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~~ *use and context of a word or phrase*, 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.
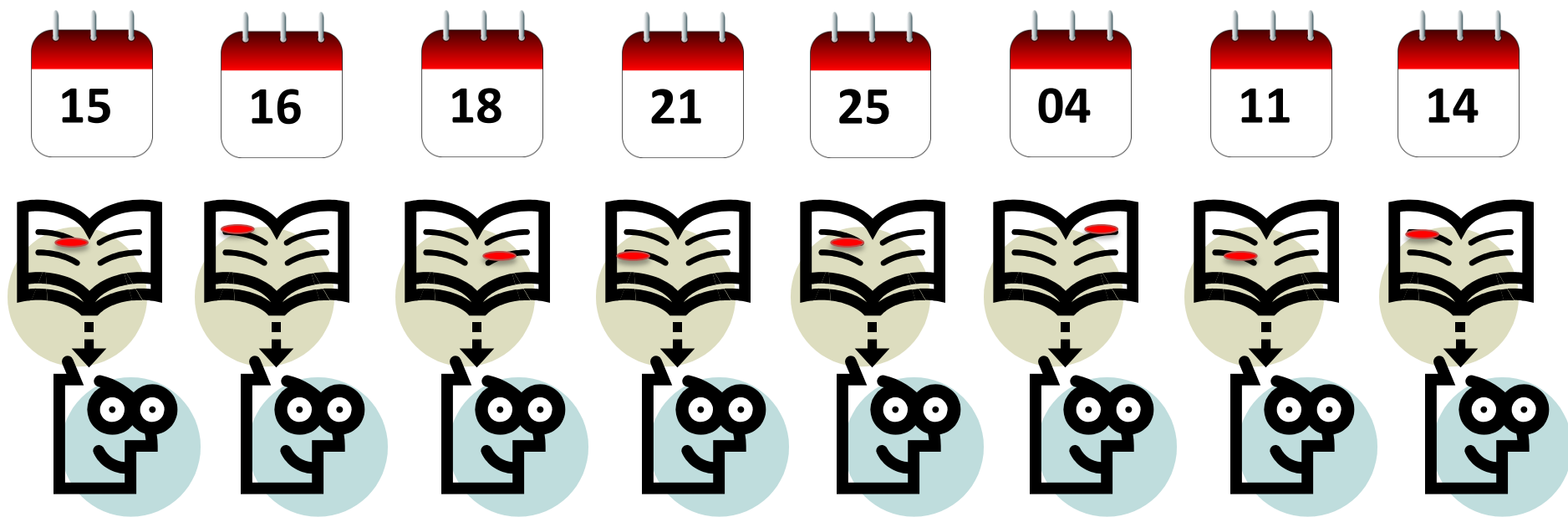
use of the word we see here is

*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?

15 **16** **18** **21** **25** **04** **11** **14**
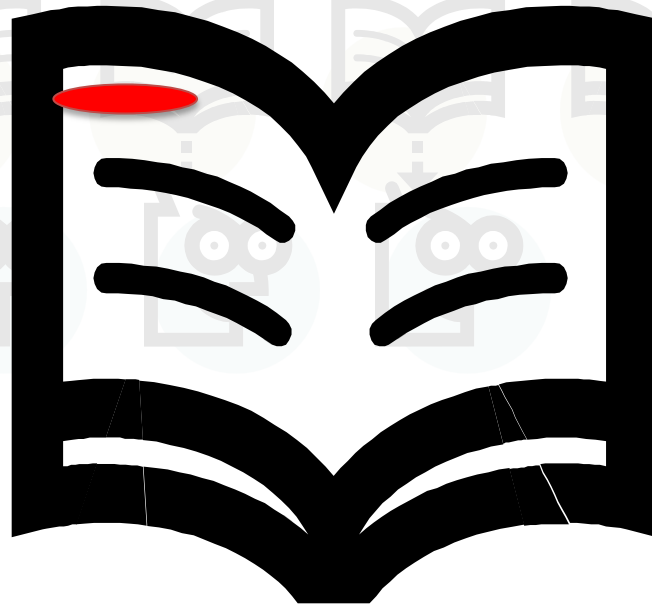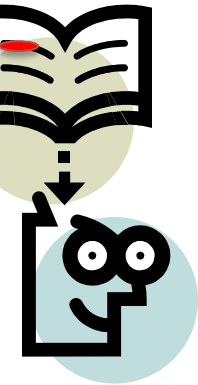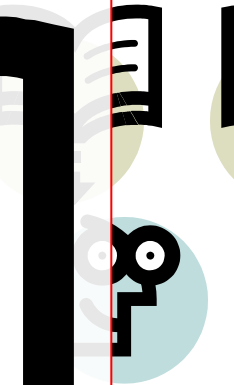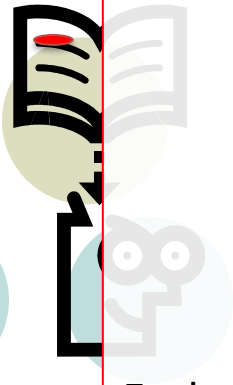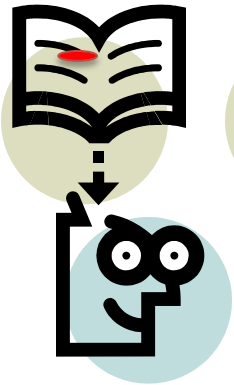
Each time *noticing*:
- The overall context;
- Other words nearby;
- The kinds of words;
- The grammatical structures;
- Other aspects of the context and co-text

