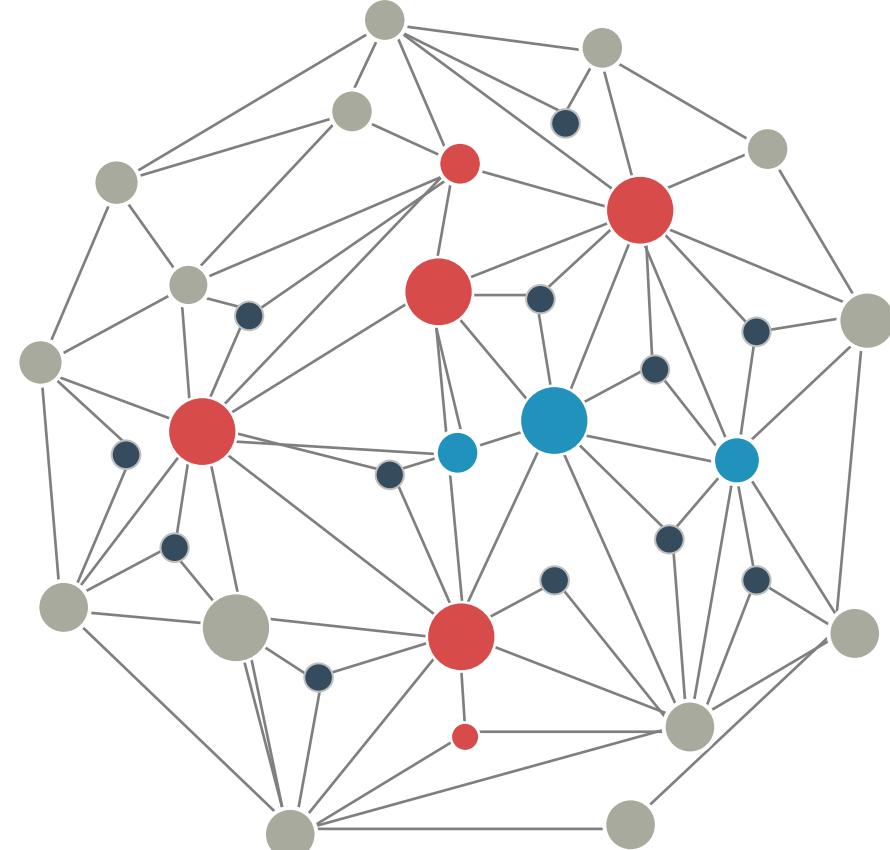


# Collocation Networks

---

**Data Visualization Using a New  
Corpus Analysis Tool #LancsBox**



Yeqing Kong | NC State University

# AGENDA

**Part 01** Data Visualization

---

**Part 02** Corpus Linguistics

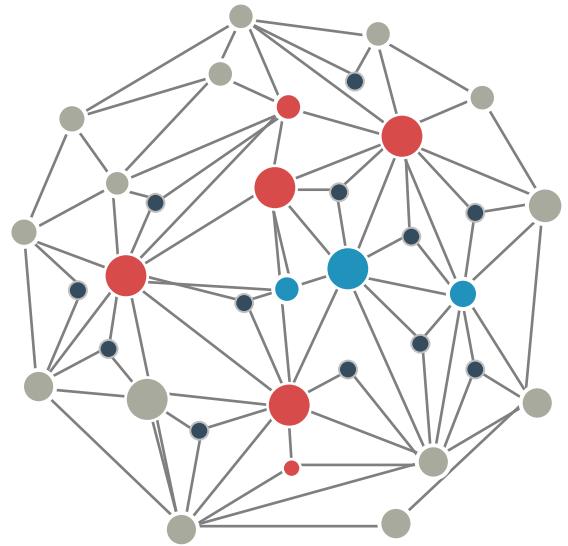
---

**Part 03** LancsBox

---

**Part 04** Q & A

---



/01

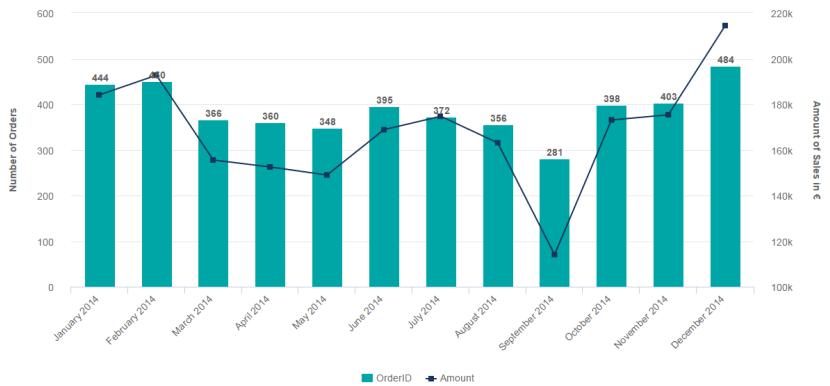
---

# Data Visualization

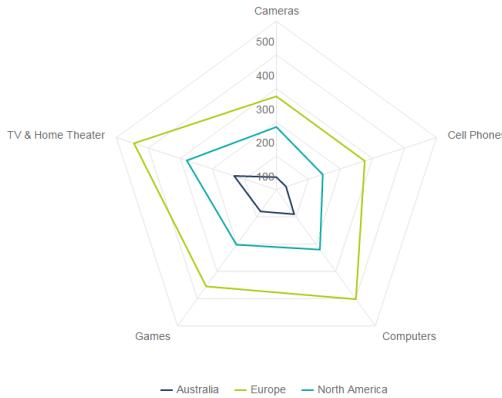
---

# What is data visualization?

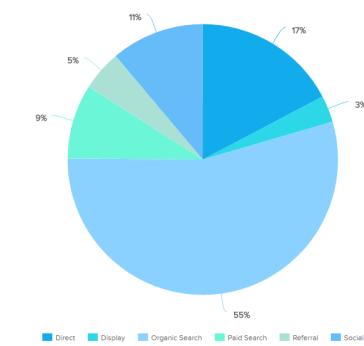
Number of Orders and Amount of Sales (last year)



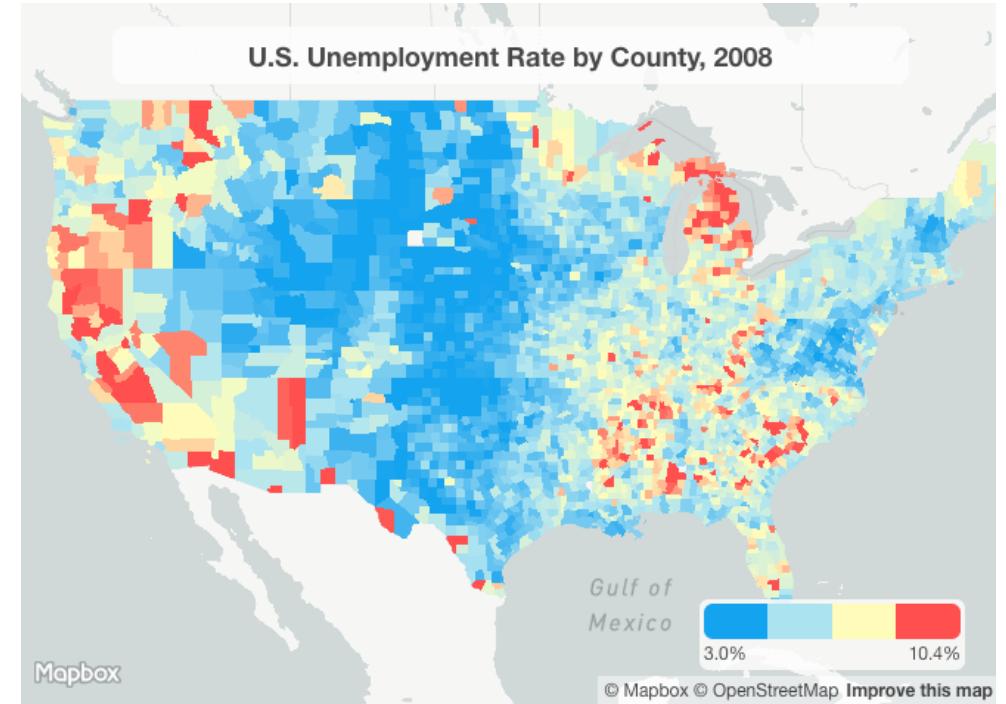
Products Sold per Category and Country (last year)



