives

- Schmid's (e.g. 2000) *attraction* and *reliance*
  - basically like CA without the fourth cell...

$$attraction = \frac{a}{a + c}$$

$$reliance = \frac{a}{a + b}$$

- Example: *give* and ditransitive construction
  - *a*: Frequency of *give* in ditransitive construction
  - *b:* Frequency of ditransitive construction
  - *c:* Frequency of *give* in all other contexts

References

- Evert, Stefan. 2005. *The statistics of word cooccurrences. Word pairs and collocations*. Stuttgart: Institut für maschinelle Sprachverarbeitung.
- Flach, Susanne. 2017. *collostructions: An R Implementation for the Family of Collostructional Methods*.
- Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536.
- Gries, Stefan Th. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics* 24(3). 385–412.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending Collostructional Analysis: A Corpus-Based Perspective on “Alternations.” *International Journal of Corpus Linguistics* 9(1). 97–129.
- Hilpert, Martin. 2010. The force dynamics of English complement clauses: A collostructional analysis. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*, 155–178. Berlin, New York: De Gruyter.
- Hilpert, Martin. 2012. Diachronic Collostructional Analysis Meets the Noun Phrase: Studying MANY A NOUN in COHA. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford Handbook of the History of English*, 233–244. Oxford: Oxford University Press.
- Küchenhoff, Helmut & Hans-Jörg Schmid. 2015. Reply to “More (old and new) misunderstandings of collostructional analysis: On Schmid & Küchenhoff” by Stefan Th. Gries. *Cognitive Linguistics* 26(3). 537–547.
- Schmid, Hans-Jörg & Helmut Küchenhoff. 2013. Collostructional Analysis and other Ways of Measuring Lexicogrammatical Attraction: Theoretical Premises, Practical Problems and Cognitive Underpinnings. *Cognitive Linguistics* 24(3). 531–577.
- 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. 2003. Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Wiechmann, Daniel. 2008. On the computation of collostruction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2). 253–290.

Heinrich Heine
Universität Düsseldorf

Thanks for your

40	prompt	3
41	hard	3
42	very	3
43	word	2
44	excellent	2
45	opinion	2
46	attention	2
47	wonderful	2
48	info	2
49	assistance	2
50	consideration	2