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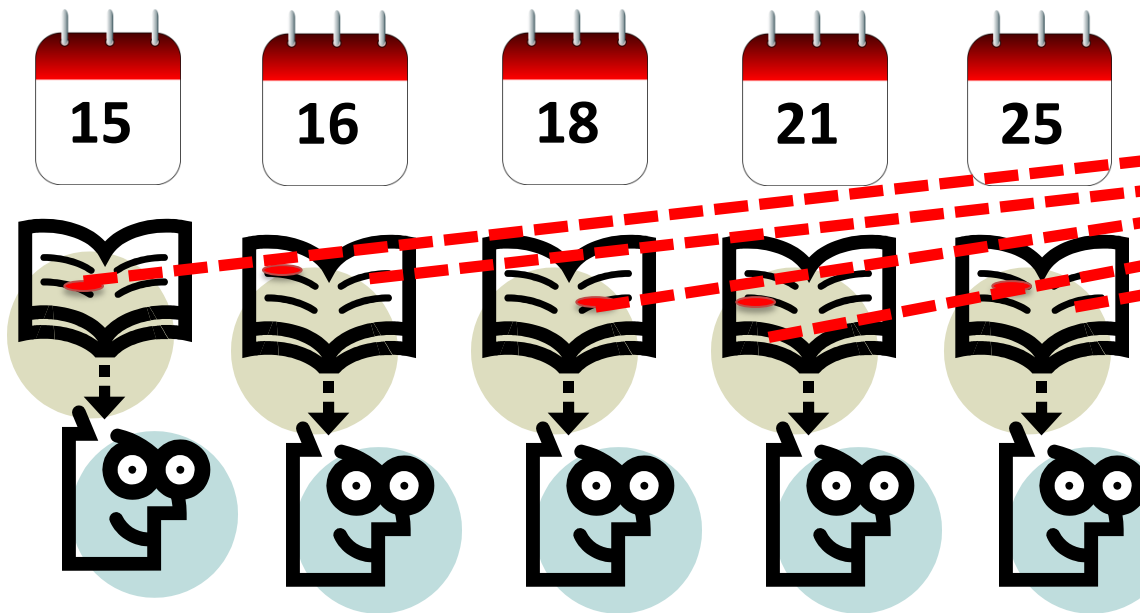
Why use concordance lines to analyze the uses of words and phrases?

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.

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.

