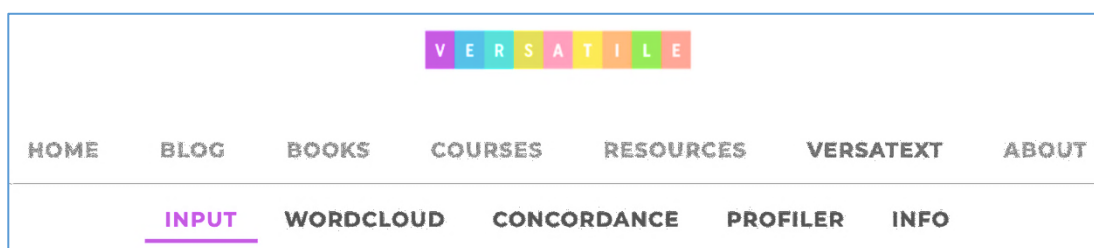


Unit 2 Meet VersaText

VersaText is an online tool that helps you explore the language of a text. It helps you focus on the vocabulary in a variety of ways. A lot of the grammar of vocabulary and other important language patterns can be observed in a text using VersaText, especially when you know what to look for.

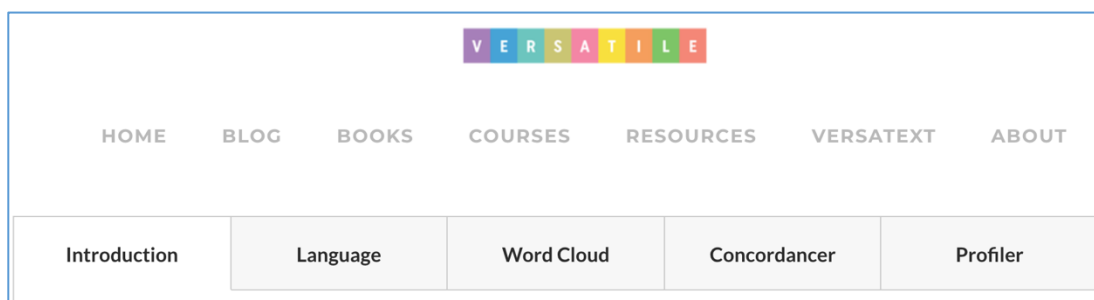


Go to www.versatile.pub and click the **VersaText** button in the navigation bar. This takes you to the **Input** page where you will see the message: *Paste in your own text or click on a preloaded text title below.* Do that. Click on the **Word cloud** button and you will see a word cloud of the text. Hover over a word in the word cloud and when it turns red, **Click on it** to see it in a concordance. Click on the **Profiler** button to see text statistics and word lists that are descriptions of the chosen text.

That is most of what you need to know about clicking in VersaText. The rest of this chapter provides background information about the program and foreground information about using it to learn about language and about English. Remember that you are free to skip sections and tasks that you don't find interesting at the moment. You are just as free to return to them later.

Info online

Click on the **Info** button (see screenshot above) to see this row of five tabs. Each one contains a page of information about various aspects of VersaText.



Texts

There are some pre-loaded texts listed under the Input field. Click on a title to set the text in the input field. The texts are described under Info's **Language** tab.

More interesting are the texts you choose yourself. As you read this list of sources, give each one a ranking of 1 to 5 for how likely you are to be interested in using such texts.

	a topic of personal interest to you – texts are available in many online 'special interest groups', fan sites, etc.
	gossip, scandals and urban myths/legends contain ideas and language that people are currently talking about.
	texts which contain interesting uses of language
	a text you have written yourself
	emails you have received: personal, professional, advertising, scams
	discussion groups like Quora, Reddit, Facebook groups
	news stories and current affairs available from many websites such as the BBC, Guardian, Wikinews.
	online magazines, e.g., Dezeen, the Advocate, Computer Music
	Wikipedia articles, Simple English Wikipedia articles. Other encyclopaedias, e.g., Britannica. Topic-based wikis, e.g., Hitchwiki, Wikivoyage.
	reviews and synopses of films and plays, which you can find at rottentomatoes.com and many online newspapers and magazines.
	required reading in a course you are doing, e.g., literature, academic text, research report
	a whole novel. Copyright free novels can be downloaded at Gutenberg.org.
	a short story: see classicshorts.com, Gutenberg.org.
	transcripts of spoken language, e.g., TED Talks and some news websites have texts both written and spoken.
	<i>your own idea</i>
	<i>your own idea</i>

Note that transcripts are mostly texts that are written to be spoken and are not natural conversation.