Sessions by Channel



U.S. Unemployment Rate by County, 2008



# Truth



# Beauty

# Beautiful Evidence

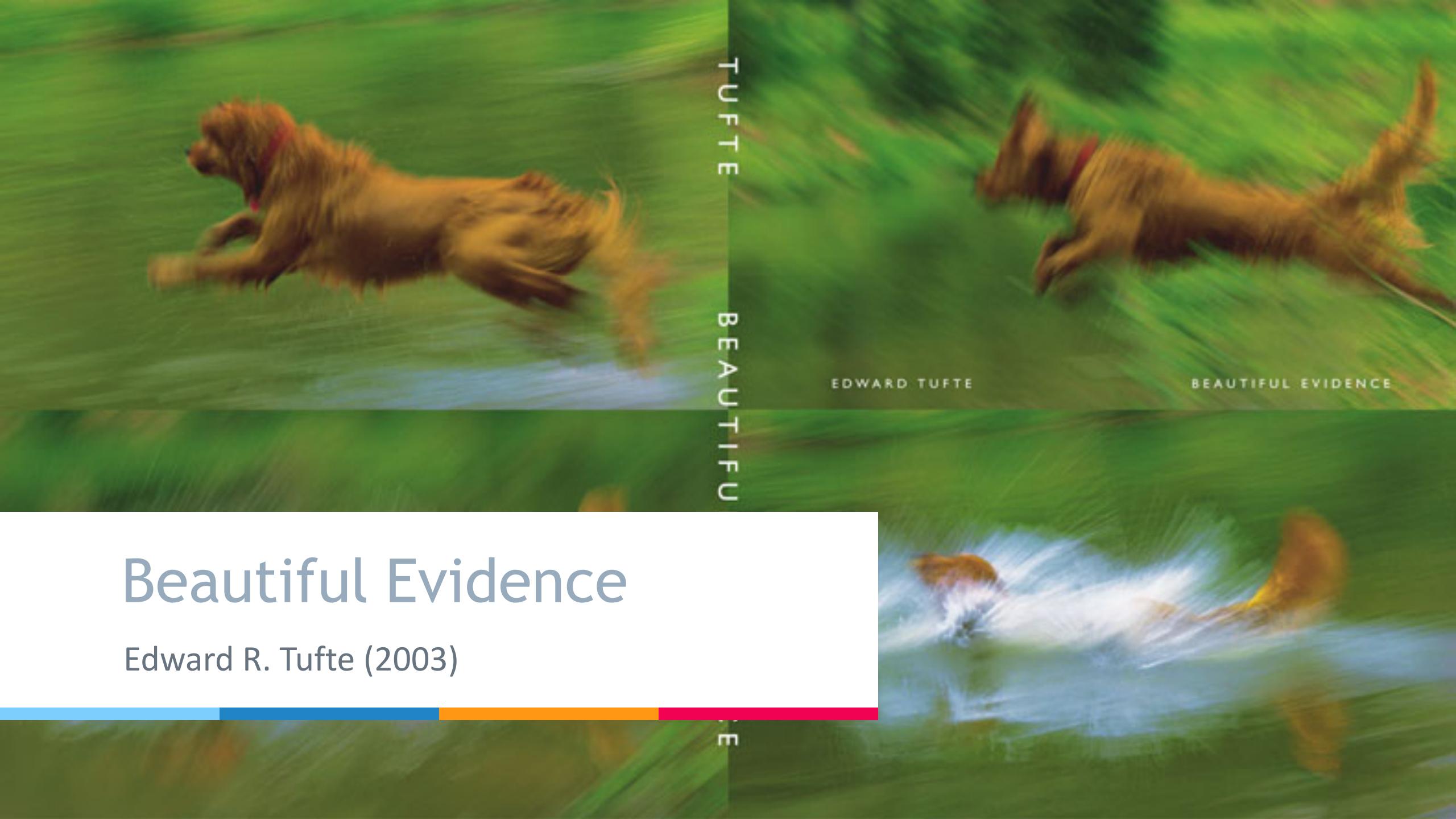
Edward R. Tufte (2003)

E

TUFT  
BEAUTIFU

EDWARD TUFT

BEAUTIFUL EVIDENCE

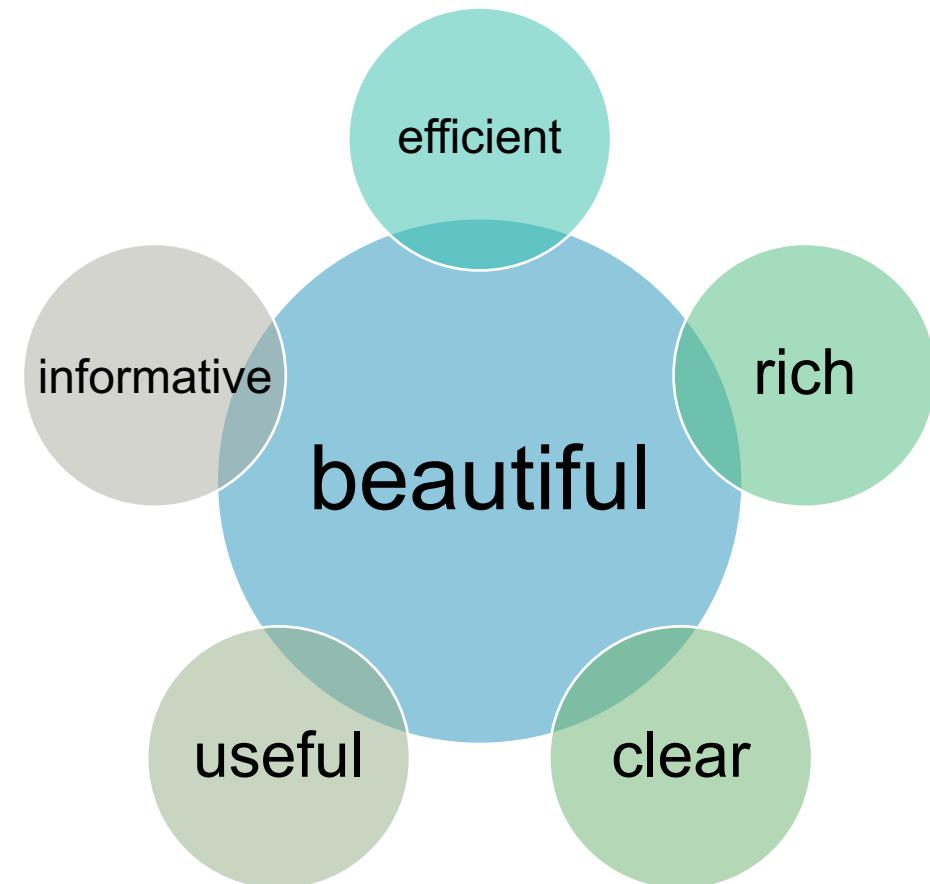


# What is “beautiful” evidence?

---

\ 'byü-ti-fəl \ adjective

*pleasing the senses or mind aesthetically*



# Principles of good data visualization

01

## Graphical Integrity

- “Show data variation, not design variation”  
(Tufte, 1983).

02

## Focus on content, comparison, and causality

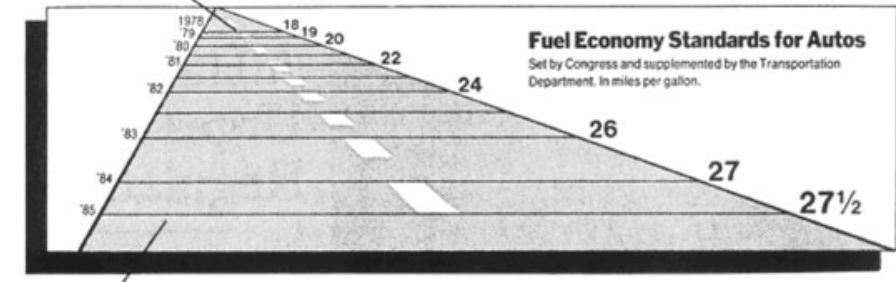
- “If picture isn’t worth a thousand words, the hell with it” (Tufte, 2013).

03

## Maximize data-ink ratio

- “Above all else show the data” (Tufte, 1983).

