

# Cambridge English

## Real language, real learners: the importance of corpus data in our profession

Annette Capel

Thun: January 26, 2014



**CAMBRIDGE**  
UNIVERSITY PRESS

## Content of this talk

- Why ARE electronic text corpora important to us?
- 20 years of the *Cambridge Learner Corpus* (CLC)
- Swiss learner evidence in the CLC
- Real English – a quick look at first language use
- The corpus-informed English Profile programme
- Using the *English Vocabulary Profile* in the classroom



## The importance of corpus data

An electronically-held text corpus is:

- vast – contains millions or even billions of words
- fast and powerful – can be searched in seconds
- highlights what is frequently being produced
- clearly shows typical patterns eg structures, phrases, collocations
- real living data (first language user or learner)



## Corpora provide real examples

Query **smartphone, apps** 3 (0.0 per million)

3 hits 213 /2? </p><p> However Orange are restricting **smartphone apps** to those approved by themselves, which  
doc#1092658 is I have already invested quite a bit in **Smartphone apps** . </p><p> Hi Arne, first of all congrats.  
doc#2105832 existing games and other graphics-based **Smartphone apps** (Windows Mobile 6 Standard) out there only

my good-child Anna. Whe took the train at **Thun** , my mother, put her suitcase under the

vilage called Riggisberg. It's about between **Thun** and Belb, in the Berner Oberland. </p><p>

Verry often we all buy t-shirts because in **Thun** are some good shops for that. </p><p> I hope

us tonight. We'll go to the Rex cinema in **Thun** , I'll take my father's car so I can pick

montains. The view is wonderful. If you stay in **Thun** in the summertime you have the possibility

clothes shop is the „New Yorker,, store in **Thun** , because the clothes there are not so expensive

most beautiful. My town is at the lake of **Thun** and very near by the montains. The view

verry often, but when i go, then I go to **Thun** with my best friends. These afternoons



## 20 years of the CLC

- Largest learner corpus in the world - over 55 million words
- 3,291,257 words of Swiss learner data (almost 6% of total)
- Covers all six CEFR levels
- General, business, legal, financial and academic exams
- Examples of strong, average and weak performance
- Error-coded data enables authors to search for common mistakes by L1



# Cambridge English

## Top 10 Swiss learner errors

<b>word</b>	<b>Freq</b>	
accomodation	147	
wich	146	
departement	138	
advertisment	123	
programm	97	
prefere	63	
verry	58	
cours	56	
adress	53	
allways	49	



# Cambridge English

## *Cambridge English: Preliminary* Extended writing task, 2010

This is part of a letter from your English penfriend.

*I live in a really busy street. What is it like where you live? If you were able to move, where would you like to live?*

Now write a letter, **answering your penfriend's questions.**

Write your letter in about 100 words.



# Cambridge English

## Swiss PET performance (B1 pass)

*Dear Pam How are you, I'm fine. I'm sorry that I couldn't write back earllyer but I was busy so never minde. I'm sorry that you live in such a busy street. Luckaly my home isn't as busy as yours, at least sommetimes. I live in a litle vilage called Riggisberg. It's about between Thun and Belb, in the Berner Oberland. You asked me if I would like to live sommewhere else well I'm not so shour. I'm quite happy where I'm now. So I would probably stay here. Write back soon I'd like to hear from you. Best wishes*



# Cambridge English

## Swiss PET performance (B1 pass)

*Dear Pam How are you, I'm fine. I'm sorry that I couldn't write back earllyer but I was busy so never minde. I'm sorry that you live in such a busy street. Luckaly my home isn't as busy as yours, at least sommetimes. I live in a litle vilage called Riggisberg. It's about between Thun and Belb, in the Berner Oberland. You asked me if I would like to live sommewhere else well I'm not so shour. I'm quite happy where I'm now. So I would probably stay here. Write back soon I'd like to hear from you. Best wishes*



## Swiss PET performance (fail)

*Dear John I live in a small village at the mountain. The street where I live is very small and the neighbour are very nice.*