Under the heading, *Parsnips and fgleaves*, (p.12), we talked about topics that may be of interest to you but are not taught in normal textbooks.

The texts can be of any length. You can even paste several texts into the input field: it can be interesting to see more examples of your key words.

Note that VersaText does not save your text, so if you want to work with it again, you need to save it on your computer.

Authentic texts

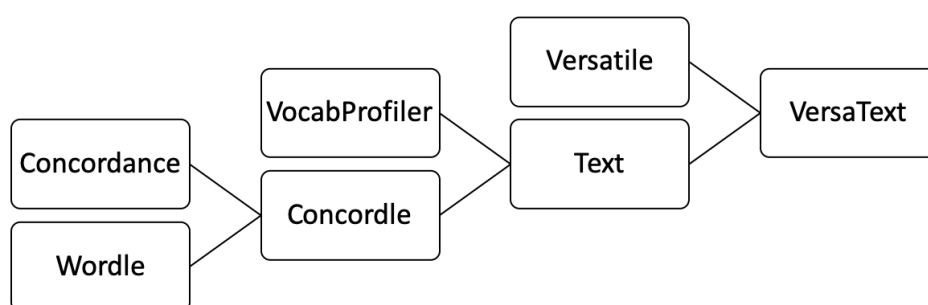
Everything listed above concerns authentic texts, not texts that have been created for language teaching. It is quite likely that one of your aims in studying English is to read “real” English – using VersaText with “real” texts is a step in that direction.

As a language learner, you can always learn something from good inauthentic texts too. For example, course books often contain the transcripts of listening texts. And many graded readers can be found on the Internet – you can search graded readers online.

A little background

VersaText is a portmanteau word blending *Versatile* and *Text*. And the tool itself blends features of two other online text tools, *Concordle* and *Vocabulary Profiler*. And then extends them both:

- *Concordle* makes a word cloud from a text and then shows how individual words are used in that text. This is also a portmanteau word blending *concordance* and *Wordle*. There are some notes and links here: bit.ly/maelt_concordle.
- *Vocabulary Profiler* provides frequency statistics about the words in the text. It is one of many tools in the widely used suite of online tools *called Compleat Lexical Tutor* – see <https://www.lextutor.ca>.



The author of this book was VersaText's matchmaker. I was using both *Concordle* and *Vocabulary Profiler* in my teaching and wanted to combine and extend the powers of both. VersaText is programmed by Vit Baisa, who was also largely responsible for SkELL, the topic of the next unit.

Word clouds in detail

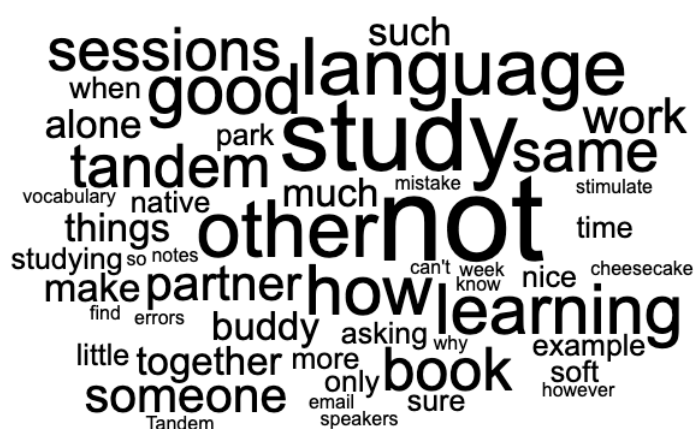
A word cloud is an image made up of the words from a text that you provide. Word clouds are popular in presentations, on webpages, coffee cups and are used in books and lessons as attractive assemblies of words. But they are not just decorative.

