

The Prime Machine: A new learner-centred corpus application as a language reference and learning tool for students.

Stephen Jeaco



As a rule, when I have ^{read or} heard some slight
indication of the ~~course of events~~, I am able to
guide myself by the thousands of other similar
cases which occur to my memory. In the
present instance I am forced to admit that the
~~facts are~~, to the best of my belief, unique.

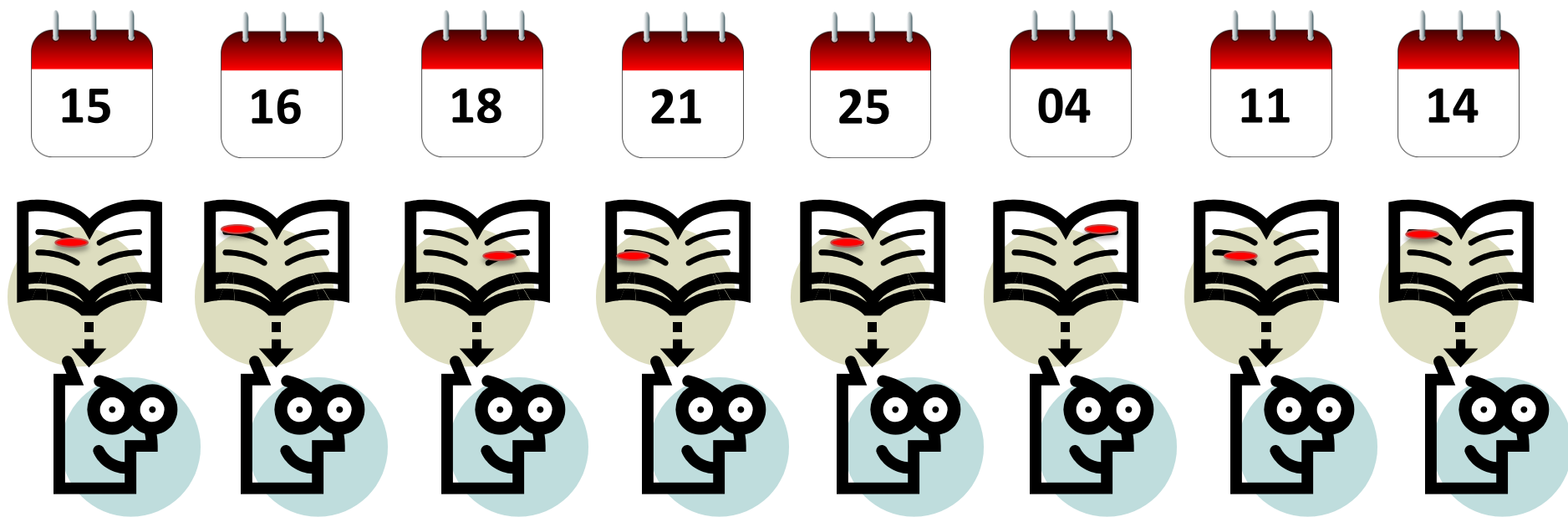
use of the word we see here is



The Adventures of Sherlock Holmes
by Sir Arthur Conan Doyle

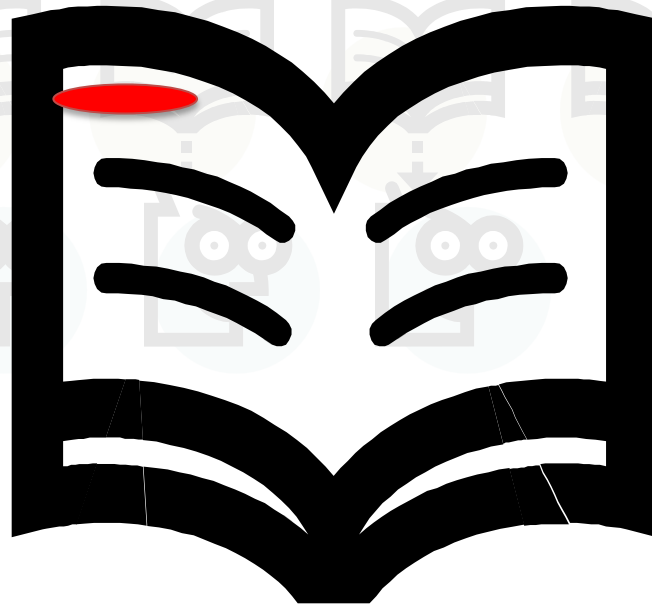
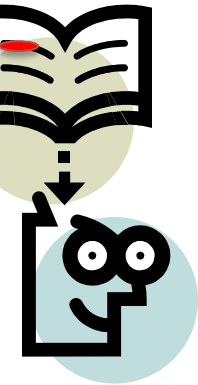
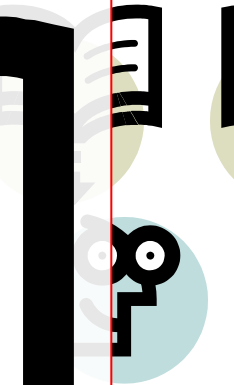
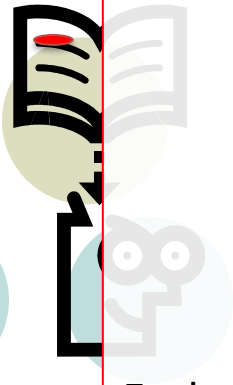
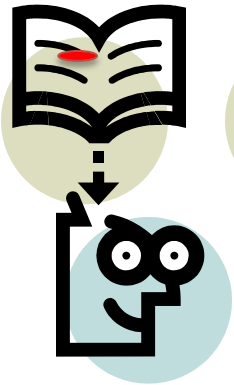
Illustration by Sidney Paget

Why use concordance lines to analyze the uses of words and phrases?