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

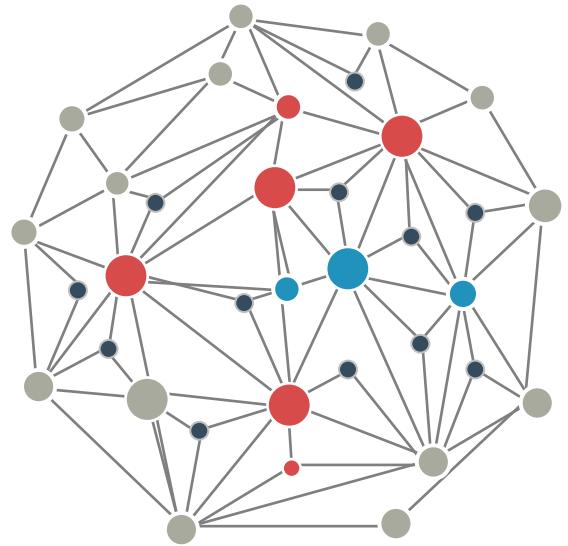


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased



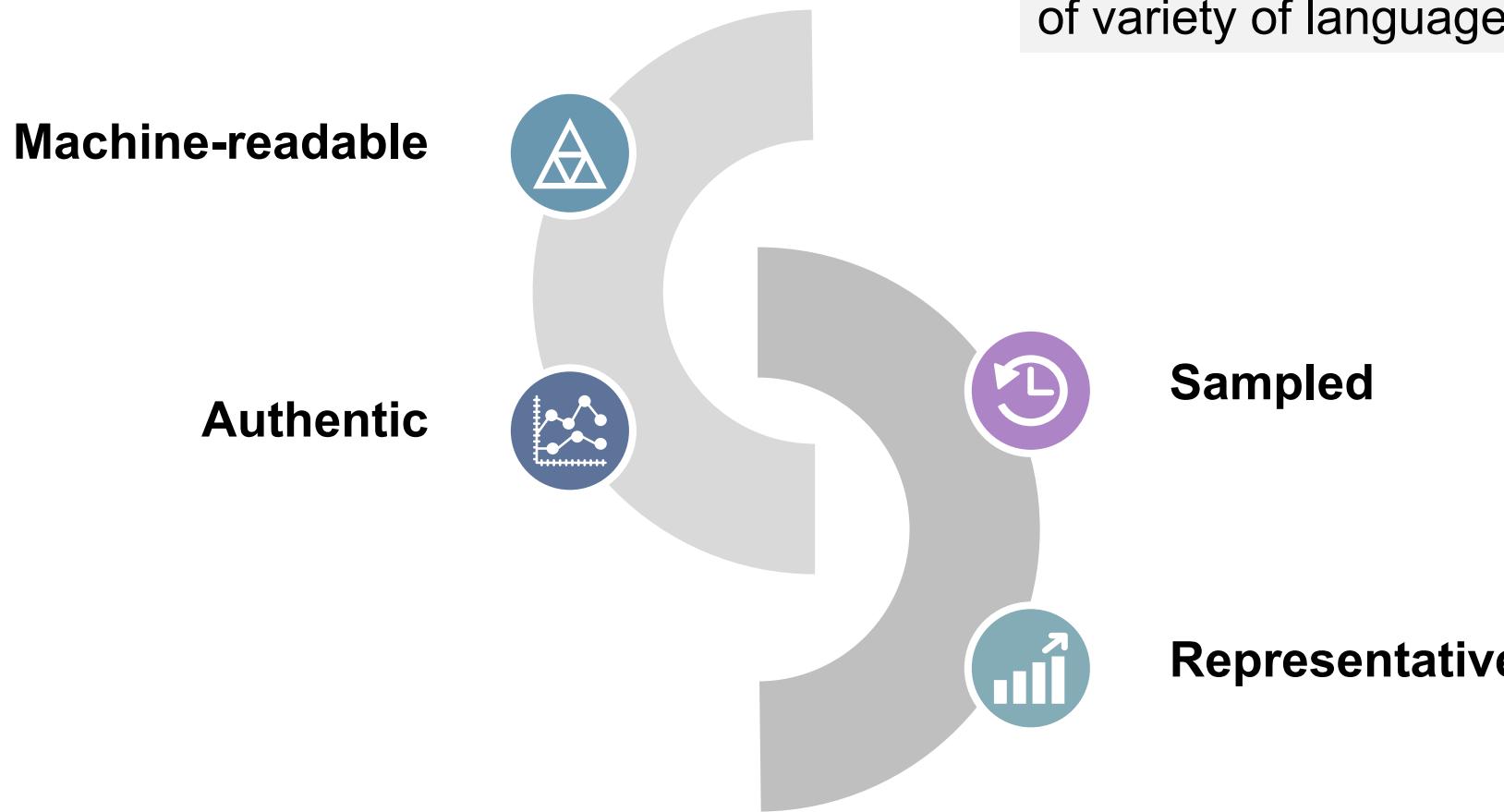
/02

---

## Corpus linguistics

---

# What is a Corpus?



A corpus is a collection of **naturally-occurring** language text, chosen to characterize a state of variety of language (Sinclair, 1991).

# Why use corpora?

---



- A corpus can be more **comprehensive and balanced**
- A corpus can show us what is **common and typical**
- A corpus can readily give us **accurate statistics**
- A corpus can store and **recall all the information** that has been stored in it
- A corpus can provide a vast number of examples in **real communication context**
- A corpus can be made **permanently accessible** to all
- A constantly updated corpus can reflect even **recent changes** in the language
- A corpus can find differences that **intuitions** alone cannot perceive

# Some important existing English corpora

- Brown Corpus
  - The British National Corpus (BNC)
  - Lancaster-Oslo/Bergen corpus (LOB)
  - Corpus of Contemporary American English (COCA)
  - Michigan Corpus of Academic Spoken English (MICASE)
  - The International Corpus of English (ICE)
  - Shakespeare's Plays



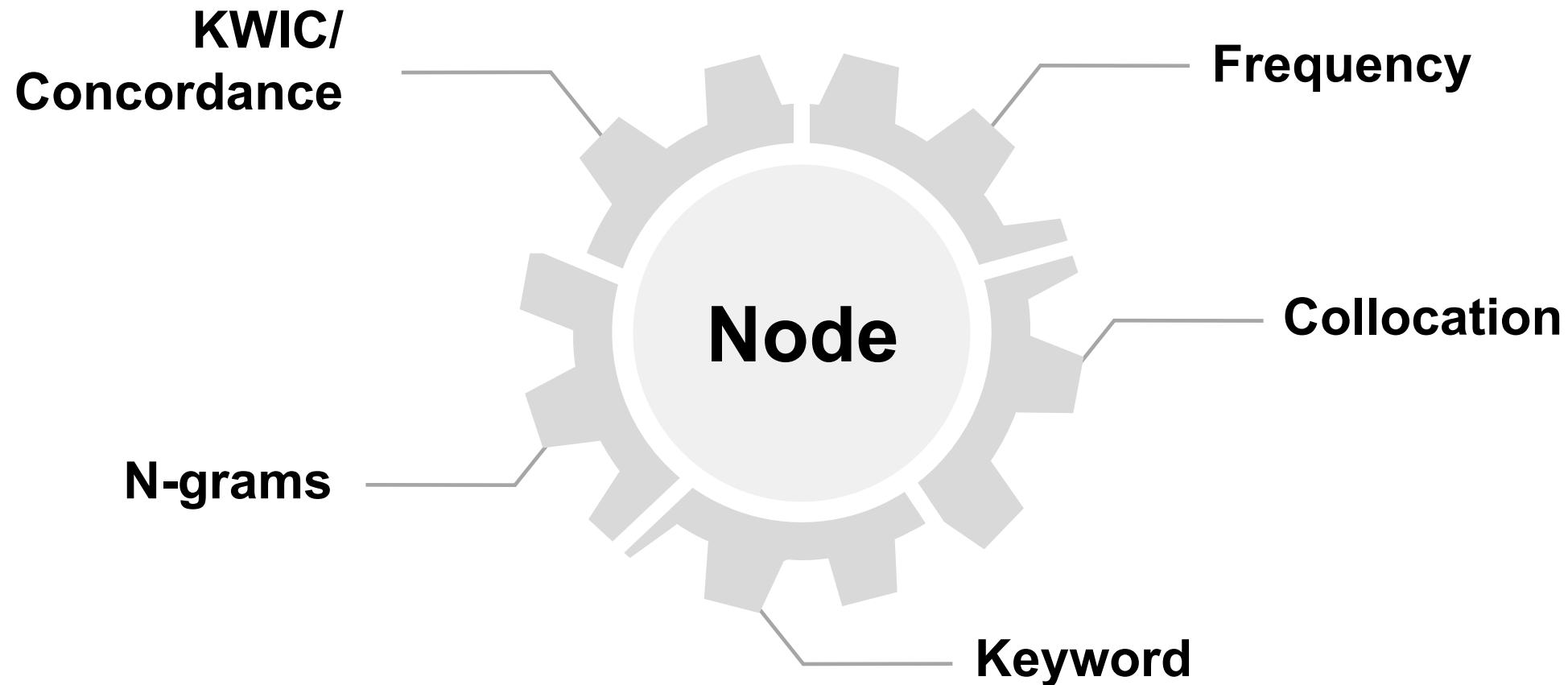
# Some important existing English corpora

- Brown Corpus
  - The British National Corpus (BNC)
  - Lancaster-Oslo/Bergen corpus (LOB)
  - Corpus of Contemporary American English (COCA)
  - Michigan Corpus of Academic Spoken English(MICASE)
  - The International Corpus of English (ICE)
  - Shakespeare's Plays



# **Key concepts in corpus linguistics**

---



## Example

---

If picture isn't worth a thousand words, the hell with it.

# Example

---

Collocation

Node

If picture isn't worth a thousand **words**, the hell with it.



Concordance

# Example

---

Collocation (*Span: 3*)

Node

If picture isn't worth a thousand words, the hell with it.