The important feature of a word cloud is that the more **frequent** a word is in the text, the bigger it appears in the word cloud. This gives you a good idea of what the text is about.

The following examples come from the pre-loaded text called Study Buddy.

Task 22. Based on the words in this word cloud, what do you think this text is about?

Your first response to this question will be a quick first impression.



But as you look more closely, you can answer these questions:

30. Does the text tell the story of someone's experiences of language learning?
31. Are the people named? And what does this tell us about the type of text?
32. Does it give students study advice?
33. Is this a text that reports research into language learning?
34. Is the text about more than one thing?

Why did you say yes or *no* to these questions?

Your response:

Task 23. Unknown words

Are there any words in the word cloud that you do not know? Remember that big words are relatively frequent in a text so it is important to know them.

Task 24. Your own observations

As you look at word clouds, you start to see things that might not seem obvious or important. You might wonder why *other* is so large and why *not* is even larger. Or why the text contains both *studying* and *learning*, both *mistakes* and *errors*, both *partner* and *buddy*. Or why *vocabulary* is the same size as *cheesecake*! What do you think?

Task 27. Colour-coding

In the table above, write the name of the colour beside the part of speech. The colours are: red, green, yellow, brown, etc.

The answer to this question can be found under the Word Cloud tab when you click **Info**.

VersaText uses **grey** when a word could be more than one part of speech in its context. Sometimes, the software is not sure. The word *study* in the *Study Buddy* text is a noun, a verb and part of the compounds *study buddy* and *study session*. Whenever you see words in grey, think about why the software was not sure. Are you sure? Can anyone be sure?

Click the **colourise** button to toggle between colour and black words.

Word cloud: word forms and lemmas

When we look up a word in a dictionary, we find its base form. Another term for this is **lemma**. The word clouds above contain **word forms**. Word forms are the base forms and the inflected forms of words, such as *learn*, *learning*, *learnt*; *session*, *sessions*; *do*, *doing*, *done*; *be*, *is*, *was*, *were*. The buttons below the word cloud let us choose what we want the word cloud to show. By default, VersaText shows word forms.

Words to show

10

20

30

40

50

60

70

80

90

100

word ☐ lemma ☒ tag ☐

Colourize ☒

Include content words

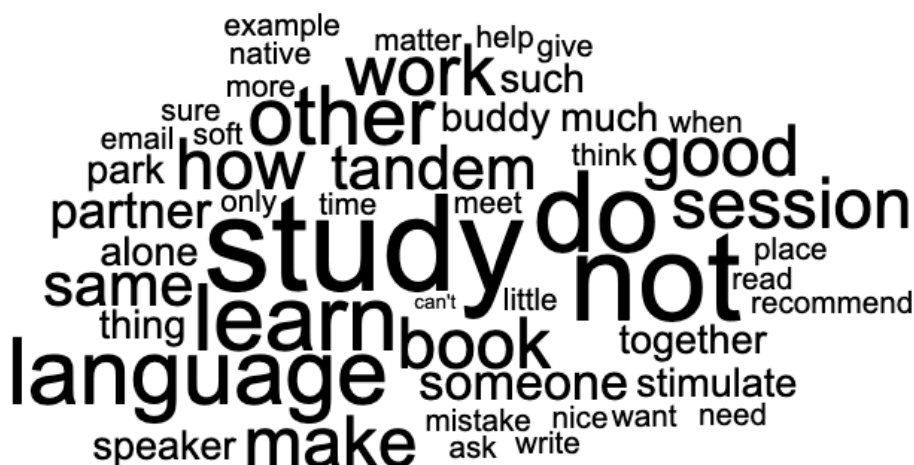
noun ☒ verb ☒ adjective ☒ adverb ☒

Include function words

preposition ☐ conjunction ☐ pronoun ☐

determiner ☐ punctuation ☐ other ☐ aux ☐

On p.43, the word cloud was made up of 50 **word forms**, but the word cloud here shows 50 **lemmas**. As we will see throughout this book, exploring words as lemmas sometimes comes in very handy.