A native speaker encounters words and phrases over time, in different contexts and with slightly different uses... but over time, he or she builds up an understanding of the meaning, use and function of each word or phrase.

Why use concordance lines to analyze the uses of words and phrases?



Each time *noticing*:

- The overall context;
- Other words nearby;
- The kinds of words;
- The grammatical structures;
- Other aspects of the context and co-text

However...

- Even language experts find it hard to make a detailed list of all the knowledge about the uses of a word or phrase.
- They do not know what they know!
- Trying to write a dictionary based on *introspection* just doesn't lead to a well-balanced or comprehensive description.

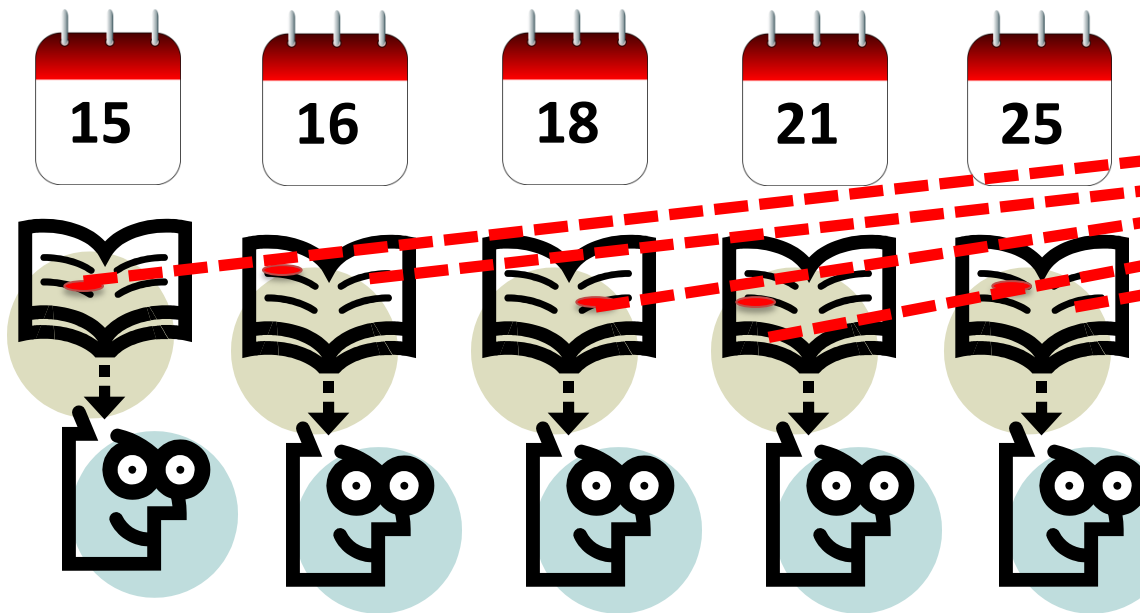
Why use concordance lines to analyze the uses of words and phrases?

A concordancer brings together hundreds (or thousands, or millions) of examples of any word or phrase, so you can look at many contexts, uses and other features all together on one screen.