*When I able to move I will move to Geneva because I was there for two weeks and it was great while I am training my french. But it's also nice in this village because of the natur. I like the people from Geneva because they've got a great heart. They've also got a green logure into Geneva: the wall of the reformater. Where will you move, if you can? Nice to hearing from you*



# Cambridge English

## Swiss PET performance (fail)

*Dear John I live in a small village at the mountain. The street where I live is very small and the neighbour are very nice.*

*When I able to move I will move to Geneva because I was there for two weeks and it was great while I am training my french. But it's also nice in this village because of the natur. I like the people from Geneva because they've got a great heart. **They've also got a green logure into Geneva: the wall of the reformater.** Where will you move, if you can? Nice to hearing from you*



## Swiss learner performance

### **Making the examiner hungry!**

*Eating has a high value in Switzerland. Starting with "Z'morge", corresponding to the usual breakfast, then "Z'mini" which would be on "at nine". After lunch ("Z'millag") we usually have another "Z'vuri" which you can compare to the English tea-time perhaps, set at four o'clock in the afternoon. After dinner ("Z'abyl") one might even have another late snack referred to as "Z'meacht", eaten at night.*



# Cambridge English (Swiss) learner performance

## Making the examiner smile!

- *After eating **old cheese** and Swiss chocolate, that was high time Mark's family went back home.*
- *So I thought to go on the **backside** of the house and to go then through the backdoor into the house.*
- *I would like to try **budgie jumping** because every time I look down from my balcony I think I'm going to fall and never come back.*
- *I have never been **in** a camel before – what a laugh!*



Cambridge English

# Swiss learner performance

**Educating the examiner!**

***Nicolas Hayek, Founder of the Swiss-based Swatch  
Group Corporation***

*For the stand at the school I suggest to put up one of the  
giant Swatches, found in many Swiss households, along with  
a miniature Smart car and a glossy picture of Mr Hayek,  
where the main achievements of his life could be reflected.*



**CAMBRIDGE**  
UNIVERSITY PRESS

# Cambridge English

## Cambridge English Corpus



- Written English
- Spoken English
- Business English
- English for Specific Purposes
- Academic English
- Learner English



CAMBRIDGE  
UNIVERSITY PRESS

# Cambridge English

## The Cambridge English Corpus

### Did you know...

.... if we assume an average reading speed of 300 words per minute (for most people this is faster than memorising or proofreading, but slower than skimming), the Cambridge English Corpus would take **88,766 hours to read!**

.... if we printed out the whole of the Cambridge English Corpus on single sided A4 paper, this huge tower of words would stand at **almost 600m** (not far behind **the tallest building in the world!**).

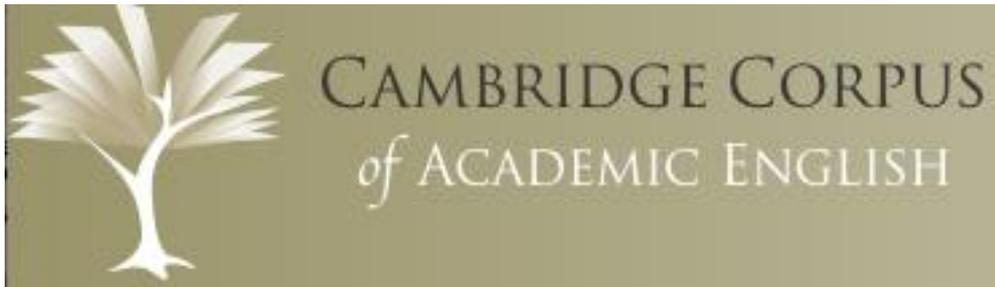


**CAMBRIDGE**  
UNIVERSITY PRESS

# Cambridge English

## The Cambridge English Corpus

- 1.6 billion words providing authentic real-life examples for authors, lexicographers and editors
- Allows researchers to monitor trends and identify new meanings, new words, phrases and idioms
- Ongoing data collection and new research eg CEPC (English Profile) and CCAE (Academic English)