Task 28. How many items are needed in a word cloud?

Open VersaText, choose a preloaded text or paste one of your own into the Input field. Then create the word cloud with ten words or lemmas only. Is this enough to answer the questions in Task 22? Experiment with different numbers of words and think about how many you need to make your observations and answer your questions. Makes some notes about this in your learning journal.

Word cloud: tags

As you can guess from the next image, tags are codes for parts of speech. VersaText analyses the text and gives every word a tag in its context. For example, those starting with N are nouns, and those starting with V are verbs. The rest of the code indicates other features such as plural, past forms, etc.

A word cloud of tags shows the relative frequencies of parts of speech in the text. A tag is the magic button that lets you see all of its words in the **concordancer**.

**Task 29. Match these tags to their full POS.**

This will involve a little guess work. You can check your answers under the Language tab at **Info** where the full set of tags is listed.

Adjective	NN	Past participle	VB
Adverb	JJ	Verb – base form	VMD
Wh- adverb	IN	3rd person sing.	VVN
Singular noun	JJS	-ing form	VVZ
Superlative adj	RB	Past tense	VVG
Preposition	WRB	Modal verb	VVD
Plural noun	NNS	Verb <i>be</i> – base form	VV

Here are some examples of lemma, word form, part of speech and tag.

Lemma	Word forms	POS	Tag
session	session	noun	N ____
	sessions	plural noun	N ____
want	want	verb – base form	V ____
	wants	3rd person singular	V ____
	wanted	past tense	V ____
	wanted	past participle	V ____

Task 30. The full tags

Using the tags in Task 29, change the one letter tags in the table above to full tags. For example, V → VVZ.

Concordance

This image is the concordance that the software creates when you click on the tag, RB, in the word cloud. The full list consists of 78 lines.

COLOURIZE

SENTENCE

Query: [tag="RB"], hits: 78.

#	Left context	KWIC	Right context
1	be at such a temperature that allows you to dress	lightly	. Turn off your phone and doorbell . Heaven
2	be any dripping taps to get under your skin .	Lastly	, give your partner some errands to run or
3	partner some errands to run or tasks to do ,	preferably	in a distant part of the house – you
4	– you might , for example , remind them how	much	you appreciated the cheese cake they baked recently
5	how much you appreciated the cheese cake they baked	recently	. Well , that's just the sort of tripe
6	you appreciated the cheese cake they baked recently .	Well	, that's just the sort of tripe that blogs
7	the cheese cake they baked recently . Well , that's	just	the sort of tripe that blogs and books on
8	is on study and concentration recommend . Has anyone	ever	given you such advice ? Advice written under those
9	recommended . Anyone who thought that study needed to be	hermetically	sealed clearly wanted to shield your mind from ideas
10	who thought that study needed to be hermetically sealed	clearly	wanted to shield your mind from ideas , inspiration
11	feet reading Proust and Kundera (in French and Czech	respectively	though not simultaneously) , at home in the
12	and Kundera (in French and Czech respectively though	not	simultaneously) , at home in the middle of

The target words run down the middle of the page, a format known as *Key Word in Context* (KWIC). These contexts allow us to observe a wide range of language features, which we will do later in this book. The concordance lines can be sorted alphabetically by clicking the column heading, **KWIC**.

The next image is the result of clicking on *tandem* in the word cloud. The concordance lines can be sorted in alphabetical order by the words on the left or right by clicking **Left context** or **Right context**. Sorting often reveals patterns in the word's usage. The patterns in the use of a word are its grammar, the very thing we need to convert our word knowledge from passive to active.

COLOURIZE

SENTENCE

Query: [word="tandem"], hits: 6.