The diagram illustrates the process of using a concordancer. On the left, a sequence of four calendar pages (15, 16, 18, 20) and four book icons are connected by a red dashed line. The line leads to a screenshot of a concordancer interface titled "The Prime Machine - Guest". The interface shows search results for the phrase "more sophisticated falsificationist", with concordance lines from various sources like "Humanities and Arts" and "Technology and Engineering". The interface includes navigation tools like "Search", "Cards", "Lines", "Graphs", "Collocations", "Tags", "Associates", and "Corpus Info."

Why use concordance lines to analyze the uses of words and phrases?

The LINES (also called KWIC) view allows you to see repeated patterns to the left and right.



The Prime Machine - Guest

Search Cards Lines Graphs Collocations Tags Assoc

	Text to the left of node	Node	Text to the right of node
1	nces amounts to a misrepresentation of th	sophisticated	falsificationist's position. More than a hint of this i
2	... that are inconsistent with them. These	sophisticated	falsificationist realizes the inadequacy of this and
3	police, falsification division support etc. for	sophisticated	technology of all kinds, there is clearly a much gre
4	... e cost-effective to pursue without access	sophisticated	technology./ Sorting Options ...
5	... n the decades of deliberate work, using e	sophisticated	technology, is that the creature skills develop no
6	... ve (1979) seeks to demolish the	sophisticated	analysis in his discussion of reorganization, but hi
7	... eadings without (Bampton 1970:193)./ A more	sophisticated	analysis is presented by Kenneth Newton in his st
8	... is a very extreme form of inductivism. Many more	sophisticated	inductivists would not wish to be associated with :
9	... here to that position. None of the modern, more	sophisticated	inductivists would wish to uphold the literal versio
10	... tudent), 11 November 1991, arose again, a more	sophisticated	analysis of what happened in 1837 would be requ
11	... 34000 blood tests, pretest information, and more	sophisticated	analysis and counselling in those in whom the resi
12	... ocess of accumulation. This is clearly a much more	sophisticated	analysis of the phenomena than the empirically in
13	... Employment and this makes feasible a much more	sophisticated	analysis than is possible with the published data e
14	... n to develop new forms of production using more	sophisticated	techniques and making improved products to sell
15	... (ii) the use of more	sophisticated	techniques to assess what listeners make of the s

D

Why use concordance lines to analyze the uses of words and phrases?

- Language experts use concordance lines to analyze the actual use of words and phrases, because then they can discover things about words which we are not consciously aware of.
- Corpora reveal patterns of typical and frequent usage.
- When a skilled language user goes against the normal expectations of how a word is used, this becomes *marked* or *foregrounded*, leading us as skilled readers and listeners to expect some deeper significance.

Concordancing for language learners

- Well, they can't read or listen to the same number of texts as a native speaker, but...
- ... using a concordancer means they are still getting *input* from real texts, and it speeds up the process of collecting multiple examples.
- It provides them with samples of language from specific kinds of contexts.
- And if they want to become an expert language user in an academic (or professional) field, they need to look at examples from that field!
- The idea is not to just provide a limited set of ready-made answers for them to look up and then immediately forget(!)

Concordancing for language learners

These issues relate to important theories of language acquisition:

1. Exposure:

- Learners should be exposed to target language in use (Krashen, 1989);
 - They need to listen to or read words and phrases in real contexts if they want to learn them properly!

2a. Noticing:

- Learners should consciously notice language features for effective intake (Schmidt, 1990);

2b. Actively engaging and thinking what and why:

- An important objective in language learning should be for learners to discover for themselves... (Bolitho et al., 2003; Tomlinson, 1994, 2008).
 - When you want to learn a second language, you need to be noticing patterns for yourself; so your brain gets that “aha!” moment for itself.
- When you look at extracts of language use, it is helpful for you to actively imagine the wider context and situation, as this improves learning (Tomlinson 1994, 2008).

Concordancing for language learners

What's more:

- Concordancing activities can provide opportunities to explore and uncover **patterns in language use** which may not be available from other resources;
- Concordancers can be used to **show differences** between similar words, through:
 - searches for pairs of words provided by the learners (Johns 1991);
 - searches for pairs of synonyms (Tsui 2004; Kaltenböck and Mehlmauer-Larcher 2005);
 - Etc.
- Bernardini (2004) suggests that with learners in the role of “traveller”, concordancing tasks can be used to meet a variety of **language teaching goals**;
- “... teachers need to be aware of how much **studying, learning** and **acquiring** are taking place simultaneously when learners are engaged in corpus-based guided discovery tasks” (Thomas 2015:17).