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 t...
...e less efficient than originally envisaged (Leeming 1985). This is partly	**due**	to the fact that reservoir construction is incomplete, ma...
...ortance of neurophysiological knowle... ...the brain. This is partly	**due**	to the fact that many significant computational questions...
...easts b... ...es. This is partly	**due**	to an increase in the amount of blood circulating in her t...
...erniia t... ...This was partly	**due**	to the we... ...cal transport system and remoteness fron...
...y defe... ...1989 was partly	**due**	to the campa... ...neral 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 capaci... ...receive is partly	**due**	to t...
...Situ... ...that this is partly	**due**	to t...
...ear... ...This was partly	**due**	to co...
...ert ... ...y. This is partly	**due**	to the flexibility of the external skeleton — it can be mod...
...h an... ...This was partly	**due**	to the West's indifference to information gathering until ...
...al ... ...l. This is partly	**due**	to the effects of the war, which has caused massive dar...
...n as... ...phasis is partly	**due**	to the background of the researchers involved (many ar...
...d 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... ...o live in areas of social depriva...
...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...
...e 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...
...heir 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
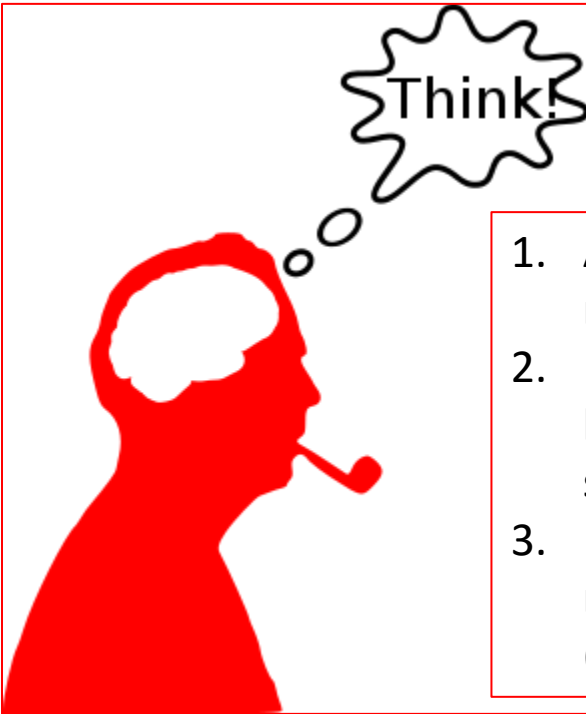
1. Are their any patterns in the use of particular verbs or nouns?
2. If you think about the kinds of action (verbs) or the kinds of object (nouns), do these seem to be semantically related?
3. Notice the grammatical words to the left and to the right. Are there particular prepositions? Are articles (*the*, *a, an)* of usually used?

Cards can help you find out more about each example

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-stuctured interview, culminating in selection of 24 firms for in-depth study, which will be singled out for longitudinal
…

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

Each card shows one sentence before and one sentence after.

Line containing node highlighted

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

Cards can help you find out more about each example

follows a pilot study carried

Social Science

Rapid -- ESRC grant abstracts. u.p.

...umber of contextual ...ernal to the ...

...study carried out in

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

Each card shows one sentence before and one sentence after.

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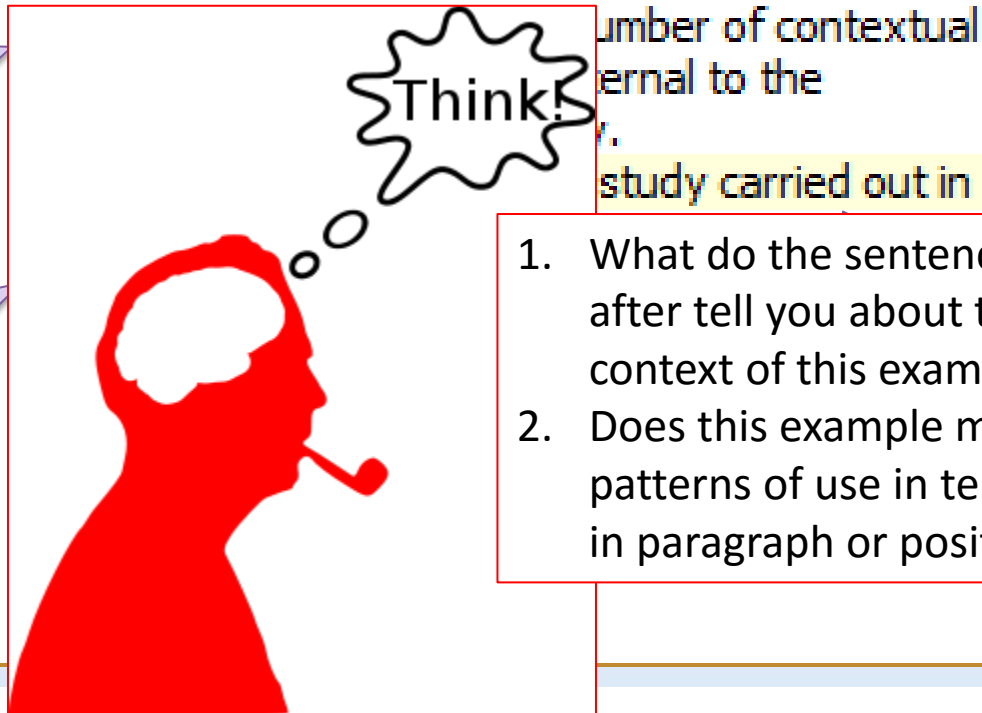
1. What do the sentences before and after tell you about the wider context of this example?
2. Does this example match any other patterns of use in terms of position in paragraph or position in text?

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

Cards can help you find out more about each example

Major Text Category and source

Caption for each card contains significant collocations

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-stuctured 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

Caption for each card contains significant collocations

follows a pilot study carried

Social Science

Rapid -- ESRC grant abstracts. u.p.