#	Left context	KWIC	Right context
1	they are help you with theirs . This is called	tandem	learning . There are many web sites on tandem
2	called tandem learning . There are many web sites on	tandem	learning as well as a Wikipedia page . Tandem
3	a good reason for making lots of notes during your	tandem	sessions . Sometimes however , a correction can be
4	enjoyable . For example , you might arrive at your	tandem	session with this book and work on some of
5	you have learnt or corrected as a result of your	tandem	sessions . And your emails are likely to contain
6	learning requires some commitment . Firstly , you and your	tandem	partner need to have enough in common to have

Task 31. Learning content and language

These six concordance lines are all that the text says about *tandem*. Maybe not all: to avoid repetition words and ideas in a text are also referred to by pronouns, hypernyms and other related words which we explore in depth in Unit 12.

Read the lines carefully to answer these questions. Make notes in your learning journal.

1. Do the six lines give you any idea of what *tandem* means?
2. Is *tandem learning* a way of studying?
3. Do you do *tandem learning* alone?
4. Is the word always part of compound nouns?
5. Why is *your* so frequently used before it?
6. Is tandem learning widely known and widely used?
7. What words related to *studying* can you find in these six concordance lines?

In addition to KWIC view, there is also **Sentence view**. Click the Sentence button. The full sentences can be sorted by length. The full sentence concordances answer the question: what did the author say about the key word you chose?

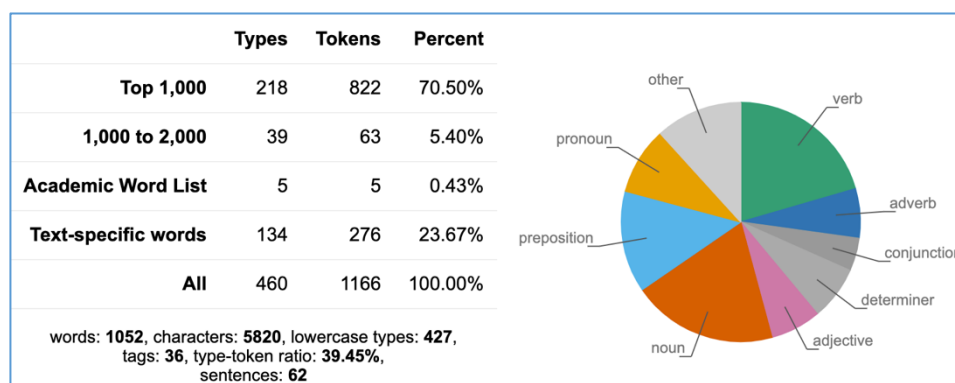
Task 32. DIY concordances

Open VersaText, choose a preloaded text or paste one of your own into the Input field. Make a word cloud then click on a “big” word to see its concordance. Sort the lines to the right context and to the left context. Are there any recurring words that form patterns with your target word?

Turn on Colourize. Are there any parts of speech that are frequent with your target word? For example, some words are typically followed by *to* infinitives, by prepositions, etc. Do you learn anything from this?

Profiler

The Profiler shows lexical statistics about the text: words, lemmas and tokens.



The Study Buddy text has 1,166 **tokens** of which 1,052 are **words**. Words are tokens with alphabetical characters only, so *123*, *AK47* and *?* are not words. A lot of the tokens are repeated, especially function words. In fact, there are only 460 different tokens in this text – **types** are the different tokens.

The **Top 1,000** words are the most frequent 1,000 words in English generally. This table says that 70.5% of all the words in the text are among the most frequently used 1,000 words of English and that only 5.4% are in the second thousand: **1,000 to 2,000**.

The **Academic Wordlist** (AWL) is a widely-used, published list of words that are not in the top 2,000 but are commonly found in many kinds of academic texts. The AWL does not include terminology. Less than 1% of the words in the Study Buddy text are AWL words. In an academic text, 15% is standard.

Text-specific words are the words in the text that are not in the other three groups.

Task 33. Text-specific words and word clouds

Why do you think that the words in a word cloud are more or less the same as the words in the text-specific words table?

Why are there no function words in the text-specific words table?

The “other” category contains items that the software couldn’t put anywhere else. This is mostly understandable but sometimes there are few surprising things on the list.