Language learning activities

Comparing results from two or more sets of concordance results can also be helpful:

- Register
 - “... the characteristics of any individual register become much more apparent when it is compared to other registers” ([Biber & Conrad, 2009, p. 8](#))

Are you ready to become a language detective?

As well as being an amazing expert in detection, Sherlock Holmes is also remembered in the stories as being an expert in disguise.

- As a detective, can you look at a hundred examples of a word or phrase and pick out some of the patterns which occur?
- As a detective, can you look at the context and co-text of examples from a corpus to see deeper meanings and typical uses of words?
- As an expert in disguise, can you dress-up your own speech or writing, so as to blend into the norms and expectations of expert language users?



The Prime Machine

A Lexical Priming Concordancer for language learners

Stephen Jeaco



You can download the software from ICE

Jeaco, S. *The Prime Machine: a user-friendly corpus tool for English language teaching and self-tutoring based on the Lexical Priming theory of language.* (2015) PhD thesis, University of Liverpool. Available from: <http://repository.liv.ac.uk/id/eprint/2014579>

Anatomy of a set of concordance lines

Text to the left of node	Node	Text to the right of node
y is generally less well developed than in England. This may be partly	due	to the fact that teacher training for further education in
less efficient than originally envisaged (Leeming 1985). This is partly	due	to the fact that reservoir construction is incomplete, mar
importance of neurophysiological knowledge of the brain. This is partly	due	to the fact that many significant computational questions
eastern b... This is partly	due	to an increase in the amount of blood circulating in her l
ernia t... This was partly	due	to the we...cal transport system and remoteness from
ly defe... 1989 was partly	due	to the campa... general practitioners against th
per cent of voters thought that if people were poor this was probably	due	to t
p between the willingness to lend and the capacity to receive is partly	due	to t
Situ... that this is partly	due	to t
ear... This was partly	due	to c
rt... This is partly	due	to the flexibility of the external skeleton — it can be moc
l ar... This was partly	due	to the West's indifference to information gathering until
al... This is partly	due	to the effects of the war, which has caused massive dar
as... emphasis is partly	due	to the background of the researchers involved (many ar
l from her mother and they were rowing a great deal. This was partly	due	to her parents' disapproval of her boy friend. In addition
all differences, which have been found in other studies, may be partly	due	to black people tend... live in areas of social deprivat
nufacturers have found it easier to apply new concepts. This is largely	due	to the fact that... types in their industry ar
creased more sharply than that of other authorities but this was partly	due	to t
: were nevertheless many women who never married. This was partly	due	to t
have been consistently higher than in the shire counties. This is partly	due	to h
their review or had made only minimal progress. This delay was partly	due	to the teachers' industrial action in the summer term of

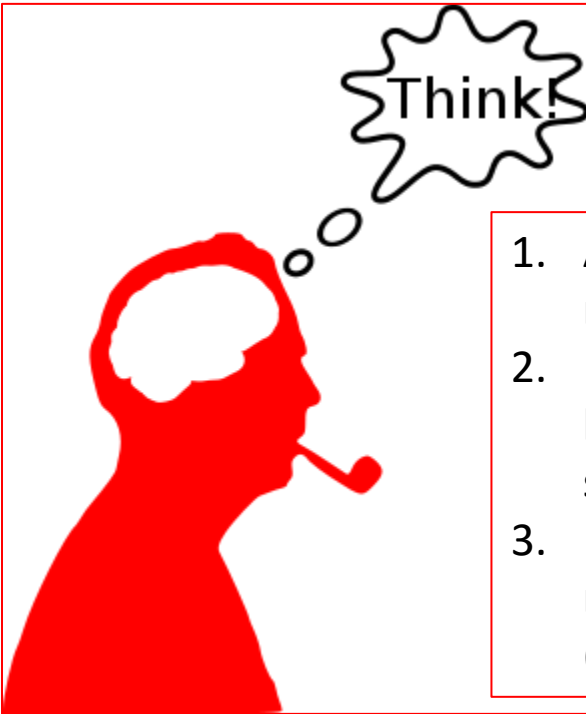
Notice the verbs to the left: *is/was/be*

Notice some common phrases: *due to the fact that*

Notice other patterns to the left: modal verbs *may*; Hedging *partly, probably*

Notice the nouns: *war, disapproval...*

Concordance lines



1. Are there any patterns in the use of particular verbs or nouns?
2. If you think about the kinds of action (verbs) or the kinds of object (nouns), do these seem to be semantically related?
3. Notice the grammatical words to the left and to the right. Are there particular prepositions? Are articles (*the, a, an*) of usually used?