**CAMBRIDGE**  
UNIVERSITY PRESS

# Cambridge English

## Monitoring language use

**nuts to butts** *idiom* informal in a way that is very crowded

*I understand the tenets of capitalism and the desire to make money. I get that more money can be made by packing people into these places ‘nuts to butts’.*

(July 2013)

*About Words*

a blog from Cambridge Dictionaries Online



CAMBRIDGE  
UNIVERSITY PRESS

## A quick look at first language data

Of course, Google is a corpus too...

[smart car UK | Facebook](https://www.facebook.com/smartcaruk)

<https://www.facebook.com/smartcaruk>

**smart car UK**. 28424 likes · 485 talking about this. Born of a collaboration between SMH, the company responsible for Swatch watches, and Mercedes-Benz



## Australians' views on the Smart car

**Online posts offer easy access to real language...**

**Emily** Think about the Earth. They are great for the environment and they are pretty stylish too. I like them,

**But you don't always get 'good models of English'!**

**Xavier** The car might help the earth but it isn't very cool for guys, **its so small and its not very big.**

**Hudson** The ugliest car that i've seen **in whole my life...**

**Whose** the designer?



## Important new meanings: *smart*

Query **smart** 69,007 > Sort **word/i 1>0~3>0** 69,007

First | Previous Page 2609 of 3,451 Go Next | Last Concordance is sorted. Jump to: ! ⇅

Web-fin-18...	operating in RFID, printed electronics, and	<b>smart</b>	packaging 3 Unisys Financial review The
Web-comm-9...	eliminate lines. One approach is called	<b>smart</b>	packaging; a product has a tag that gives
2687867	Paul says. "We choose the best quality,	<b>smartest</b>	packaging and sell it at the most competitive
2300381	top hat. </p><p> "The Grand" does offer some	<b>smart</b>	packaging around the improv bits, like
1542200	Inflatable rooms </p><p> TRENDWATCH </p><p>	<b>Smart</b>	packaging </p><p> Ask any student of consumer
1122333	We were impressed with the Safeway egg's	<b>smart</b>	packaging but the quality of the chocolate
1115165	keeping tabs on the product's location,	<b>smart</b>	packaging can also hold information such
Web-comm-8...	high as 40% of a product's selling price.	<b>Smart</b>	packaging decisions can help reduce costs
1115165	could detect when a specific product in	<b>smart</b>	packaging has been picked up by a shopper
1408514	will eventually be topped by a V6. Boot:	<b>Smart</b>	packaging has ensured there's no wheelarch
1797407	the most ruthless staff-cutting program.	<b>Smart</b>	packaging is half the battle. </p><p> 2.
1558198	insurance </p><p> Travel Insurance </p><p>	<b>Smart</b>	packaging makes Hawaiian Tropic's 60+ SPF
Web-comm-9...	Briefing Presentation on Global Active &	<b>Smart</b>	Packaging Market Trends, Forecast & Applications
Web-comm-9...	Briefing Presentation on the Global Active &	<b>Smart</b>	Packaging Market Trends, Forecast & Applications
1542201	</p><p> 52 The Hottest Thing on DVD? Old TV	<b>Smart</b>	packaging of old series, led by the newly
2210261	individuals trade directly with one another.	<b>Smart</b>	packaging refrigerates food using a tiny
1115165	Everyone is impressed by star names." </p><p> The	<b>smart</b>	packaging that lets Diageo keep its bottle
1115165	said. "But there are many possibilities."	<b>Smart</b>	packaging uses what is known as a radio
Web-food-2...	overnight or a week-long getaway. For the	<b>smart</b>	packer it would even be totally suitable
Web-med-12...	vacation - good health. Early preparation and	<b>smart</b>	packaging can help make your trip a healthy