Think!

umber of contextual
ernal to the

study carried out in

- 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 <u>thinking</u> should be helping you actually learn how to use these words!
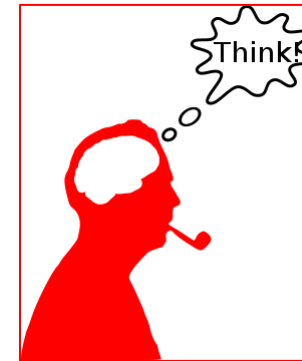
Search support features

- 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

Making queries

- Suggesting alternative combinations

Cards and Lines

43 tendencies

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

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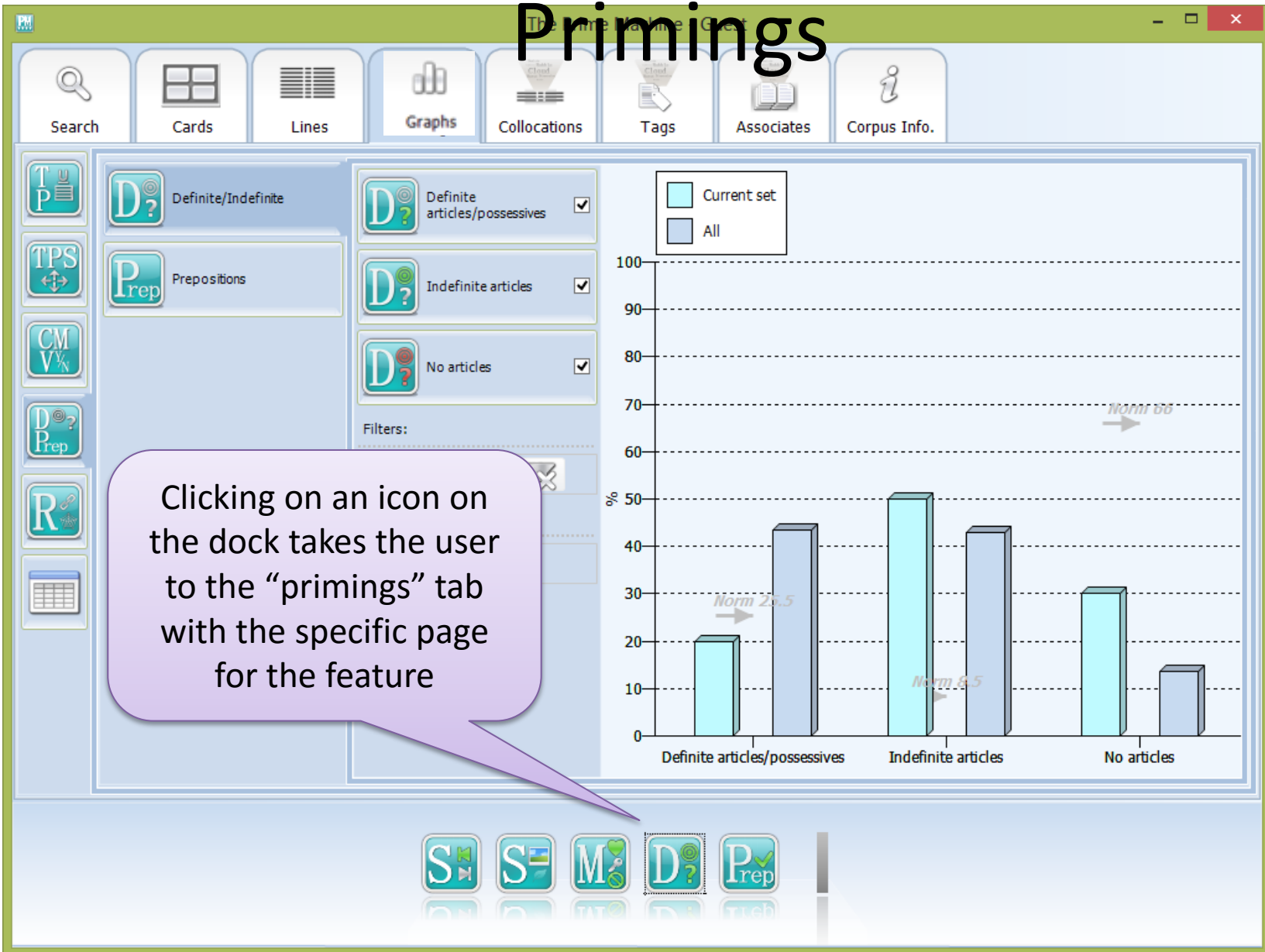
Clicking on an icon on the dock takes the user to the "primings" tab with the specific page for the feature

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. McEnery (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. McEnery (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.