Cards can help you find out more about each example

... dots indicate **not** beginning of new paragraph

Each card shows one sentence before and one sentence after.

- Position in paragraph;
- Position in sentence;
- More context.

follows a pilot study carried

Social Science

Rapid -- ESRC grant abstracts. u.p.

... Strategy in relation to a number of contextual factors--both internal and external to the firms--will come under scrutiny.

The research follows a pilot study carried out in 1985/86 and supported by the ESRC. The principal methods of investigation will be survey and semi-structured interview, culminating in selection of 24 firms for in-depth study, which will be singled out for longitudinal

...



Line containing node highlighted

Cards can help you find out more about each example

follows a pilot study carried

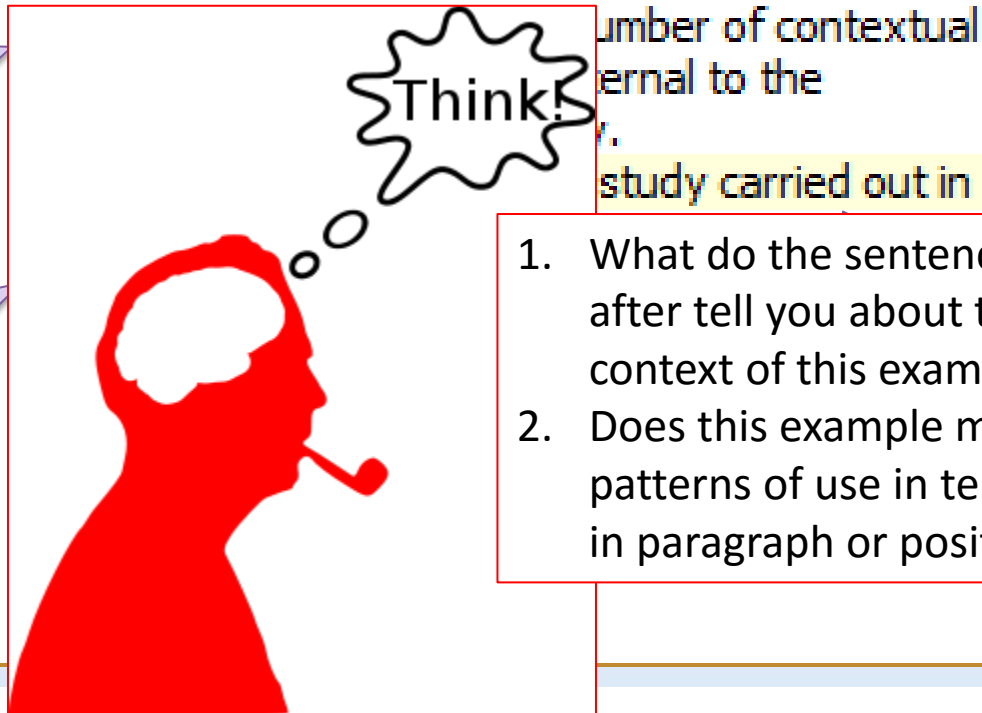
Social Science

Rapid -- ESRC grant abstracts. u.p.

... dots indicate **not** beginning of new paragraph

Each card shows one sentence before and one sentence after.

- Position in paragraph;
- Position in sentence;
- More context.



1. What do the sentences before and after tell you about the wider context of this example?
2. Does this example match any other patterns of use in terms of position in paragraph or position in text?

Cards can help you find out more about each example

Major Text Category and source

follows a pilot study carried

Caption for each card contains significant collocations

Social Science

Rapid -- ESRC grant abstracts. u.p.

... Strategy in relation to a number of contextual factors--both internal and external to the firms--will come under scrutiny.

The research follows a pilot study carried out in 1985/86 and supported by the ESRC. The principal methods of investigation will be survey and semi-structured interview, culminating in the selection of 24 firms for in-depth study, three of which will be singled out for longitudinal analysis.

...



- Collocations in KWIC;
- Major Text Category;
- Citation-style source.