# Cambridge English

- A1
- A1-A2
- A1-B1
- A1-B2
- A1-C1
- A1-C2
- A2 only
- B1 only
- B2 only
- C1 only
- C2 only

Browse A-Z

OR

Enter a word or phrase

smart

## ADVANCED SEARCH



- Category
- Part of speech
- Grammar
- Usage
- Topic
- Prefix
- Suffix

Clear filters

Hide culturally sensitive words

Search

English Profile Home

## ► ADJECTIVE

### STYLISH

**B1** having a clean, tidy and stylish appearance

**Dictionary examples:**

*Guy **looks** very smart in his new suit, doesn't he?*

*I need a smart jacket for my interview.*

*She works in a very smart **new** office overlooking the River Cam.*

**Learner example:**

*I don't like smart clothes, but for special days I have [a] very nice long black skirt.*

### CLEVER

**B1** intelligent

**Dictionary examples:**

*Gemma's teacher says she's one of the smartest kids in the class.*

*Why don't you fix it if you're so smart?*

*I'm not smart enough to understand computers.*

*He's smart enough to know he can't run the business without her.*

*Quitting that job was the smartest **move** I ever made.*

**Learner example:**

*He is [a] very kind, smart, and good person, besides, [he] is a good son.*

### TECHNOLOGY

**C1** using computers or information stored in electronic form

**Dictionary examples:**

*a smart phone*



CAMBRIDGE  
UNIVERSITY PRESS

# Cambridge English

## English Vocabulary Profile



British English

American English

Choose level:

- A1
- A1-A2
- A1-B1
- A1-B2
- A1-C1
- A1-C2
- A2 only
- B1 only
- B2 only
- C1 only
- C2 only

Browse A-Z

OR

Enter a word or phrase

ADVANCED SEARCH

Hide culturally sensitive words

Search

English Profile Home

About English Vocabulary Profile

Word of the Week

Help

Feedback

### About the English Vocabulary Profile

Welcome to the pilot version of the full six-level **English Vocabulary Profile**. This version covers levels A1-C2 of the Common European Framework of Reference (CEFR) for British and American English. If you have any comments, questions or requests, the EVP team would love to hear from you! **Get in touch** now.

- What is the English Vocabulary Profile (EVP)?
- Why is the EVP important to me?
- How has the EVP been created?
- What does an EVP entry include?
- How can I work with the EVP?

What is the EVP?

Why is the EVP important?

How has it been created?

What does an EVP entry include?

How can I work with the EVP?

### What is the English Vocabulary Profile?

The EVP shows, in both British and American English, which words and phrases learners around the world know at each level - A1 to C2 - of the CEFR. Rather than providing a syllabus of the vocabulary that learners *should* know, the EVP project verifies what they *do* know at each level. CEFR levels are assigned not just to the words themselves, but to each individual meaning of these words. So, for instance, the word *degree* is assigned level A2 for the sense TEMPERATURE, B1 for QUALIFICATION, B2 for AMOUNT and C2 for the phrase *a/some degree of (sth)*. The capitalized guidewords help the user to navigate longer entries, and phrases are listed separately within an entry.

Find out more about the CEFR.

	Common European Framework of Reference for Languages	General English
Proficient user	C2 Mastery	CPE
	C1 Effective Operational Proficiency	CAE
Independent user	B2 Vantage	FCE
	B1 Threshold	PET
Basic user	A2 Waystage	KET
	A1 Breakthrough	



CAMBRIDGE  
UNIVERSITY PRESS



## Corpus-informed research into idioms

- Productive use of idioms is mainly a feature of the C levels
- Inclusion in the English Vocabulary Profile was dictated by first-language corpus frequency and finding examples in the Cambridge Learner Corpus
- Not included: *raining cats and dogs, nuts and butts*
- Included: *behind closed doors, down to earth*



## First-language frequency

Query **behind, closed, doors** **4,156** (2.2 per million)

Page  of 208   | [Last](#)