Concordance

# Measures of Collocation

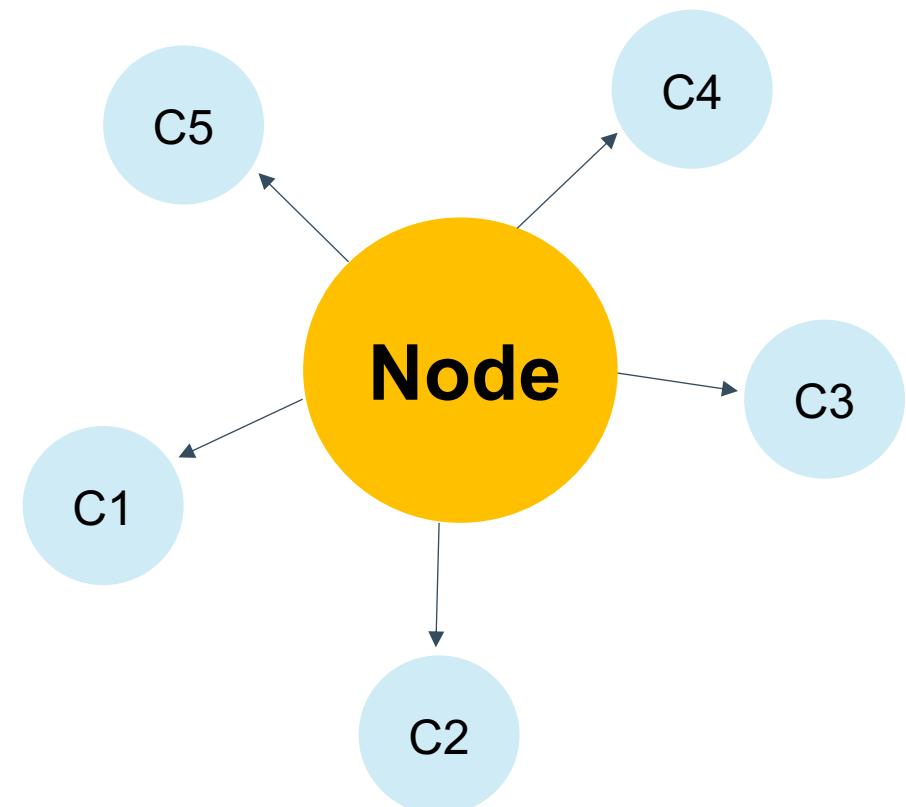
---

## Criteria for identifying collocates

- distance → span (e.g., L3, R3)
  - frequency
  - exclusivity
  - dispersion
  - directionality
  - type-token distribution
  - connectivity
- 
- statistic (e.g., MI, LL)