Type-token ratio

The type-token ratio (TTR) is the number of tokens divided by the number of types times one hundred. However, the TTR varies very widely in accordance with the length of the text: a 1,000-word article might have a TTR of 40%, while a shorter one might reach 70%. There are procedures for standardising TTR, which we will not study in this book.

On the website, all of the lemmas in the text are listed in their part of speech groups under the word profiler. Here is the set for Study Buddy’s text-specific words.

Text-specific words

PoS	Items
noun	shoulder desk waterfall forest creature height temperature dress doorbell plumber tap housemate cheesecake tripe blog shield inspiration stimulation zone buddy homework essay wikihow com rhythm piano hippy sweater shore sheep wool d tandem wikipedia beginner existent cup tea squirrel pigeon teenager skateboard kid delight café wi fi nobody limitation e scrutinize furniture t don
verb	studying light shine hand must should shop drip remind has bound everything distract irritate can thing park stimulate ponder knit spin something arriving squeal fossilize misuse correct focussed nothing anything
adjective	quiet tidy gentle comfortable nearby distant bland warm conducive favourite advanced non many slow frustrating boring routine welcome productive enjoyable soft
adverb	lightly lastly how where actively fruitfully agreeably gladly ply firstly lovely why tacitly likely
other	., ' ? advice ' : two – i

Pie Chart

The pie chart shows the relative frequencies of parts of speech in the text.

Show lemmas

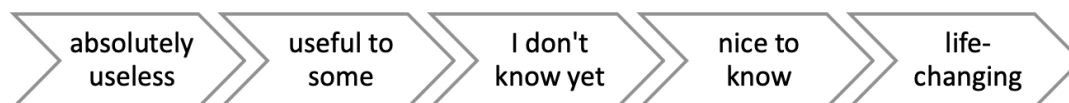
At the bottom of the Profiler page, there are two buttons: **Show Lemmas** and **Analysis**. *Show lemmas* lists all the lemmas in the text. Each column can be sorted by clicking on its heading. This screenshot is the top of the list, sorted by frequency. In Task 25 we asked what the most frequent words in a text were. Now we know!

Lemma	Freq ↓	HPM	Gen. HPM	Score
.	74	52631.58	39298.54	1.34
be	67	47652.92	31014.83	1.54
,	49	34850.64	45973.99	0.76
you	44	31294.45	3545.94	8.82
and	43	30583.21	24847.11	1.23
to	36	25604.55	22196.00	1.15
your	35	24893.31	1355.40	18.35
the	34	24182.08	56157.09	0.43
of	28	19914.65	26614.01	0.75
a	28	19914.65	18539.64	1.07
in	20	14224.75	18678.92	0.76
that	18	12802.28	8735.93	1.47
with	16	11379.80	6341.80	1.79
as	16	11379.80	6322.10	1.80
have	13	9246.09	8472.19	1.09

Analysis

Show Analysis shows all the words in the text in the order of the text. You can read the text vertically. The columns cannot be sorted.

The columns show how VersaText has analysed the text to provide the colours, tables and charts. If you find something strange in the word cloud, for example, you can find it in this table to see what VersaText has done. No programs that analyse text are perfect.

Task 34. How would you rank the usefulness of the information that Profiler provides?

Your response:

Time to play

Now that you have met all the tools that VersaText offers, go play! You can use the pre-loaded texts first and then paste in some other texts like those suggested on p.41. You are bound to make some interesting discoveries about the texts, how their key words express the text's messages, and about how these words combine with other words. As you look at a number of texts, you might discover things about English in general.

Make notes in your learning journal now and look at them again as you do the tasks in this book.

Reflection

This Unit opened with the sentence: *VersaText is an online tool that helps you explore the language of a text.* When you look at word clouds there are many things you can discover about a text. But not everything. And when you look at concordances, there are many things you can discover about the words. But not everything.

These ways of looking at the language in texts are quite different from the activities found in course books. Don't throw the baby out with the bath water. Combine the best of both worlds.

This brings to an end the introduction to VersaText. In conjunction with SkELL, it will be used throughout this book to study grammar, vocabulary and everything in between.