Cards can help you find out more about each example

Major Text Category and source

follows a pilot study carried

Caption for each card contains significant collocations

Social Science

Rapid -- ESRC grant abstracts. u.p.

number of contextual
external to the
study carried out in

Think!



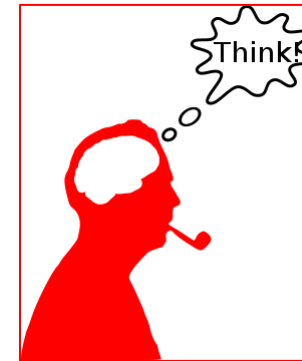
- Collocations in KWIC;
- Major Text Category;
- Citation-style source.

1. Does the Major Text Category suggest a specialized use?
2. Does this example come from the same text or a different text from the other examples?
3. Does the Major Text Category or citation give you clues about how the corpus was constructed? Is there anything to suggest you're not getting a balanced overview?
4. How does the Caption help you to spot collocations?

And remember:

According to theories of language acquisition...

... all that thinking should be helping you actually learn how to use these words!



Search support features

- Auto-complete
- Corpus-specific spell checking
- Suggestions for comparisons
- Compare mode

- Viewing results

Making queries

- Suggesting other word forms
- Suggesting similar words
- Suggesting collocations and multi-word units

The screenshot shows a search interface with the following elements:

- Navigation tabs: "Words and phrases" (selected), "Tags", and "Compare with another corpus".
- Search input: "consequence" with a "Search" button.
- Compare input: An empty input field with a "Compare" button.
- Search results (left):
 - consequences (1701)
 - consequence (1406)
 - inevitable consequence (24)
 - direct consequence (33)
 - unintended consequence (7)
 - partly .. consequence (13)
 - natural consequence (17)
- Compare results (right):
 - consequences (1701)
 - aftermath (139)
 - upshot (42)
 - event (2364)
 - moment (1271)
 - import (209)

Making queries

- Suggesting alternative combinations

Words and phrases Tags Compare with another corpus

No priming summary available. Try with |, _ or *

outcome longterm Search

Check other corpora

outcome (1340)
outcomes (419)

longterm outcome (8)
outcome longterm (>0)

Detailed description: This screenshot shows a search interface with a search bar containing 'outcome longterm'. Below the search bar, there are two lists of suggestions. The first list contains 'outcome (1340)' and 'outcomes (419)'. The second list contains 'longterm outcome (8)' and 'outcome | longterm (>0)'. The interface includes tabs for 'Words and phrases', 'Tags', and 'Compare with another corpus', and a 'Check other corpora' button.

Words and phrases Tags Compare with another corpus

No priming summary available. Try with |, _ or *

president united states Search

Check other corpora

states (5438)
statesman (61)
statesmen (58)
statesmanship (7)
statesmanlike (2)

president united states (>0)
president * united * states (>0)

Detailed description: This screenshot shows a search interface with a search bar containing 'president united states'. Below the search bar, there are two lists of suggestions. The first list contains 'states (5438)', 'statesman (61)', 'statesmen (58)', 'statesmanship (7)', and 'statesmanlike (2)'. The second list contains 'president | united | states (>0)' and 'president * united * states (>0)'. The interface includes tabs for 'Words and phrases', 'Tags', and 'Compare with another corpus', and a 'Check other corpora' button.

Cards and Lines

The Prime Machine - Guest

Search Cards Lines Graphs Collocations Tags Associates Corpus Info

	Text to the left of node	Node	Text to the right of node
1	ing personal security issues./ The study, which follows on from a	pilot	study funded by ESRC, involves the Departments of Civil Engine
2	reater level of industrial competitiveness. This research follows a	pilot	study, conducted in 1984, supported by the ESRC./ Informatio
3	al to the firms--will come under scrutiny./ The research follows a	pilot	study carried out in 1985/86 and supported by the ESRC. The pr
4	about, the advantages of renewable energy./ The findings of a	pilot	study carried out by NATTA on attitudes to the Severn Barrage i
5	ice to reconstruct the landscape in detail./ This survey follows a	pilot	study of the West Midlands in the Anglo-Saxon period in which p
6	ith two videotape material cassettes./ The project arose from a	pilot	study carried out in 1982-3 with the support of the Nuffield Four
7	: their children's language development./ The study follows on a	pilot	study in 1983-84 supported by the University of York. The prese
8	n the language development./ This study extends the work of a	pilot	study funded by ESRC (C/00/23/2220) which examined the inter
9	illy managed is to be explored./ The research, which builds on a	pilot	study carried out in 1984, involves a depth case study of a large
10	tish education system./ The current project was preceded by a	pilot	study carried out in 1980-81, supported by the former SSRC.
11	Subjects, methods, and results/ A	pilot	study was carried out to test the hypothesis that British trained
12	Methodology/ Initially a small database would be created as a	pilot	study. This database would be created by three S3 pupils from a
13	n graft thrombi; moreover, it seems to be less immunogenic. In a	pilot	study with 10 mg recombinant staphylokinase given over 30 min
14	triple therapy regimen which had previously been evaluated in a	pilot	study./ Methods ...
15	of laser palliation for advanced rectal and rectosigmoid cancer: a	pilot	study Abstract ...