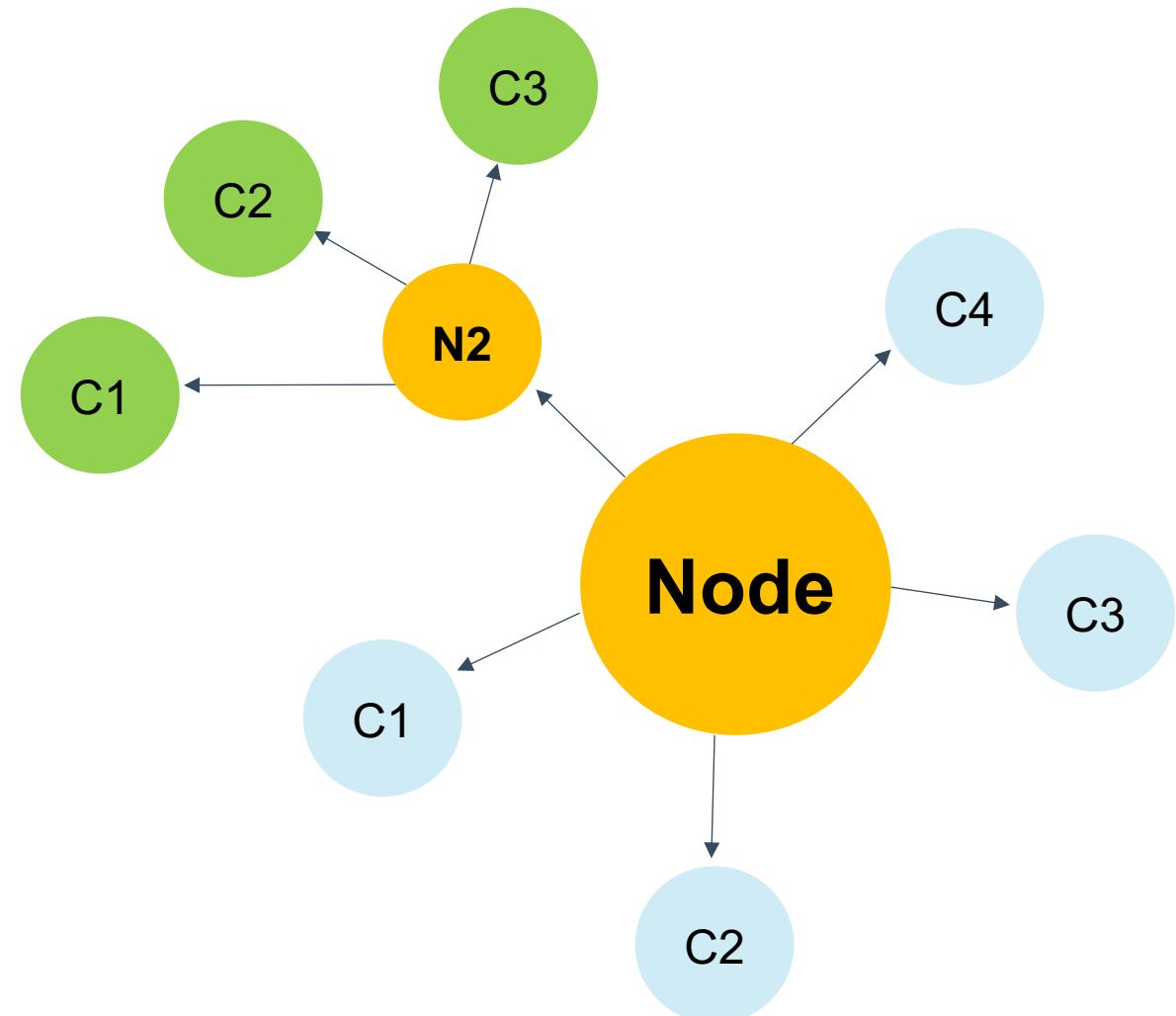
# Collocation Networks

---



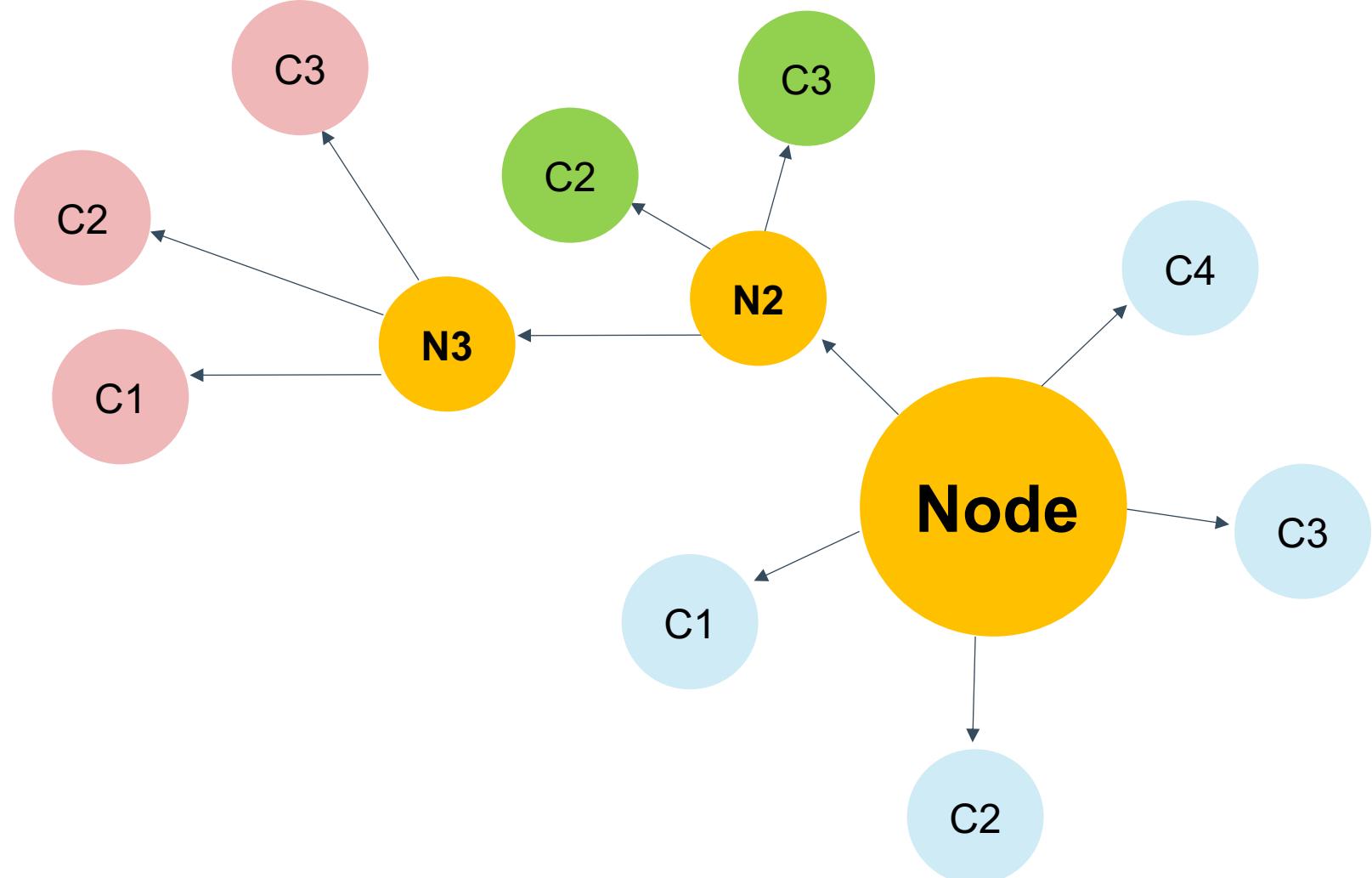
# Collocation Networks

---



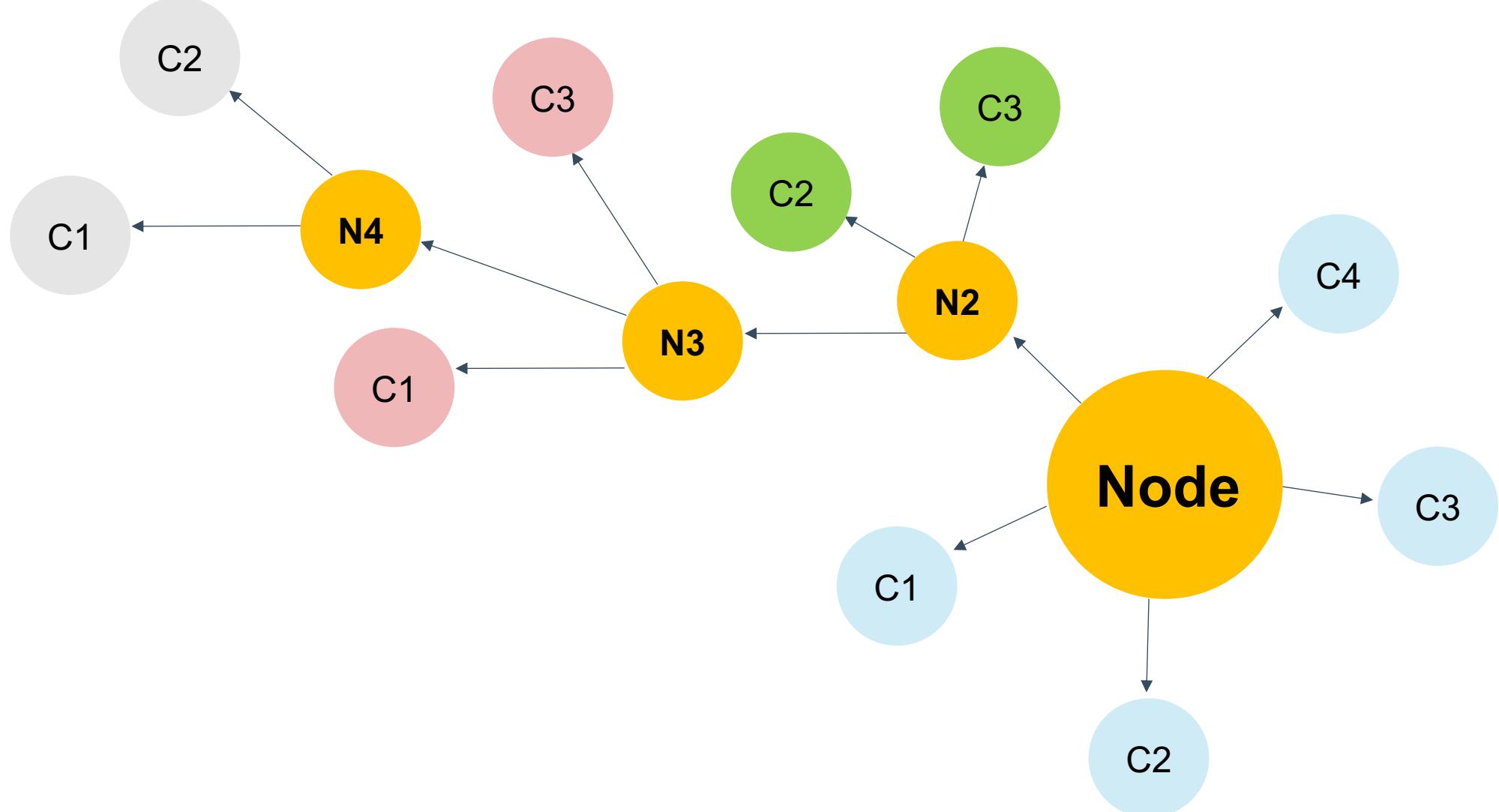
# Collocation Networks

---



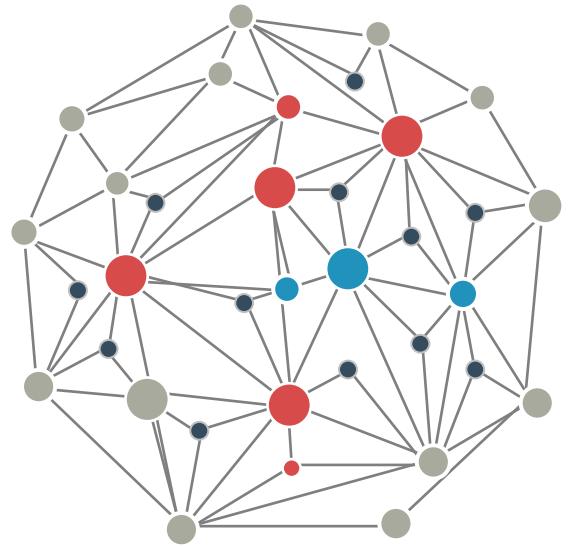
# Collocation Networks

---









/03

---

## LancsBox

---

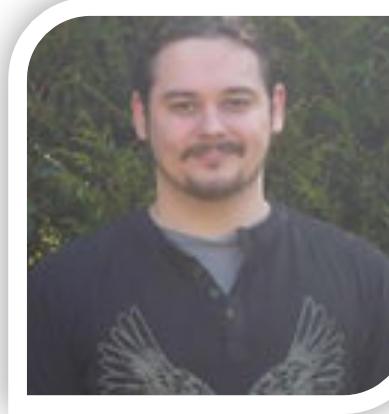
# Developer Team

Linguistics and  
English Language

Lancaster  
University



Dr. Vaclav Brezina  
(Project Lead)



Matt Timperley  
(Software Developer)



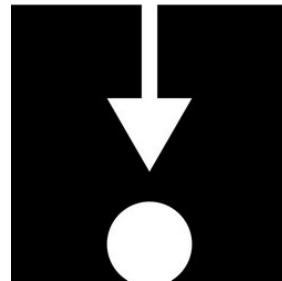
Professor Tony McEnergy  
(Adviser)

# AntConc vs. LancsBox

---

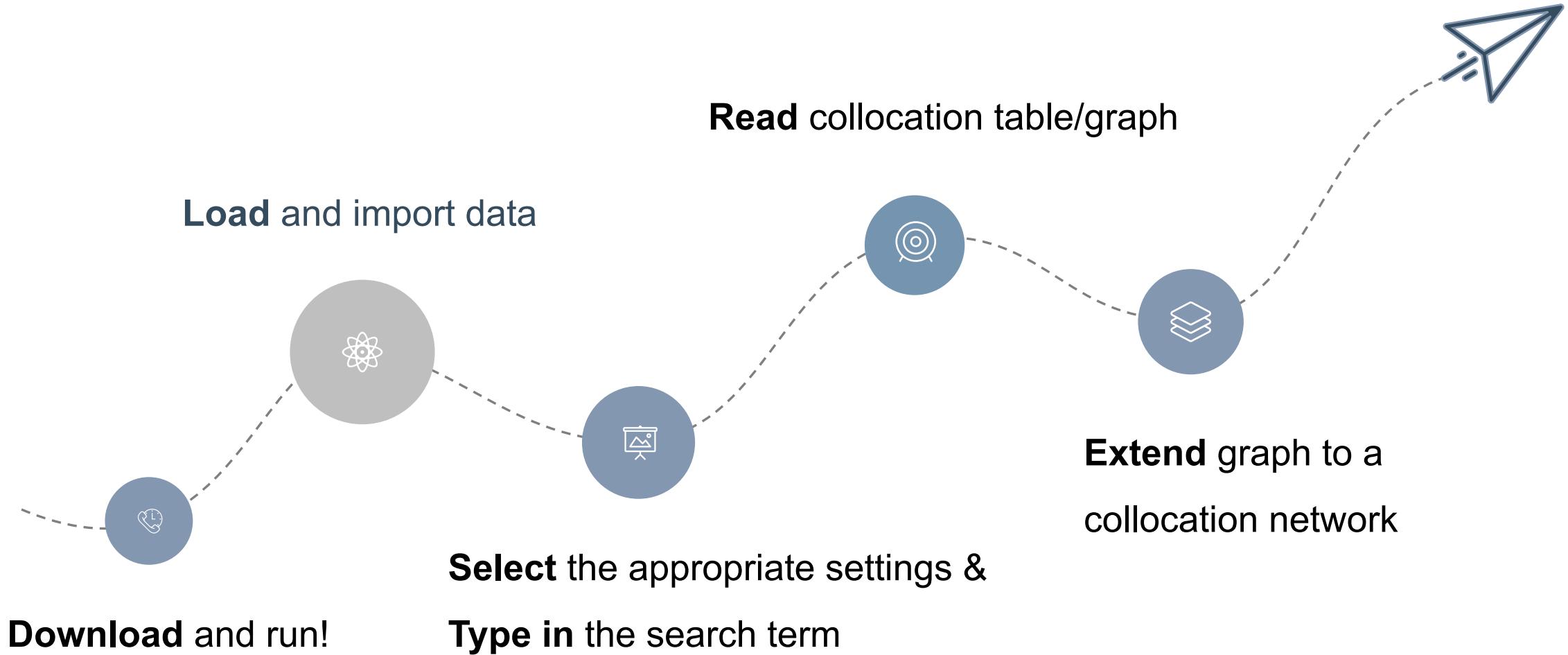


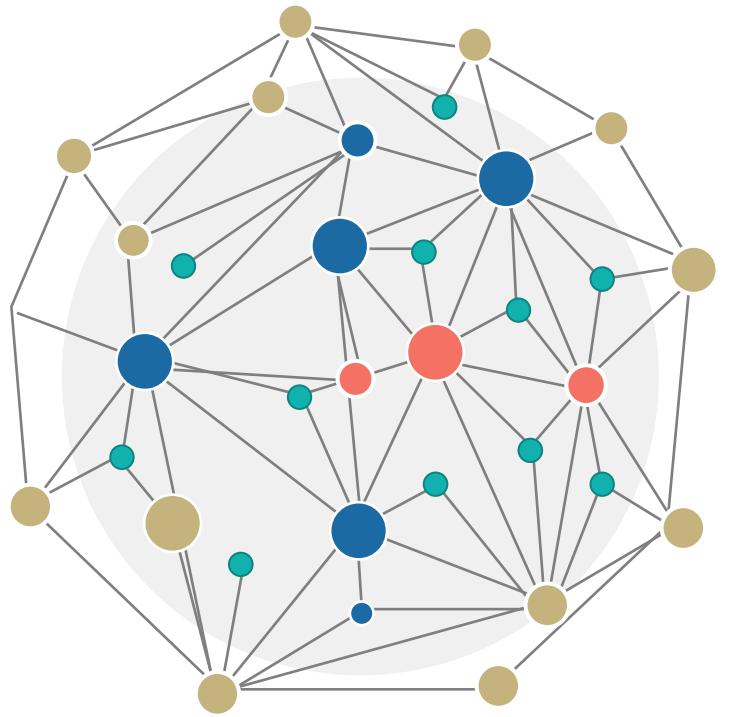
- KWIC/Concordance
- Words list/Keyword list
- Frequency
- Collocation
- N-grams



- KWIC/Concordance
- Words list/Keywords list
- Frequency
- Collocation
- N-grams
- Part-of-speech tagging
- Collocation graph/network (GraphColl)

# How to produce a collocation graph?





# LancsBox Demo

# Focus on content

The screenshot shows the #LancsBox v 3.0 software interface. On the left, there's a sidebar with 'Corpora' selected. It has sections for 'Name:' (set to 'Corpus 2'), 'Load data' (with 'Corpus' and 'Word List' options), 'Download' (with 'Corpus' and 'Word List' options), and 'Language' (set to 'English'). Below these are buttons for 'Delete' and 'Import!'.

In the center, large pink text reads 'txt, xml, pdf, doc(x), zip...'. To its right is a panel titled 'Import Options'.

On the right, a list of files is shown:

- A\_Press\_report.txt**  
Format: Text, 88805 tokens, 12068 types, 11195 lemmas
- B\_Press\_edit.txt**  
Format: Text, 54367 tokens, 7902 types, 7028 lemmas
- C\_Press\_review.txt**  
Format: Text, 34289 tokens, 7699 types, 7240 lemmas
- D\_Religion.txt**  
Format: Text, 34257 tokens, 5397 types, 4701 lemmas
- E\_Skills.txt**  
Format: Text, 76613 tokens, 10799 types, 9721 lemmas
- F\_Pop\_lore.txt**  
Format: Text, 88742 tokens, 12267 types, 10991 lemmas

At the bottom, it says 'Loaded corpus: LOB' and there are navigation icons (Σ, Ι, ?, !, †).

# Focus on content

#LancsBox v 3.0

Corpora KWL... X

Search Term {love} Occurrences 389 (3.86) Texts 15 Corpus LOB Context 17 Display Text

| Index | File           | Left  | Node   | Right   |
|-------|----------------|---|--------|---|
| 1     | A_Press_revi   | the creative idea, like a poet." The idea for" Romeo and Juliet" was the irresponsibility of young        | love   | pushed into tragedy by Shakespeare." Othello" is the" sentimental" tragedy of a cultivated, brave man who     |
| 2     | A_Press_revi   | tragedy by Shakespeare." Othello" is the" sentimental" tragedy of a cultivated, brave man who comes to    | love   | too late, and does not know what to do with it." It is a tragedy not to                                       |
| 3     | A_Press_revi   | know what to do with it." It is a tragedy not to know what to do with                                     | love." | Zeffirelli does not mention the colour of Othello's skin, but his knife traced another geometrical pattern. H |
| 4     | A_Press_revi   | American teenage magazine Mademoiselle. Christopher Logue writes fierce, noisy poems about war,           | love,  | and Logue. Son of a Southampton civil servant, he was brought up by Jesuits." I now believe                   |
| 5     | A_Press_revi   | strands on his flossy gold beard.   |        |   |
| 6     | A_Press_revi   | were shadows even for a headmistress, with the care an  |        |   |
| 7     | B_Press_edit   | apt to be hard and unbending. The diligent teacher  |        |   |
| 8     | C_Press_revage | the reasons( real enough though embroidered later) for La   |        |   |
| 9     | C_Press_revi   | days, who had travelled widely with him, been   |        |   |
| 10    | C_Press_revi   | in him. There is scarcely a woman in his life besides Te  |        |   |
| 11    | C_Press_revi   | impossible and turned out what certainly must be on   |        |   |
| 12    | C_Press_revi   | shares her room and gives her his forlorn gift of companionshi  |        |   |
| 13    | C_Press_revi   | of companionship and sympathy—" you need someone to love  |        |   |
| 14    | C_Press_revi   | and yet they want her to stay; he does  |        |   |
| 15    | C_Press_rev    | platinum-plated Hollywood carrot was dangled before his nose  |        |   |
| 16    | C_Press_rev    | known in Tin Pan Alley, likes songs with a r  |        |   |
| 17    | C_Press_rev    | Kinky bullet opened their season at Covent Garden last  |        |   |
| 18    | C_Press_rev    | invitations, the place is full of American teen   |        |   |
| 19    | C_Press_rev    | and direct play. A poet who is drunk doesn't necessarily  |        |   |
| 20    | C_Press_rev    | to play whores in order to express  |        |   |
| 21    | C_Press_rev    | in his most satisfying sculptures to date. The Subj   |        |   |
| 22    | C_Press_rev    | By PHILIP TOYNBEE THE ambiguous title reveals, by   |        |   |
| 23    | C_Press_rev    | set up by an opposition of character  |        |   |
| 24    | C_Press_rev    | but Kinglake was animated also with hatred of Napole  |        |   |
| 25    | C_Press_rev    | his absence." I can not say that  |        |   |
| 26    | C_Press_rev    | absence of the Major, and enjoy with him here-after m   |        |   |
| 27    | C_Press_rev    | This oddity in his nature appears again and   |        |   |
| 28    | C_Press_rev    | a hundred letters and post-cards between 1921   |        |   |
| 29    | C_Press_rev    | revised with theories and dreams about animals than they are with actual, living animals. Others of them  | love   | only those aspects of the animal that suit their professional interests— an extreme( and witty) example is    |
| 30    | C_Press_rev    | animal that suit their professional interests— an extreme( and witty) example is the Zoo pathologist, who | loves  | animals most when they are dead, dissection being his forte. These persons, let us say, are the               |
| 31    | C_Press_rev    | H. D.'s "Bid Me to Live, a small, handsomely-produced volume, is described as" a madrigal of war-time     | love   | and death in the London of 1917." It recounts in short, hectic and often verbless sentences the               |

Text view

Search Term ^Qlove... Occurrences 44 (12.83) Corpus LOB Text Lemma

Line Text

118 those\_DT who\_WP fear\_WD that\_IN that antonioni\_NP could\_MD never\_RB follow\_VW l'aventura\_NP with\_IN another\_DT masterpiece\_NN can\_MD rest\_VV easy\_JJ he\_PP have\_VHZ do\_VVN the\_DT near\_JJ impossible\_JJ and\_CC turn\_WD out\_RP what\_V must\_MD be\_VB one\_CD of\_IN the\_DT great\_JJS study\_NNS of\_IN the\_DT renewal\_NN of\_IN love\_NN that\_IN that the\_DT screen\_NN have\_VHZ ever\_RB see\_VN less\_RBR obviously\_RB complex\_JJ than\_IN his\_PP\$ last\_JJ film\_NN la\_NP notte\_NP will\_MD undoubtedly\_RB have\_VI more\_JJR popular\_JJ appeal\_NN but\_CC this\_DT be\_VBZ in\_IN no\_DT way\_JJ a\_DT reflection\_NN on\_IN its\_PP\$ seriousness\_JJ