follows .. a pilot study funded

Social Science

Rapid -- ESRC grant abstracts.
u.p.

... This more theoretical aspect should provide a better basis for urban designers and transport planners considering personal security issues.

The study, which follows on from a pilot study funded by ESRC, involves the Departments of Civil Engineering and Law.

T T P S S C D R

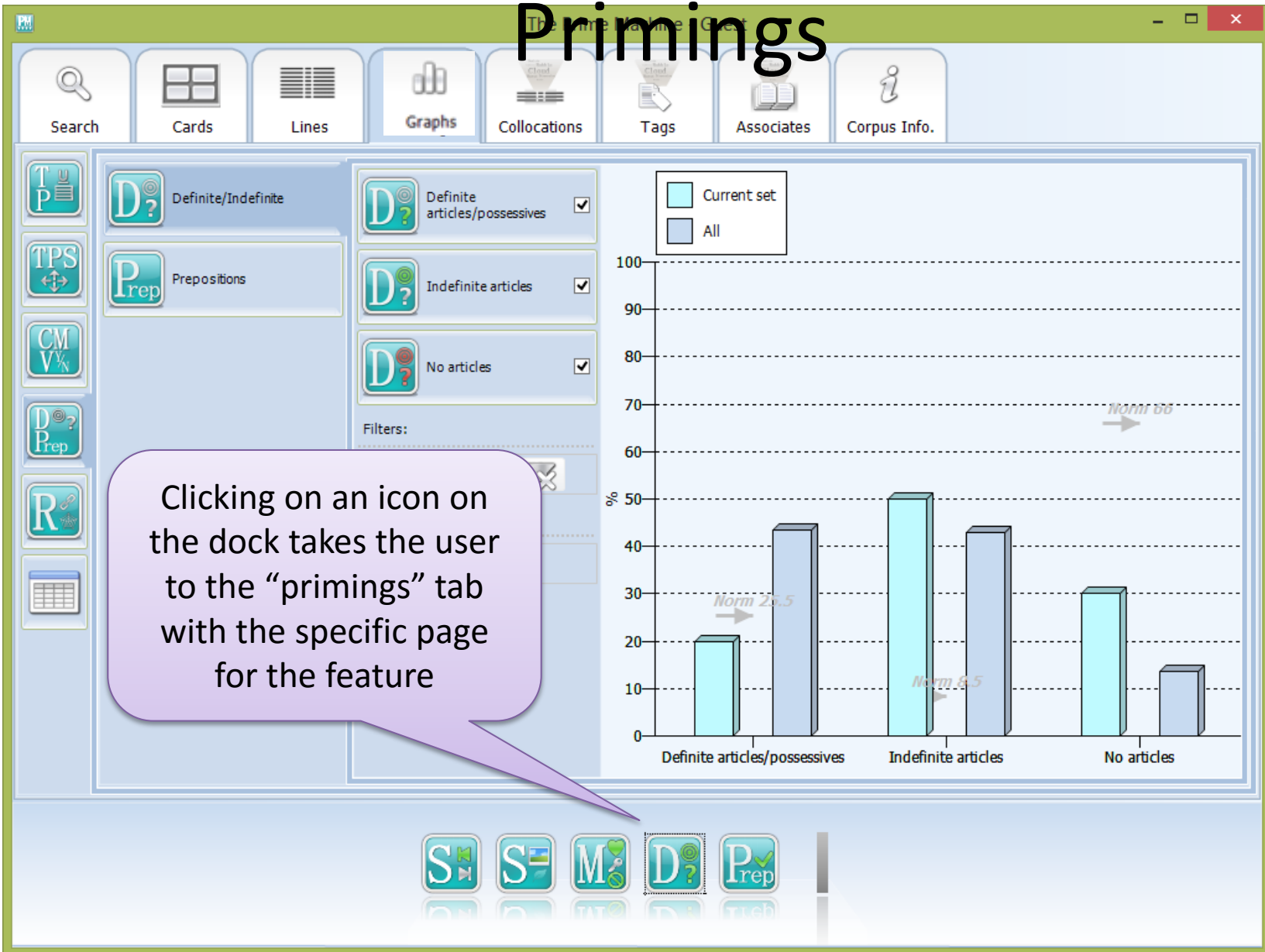
Options

43 tendencies



- Simple vs. Complex sentences;
- Position in text/paragraph/sentence;
- Modality;
- Articles & prepositions;
- Repetition;
- Voice
- Polarity (Is “not” present?)

Primings



Clicking on an icon on the dock takes the user to the "primings" tab with the specific page for the feature

TDF

- ✓ Server set up in Central Building
- ✓ Range of corpora available:
 - The British National Corpus (also split into sub-corpora)
 - Collections of academic journals
 - Spoken corpora
 - Learner corpora
 - Newspapers (from the web)
 - Literature (novels by 12 authors from 18th/19th C.)
 - Specialist corpora for ECC FYP projects

TDF

- ✓ Use of *The Prime Machine* in some ECC modules and as a tool for LC continuing support classes
- ✓ Started collecting log data on actual usage
- ✓ Analyzed some specific features through questionnaire/evaluation task (March/April 2016)

TDF

To do:

- Invite LC staff and students to use the software
- Analyze log data and additional questionnaire feedback
- Add to the range of corpora
 - Obtaining more resources
 - Building corpora through targeted web-crawling
 - Building new corpora through SURF project (?)

References

- Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15-36). Amsterdam: John Benjamins.
- Biber, D., & Conrad, S. M. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Bolitho, R., Carter, R., Hughes, R., Ivanič, R., Masuhara, H., & Tomlinson, B. (2003). Ten questions about Language Awareness. *ELT Journal*, 57(3), 251-259.
- Gabel, S. (2001). Over-indulgence and under-representation in interlanguage: Reflections on the utilization of concordancers in self-directed foreign language learning. *Computer Assisted Language Learning*, 14(3-4), 269-288.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301-319.