119 nis\_PP\$ flood\_NN of\_IT maintain\_NN with\_IT the\_DT carra\_NP have\_VI never\_RB the\_VBN re\_BR resulting\_exchanging\_G the\_DT rock\_IS IN city\_IP the\_DT skyscraper\_NN of\_IT military\_NN but\_CC nis\_PP\$ society\_Nnve\_Vbz the\_DT same\_JJ how\_RB even\_RB clear\_JJ but\_CC touch\_VvD with\_IN a\_DJ mercifully\_JJ compassion\_NN which\_WDT be\_VBZ a\_DT strong\_JJ sign\_NN of\_IN the\_DT maturity\_NN of\_IN his\_PP\$ ultimate\_JJ artistic\_JJ vision\_NN

120 strangely\_RB enough\_RB the\_DT berlin\_NP audience\_NN receive\_WD the\_DT film\_NN with\_IN extreme\_JJ coolness\_NN much\_RB prefer\_VWG jean-luc\_NP godard's\_NP|POS disappointing\_JJ une\_NP femme\_NP est\_NP une\_NP femme\_NP a\_DT ninety-c minute\_NN hymn\_NN to\_TO vogue\_NP cahiers\_NP du\_NP cinema\_NP and\_CC the\_DT bad\_JJS aspect\_NNS of\_IN the\_DT american\_c cinema\_NN

121

122

Filtering complete

▲ ▲ ▲

|Σ| |Π| |Π|

# Focus on comparison

| Left   | Node  |
|--|---|
| es 234 (2.29)  | Texts 15  |
|  | Corpus Brown  |
|  |   |
| know what to do with it. "It is a tragedy not to know what to do with Mademoisell Christopher Logue whoes fierce, noisy puppets about whoes on his fess bold and his loud group in flue the day, headmistresses with were an administr on the school. But she has unbending. The gentle teacher out, her heart to is work because though embroidered later) for Lady Byron's leaving him, the scandal of his travelled widely with him, been fascinated by him to a point that looks like a woman in his life besides Teresa Guiccioli, last and most reasonable out what certainly must be one of the greatest studies of the renewal of his forlorn gift of companionship and sympathy—"you need someone to love—"you need someone to love you while you are looking for someone to at they want her to stay he does not know until almost too late whether he | love: Zemirelli does not "mention the colour" of Othello and Logue. Son of Southampton will serve the note - an note of the Furies in anger the e. It was never a dominant affair, it, a. I think like to think, the other affair w/ his half-sister, Augusta Leigh, the had fanned the enthusiasm which had sent who does not affect the modern reader with love that the screen has ever seen. Less obvious love you while you are looking for someone to love". Miss Dora Bryan plays the mother as a flamboyant or loathes her. Aroughly the dramatic h |
|  |   |

#LancsBox v 3.0

Corpora KWIC Whelk GraphColl Words Text

Search Term {love} Occurrences 389 (3.86) Texts 15 Corpus LOB Context 17 Display Text

| Index | File           | Left   | Node   | Right  |
|-------|----------------|--|--------|--|
| 1     | A_Press_repi   | the creative idea, like a poet." The idea for "Romeo and Juliet" was the irresponsibility of young     | love   | pushed into tragedy by Shakespeare." Othello" is the "sentimental" tragedy of a cultivated, brave man who      |
| 2     | A_Press_repi   | tragedy by Shakespeare. "Othello" is the "sentimental" tragedy of a cultivated, brave man who comes to | love   | too late, and does not know what to do with it." It is a tragedy not to  |
| 3     | A_Press_repi   | know what to do with it." It is a tragedy not to know what to do with                                  | love." | Zeffirelli does not mention the colour of Othello's skin, but his knife traced another geometrical pattern. He |
| 4     | A_Press_repte  | American teenage magazine Mademoiselle Christopher Logue who is fierce, noisy poems about wa           | love,  | and Logue. Son of a Southampton civil servant, he was brought up by Jesuits." I now believe                    |
| 5     | A_Press_repi   | strands on his household and his friends group in files one day.                                       | love,  | he wrote - can not see it unless it is there. And every tear I brush aside I find                              |
| 6     | A_Press_repi   | were shadows even for a headmistress with an administration of school. But she ha                      | loved  | the boy was never a component for the head of a community of 400 or  |
| 7     | B_Press_edit   | apt to be hard and unbending. The gentle teacher puts his heart into his work because he               | loves  | it, and that is how like to think together today doing their work. In the                                      |
| 8     | C_Press_revage | the reasons (real enough though embroidered later) for Lady Byron's leaving him, the scandal of his    | love   | affair with his half-sister, Augusta Leigh, the question of the paternity of Medora Leigh her daughter, the    |
| 9     | C_Press_revi   | days, who had travelled widely with him, been fascinated by him to a point that looks like             | love   | had fanned the enthusiasm which had sent Byron finally to Greece, and suffered years of loyal exasperati       |
| 10    | C_Press_revi   | in him. There is scarcely a woman in his life besides Teresa Guiccioli, last and most reasonable       | love,  | who does not affect the modern reader with nausea. The Countess Guiccioli was by birth a Gamba;                |
| 11    | C_Press_rev    | impossible and turned out what certainly must be one of the greatest studies of the renewal of         | love   | that the screen has ever seen. Less obviously complex than his last film, "( La Notte" will undoubtedly        |
| 12    | C_Press_revi   | shares her room and gives her his forlorn gift of companionship and sympathy—" you need someone to     | love   | you while you are looking for someone to love". Miss Dora Bryan plays the mother as a                          |
| 13    | C_Press_rev    | of companionship and sympathy—" you need someone to love you while you are looking for someone to      | love". | Miss Dora Bryan plays the mother as a flamboyant, down-to-earth sensualist who lacks perception but is         |
| 14    | C_Press_rev    | and yet they want her to stay he does not know until almost too late whether he                        | loves  | her or loathes her. Arguably the dramatist has committed a technical error in allowing Irene to sneak          |

Search Term love Occurrences 234 (2.29) Texts 15 Corpus Brown Context 7 Display Text

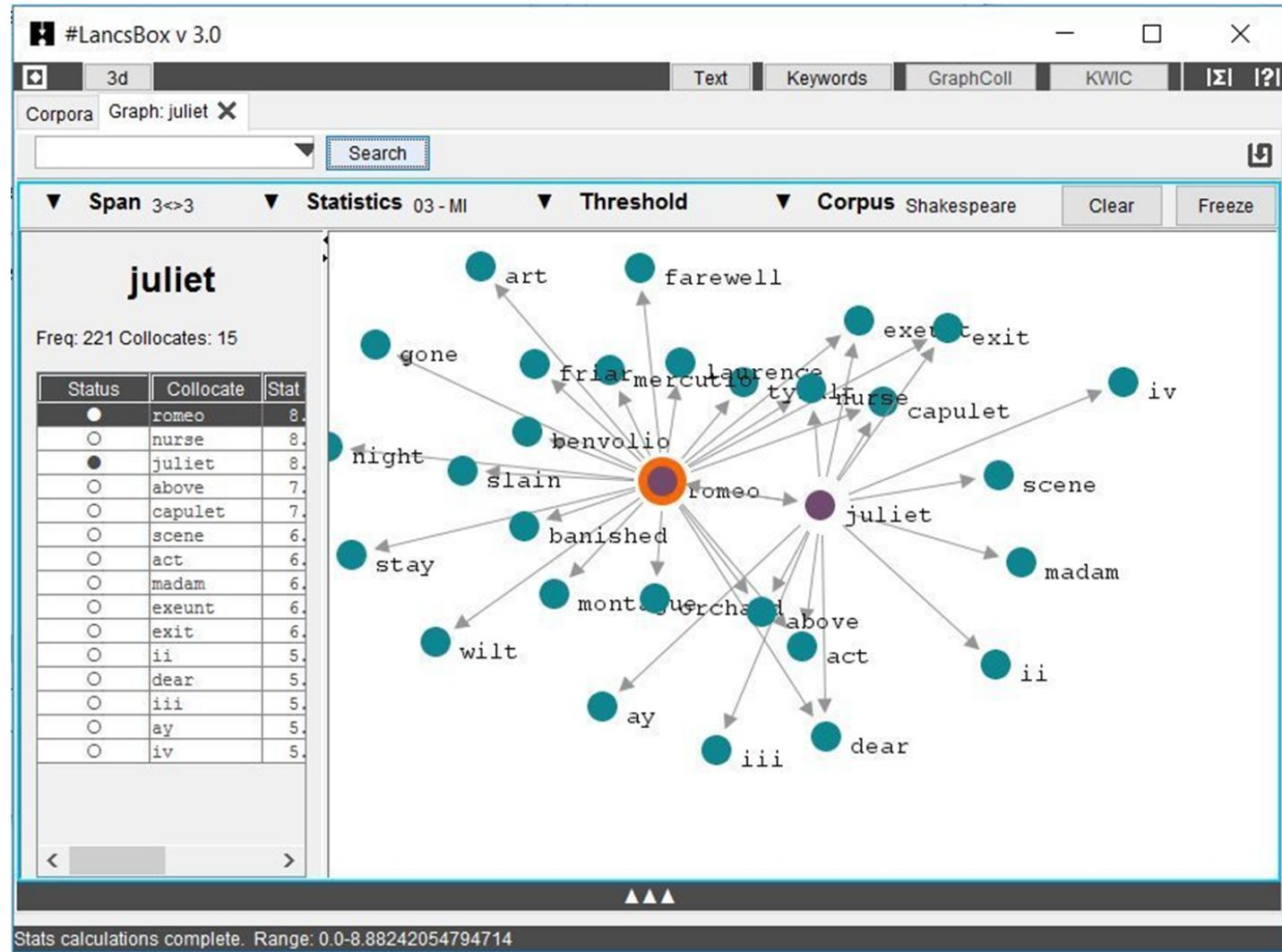
| Index | File         | Left   | Node  | Right  |
|-------|--------------|--|-------|--|
| 1     | A_Press_repi | state". Sunday he had added, "We can             | love  | Eisenhower the man, even if we considered    |
| 2     | A_Press_repi | there never is any need. The boys                | love  | it". Mrs. John Charles Cottier is chairman   |
| 3     | A_Press_repi | using puppets for lamp bases. Most children      | love  | the animated puppet faces and their flexible |
| 4     | B_Press_edit | county officials. SOMEWHERE_SOMEBODY IS BOUND TO | LOVE  | US G. Mennen Williams is learning the        |
| 5     | B_Press_edit | his past sins covered by God's                   | love. | It also implies that a man wants             |
| 6     | B_Press_edit | " infinite " how do you know you                 | love  | somebody else to be married?" he asked.      |
| 7     | B_Press_edit | ever been...oxing, regardless of being in        | love  | the usual response to include                |
| 8     | B_Press_edit | you in him, older brothers of                    | love, | is interesting. Does he care about           |
| 9     | B_Press_edit | when I can give a message of                     | love  | and comfort to other men and I               |
| 10    | B_Press_edit | be. We expect bestowal of God's                  | love  | through him. But how little love we          |
| 11    | B_Press_edit | 's love through him. But how little              | love  | we give him. The church truly is             |
| 12    | B_Press_edit | mother loves him, he clings to that              | love  | as a ballast. It motivates his behavior.     |
| 13    | B_Press_edit | needs constant reassurance of his mother's       | love  | for him and what she expects of              |