- Gries, S. T. (2013). 50-something years of work on collocations. *International Journal of Corpus Linguistics*, 18(1), 137-165.
- Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. In T. Johns & P. King (Eds.), *Classroom Concordancing* (Vol. 4, pp. 1-13). Birmingham: Centre for English Language Studies, University of Birmingham.
- Johns, T. (2002). Data-driven Learning: The perpetual change. In B. Kettemann, G. Marko & T. McEney (Eds.), *Teaching and Learning by Doing Corpus Analysis* (pp. 107-117). Amsterdam: Rodopi.
- Kaltenböck, G., & Mehlmauer-Larcher, B. (2005). Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL*, 17(01), 65-84.
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: additional evidence for the Input Hypothesis. *The Modern Language Journal*, 73(iv), 440-464.
- Mair, C. (2002). Empowering non-native speakers: the hidden surplus value of corpora in Continental English departments. In B. Kettemann, G. Marko & T. McEney (Eds.), *Teaching and Learning by Doing Corpus Analysis* (pp. 119-130). Amsterdam: Rodopi.
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129-158.
- Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stevens, V. (1995). Concordancing with language learners: Why? When? What? *CAELL Journal*, 6(2), 2-10.
- Sun, Y.-C. (2003). Learning process, strategies and web-based concordancers: a case study. *British Journal of Educational Technology*, 34, 601-613.
- Thomas, J. (2015). *Discovering English with Sketch Engine: Versatile*.
- Tomlinson, B. (1994). Pragmatic awareness activities. *Language Awareness*, 3(3-4), 119-129.
- Tomlinson, B. (2008). Language acquisition and language learning materials. In B. Tomlinson (Ed.), *English Language Learning Materials: A Critical Review* (pp. 3-13). London: Bloomsbury Publishing.
- Tsui, A. B. M. (2004). What teachers have always wanted to know - and how corpora can help. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 39-61). Amsterdam: John Benjamins.
- Yeh, Y., Liou, H.-C., & Li, Y.-H. (2007). Online synonym materials and concordancing for EFL college writing. *Computer Assisted Language Learning*, 20(2), 131-152.