Filtering complete. ITSLIB11

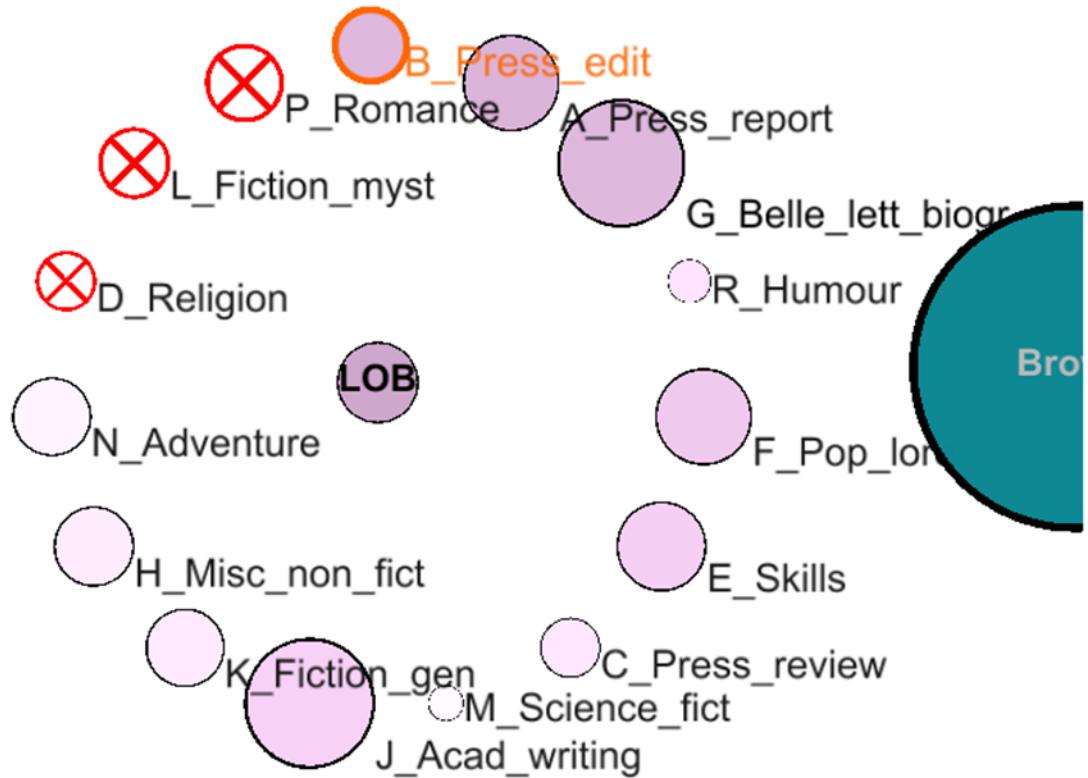
# British English

# American English

# Focus on visualization



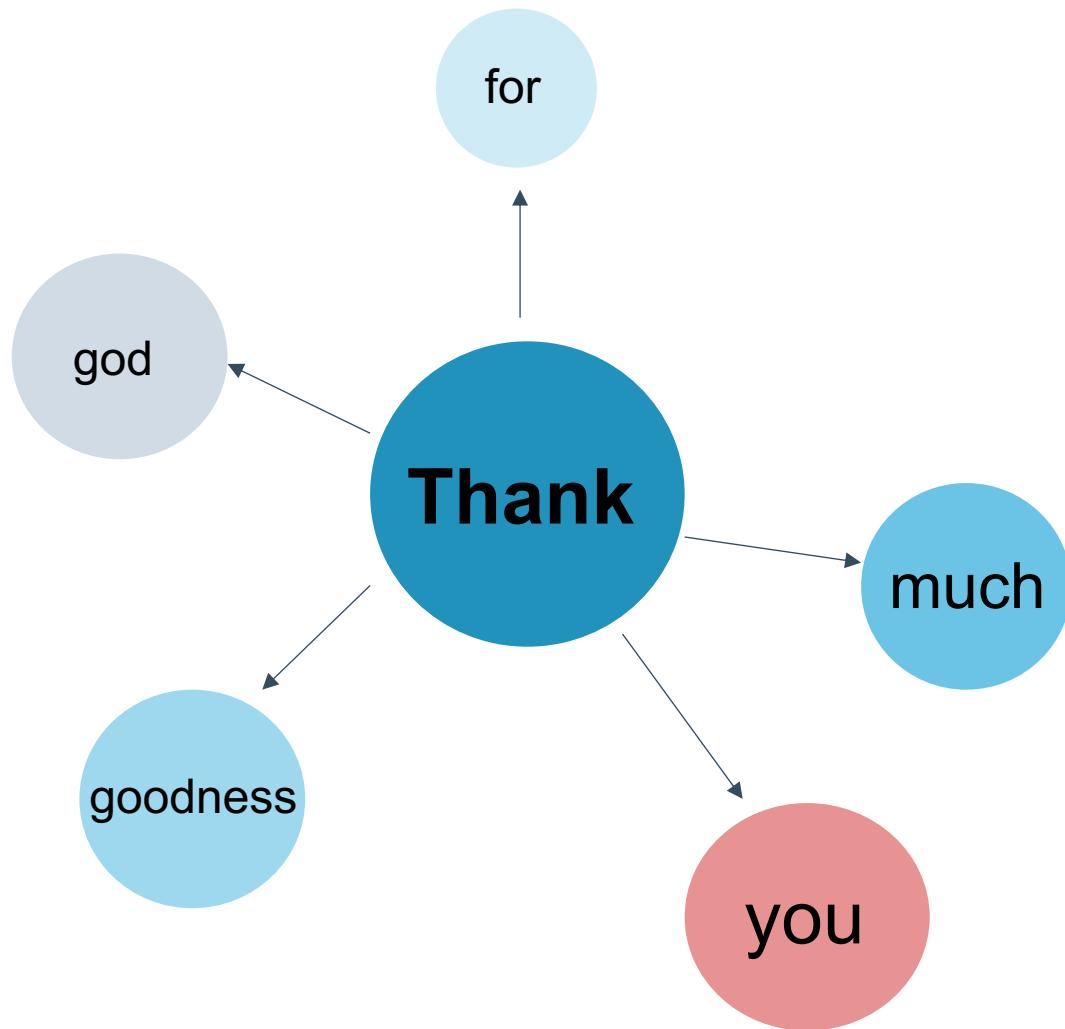
# Focus on visualization (continued)



# Some logistic issues

---

- **Where to download?** <http://corpora.lancs.ac.uk/lancsbox/>
- **How to cite #LancsBox?**
  - Brezina, V., Timperley, M., Gablasova, D., McEnery, T. (in prep.) #LancsBox: A new-generation corpus analysis tools for researchers, students and teachers.
  - Brezina, V., McEnery, T., & Wattam, S. (2015). [Collocations in context: A new perspective on collocation networks](#). *International Journal of Corpus Linguistics*, 20(2), 139-173.
- **User Guide:** [#LancsBox 4.0 manual](#)
- **More materials** (video lectures, exercises, slides etc.) are available on the #LancsBox website:  
<http://corpora.lancs.ac.uk/lancsbox/materials.php>



**Any questions?**



@YeqingKong



ykong2@ncsu.edu