<b>1541038</b>	yet covet them, with hypocritical lust, <b>behind closed doors</b> . </p><p> 8 Open Water A scuba diving couple
<b>1541930</b>	sniffing "the fine white fuses of pleasure" <b>behind closed doors</b> . "Nick loved the way coke took off the
<b>1542532</b>	Even though Weisbrod says he was "giddy <b>behind closed doors</b> ," he didn't let on in public, keeping in
<b>1543700</b>	Press, 228 pages. Paper, \$19.95. This is a " <b>behind closed doors</b> " look at freelancing from the creator of
<b>1544227</b>	For it is now clear that Soviet officials, <b>behind closed doors</b> , and from the Politburo down to the local
<b>1544444</b>	Congressmen tell TIME there was a bitter debate <b>behind closed doors</b> over whether to change the rule for the
<b>1545359</b>	pick up a net of four seats in the chamber. <b>Behind closed doors</b> , Reid is working to assemble his own leadership
<b>1547699</b>	avoid their faculty lounge because the talk <b>behind closed doors</b> is offensive to them, full of negative
<b>1547699</b>	want to change things quietly, politely, <b>behind closed doors</b> . I understand that feeling, having a certain
<b>1548201</b>	of them apparently negotiated by Cheney <b>behind closed doors</b> with lobbyists from Enron and other energy
<b>1548461</b>	representative of al-Sadr's father: "When he sits <b>behind closed doors</b> , he is seen as a man who makes no mistakes
<b>1548461</b>	. He is still most comfortable operating <b>behind closed doors</b> ; he hasn't conducted Friday prayers for
<b>1551609</b>	practice of medicine, the stuff that goes on <b>behind closed doors</b> , is often gruesome, gory, and messy. Saws
<b>1551835</b>	on his work, not his life. He rehearses <b>behind closed doors</b> and rarely gives interviews; no book has
<b>1552687</b>	Bryant in court for first time; testifies <b>behind closed doors</b> about her sexual history </p><p> JUNE 25,
<b>1553882</b>	dire financial straits - typically retreat <b>behind closed doors</b> to broker a deal, usually with little input
<b>1557716</b>	and Jake disappear into the dressing room, <b>behind closed doors</b> ...hmmm! We'll leave this one to you...
<b>1558722</b>	were regularly meeting with EPA officials <b>behind closed doors</b> to promote a different approach, according
<b>1559310</b>	's been going on and who's been doing it <b>behind closed doors</b> ." No reference is made to Hillary Clinton
<b>1559571</b>	trust. But the government's tendency to deal <b>behind closed doors</b> with all security matters is only fueling

Page  of 208   | [Last](#)



# Cambridge English

## Idioms at EVP entry for *behind*

### behind sb's back

**C2** If you do something behind someone's back, you do it without them knowing, often in an unfair way.

**Dictionary example:**

*Have they been saying things about me behind my back?*

**Learner example:**

*The food must be perfect, too, or else her mother-in-law and my father's sisters will think that she's not a proper wife, and will probably say insulting things about her behind her back.*

*Certificate of Proficiency in English; C2; Greek*

### behind closed doors

**C2** privately and not in public

**Dictionary example:**

*Most of the deals were done behind closed doors.*

**Learner example:**

*Decisions which affect the whole world are made behind closed doors and are top secret.*

*Certificate of Proficiency in English; C2; Russian*

### behind the scenes

**C2** If something happens behind the scenes, it happens secretly.

**Dictionary example:**

*A lot of hard work has been going on behind the scenes.*

**Learner example:**

*Ms. Bolam knew a lot of things going on behind the scenes, [which were] not [known to] the patients.*

*Certificate of Proficiency in English; C2; German*



# Cambridge English

## Using the EVP in the classroom

A1-A2       A2 only  
 A1-B1       B1 only  
 A1-B2       B2 only  
 A1-C1       C1 only  
 A1-C2       C2 only

[Browse A-Z](#)

OR

Enter a word or phrase

**ADVANCED SEARCH** 

Category

Core results:

- **down to earth** C1
- **down under** C2
- **down the drain** C2
- **down and out** C2
- **be down in the dumps** C2

**down** was also found at:

- **put your foot down** C2
- **let your guard down** C2
- **let your hair down** C2
- **a stroll/trip/walk down memory lane** C2
- **send shivers/a shiver down your spine** C2



# Cambridge English

## English Profile is...

A long-term collaborative research programme aiming to:

- understand what the CEFR actually means for English
- investigate what learner English is really like
- develop reference descriptions for each CEFR level

With individual projects focusing on vocabulary, grammar, functions, spoken English and pronunciation



Education and Culture DG

Lifelong Learning Programme



**CAMBRIDGE**  
UNIVERSITY PRESS

# Cambridge English

## The English Grammar Profile project

Using the Cambridge Learner Corpus data, with some

comparison of first-language frequency

- describes gradual mastery of structures
- highlights form and use separately
- Can do statements at each CEFR level
- new insights into actual learner ability



Education and Culture DG

Lifelong Learning Programme



**CAMBRIDGE**  
UNIVERSITY PRESS

# Cambridge English

Future continuous	A2	FORM: AFFIRMATIVE WITH <i>WILL</i>	Can use the affirmative form with <i>will</i> .
Future continuous	A2	USE: FUTURE ARRANGEMENTS	Can use the future continuous with <i>will</i> to talk about an event or action in progress at a specified time in the future.
Present continuous for future use	A2	USE: FUTURE ARRANGEMENTS	Can use the present continuous with a limited range of verbs to talk about future arrangements.
Present continuous for future use	A2	USE: QUESTIONS ABOUT THE FUTURE	Can use the present continuous with a limited range of verbs to ask about future plans.
Present simple for future use	A2	FORM: WITH <i>WHEN</i>	Can use the present simple with <i>when</i> to refer to the future.
Present simple for future use	A2	USE: TIMETABLES AND PLANS	Can use the present simple to talk about timetabled events in the future.
Present simple for future use	A2	USE: <i>HOPE</i>	Can use the present simple with <i>I hope</i> to talk about hopes and wishes for the future.



# Cambridge English

Time to end this talk... with closes  
from learners around the world

- *Yours fried,*
- *Kind degrades,*
- *I'm looking forward to seeing to you  
soon.*
- *Best witches and loaves,*  
*Annette*



# Cambridge English

## EnglishProfile

CEFR for English

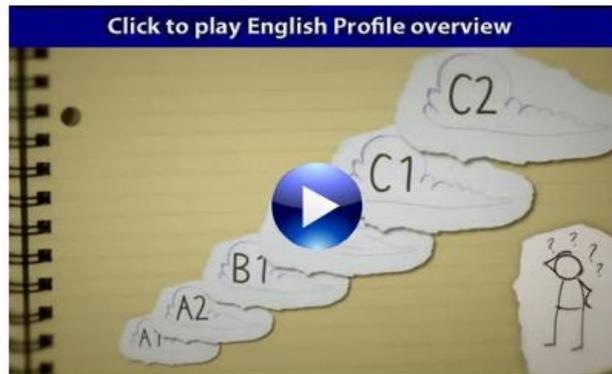
- Home
- About
- Events calendar
- Community
- Get involved
- Resources
- Corpus
- For researchers
- Teachers' area



### Welcome to English Profile

English Profile is a collaborative programme endorsed by the Council of Europe, designed to create a 'profile' or set of [Reference Level Descriptions](#) for English. These will provide detailed information about what learners 'can do' in English at each of the six levels of the Common European Framework of Reference (CEFR), offering a clear benchmark for progress for English language learners.

Once you've had a look around our site, please [take our survey](#) and let us know how we can improve it! It only takes a couple of minutes to complete, and as a thank-you, you'll be entered into a prize draw where you can win [a £50 Amazon giftcard](#).



This site gives the background to the English Profile programme and provides access to our resources, including the [English Vocabulary Profile](#), which is a searchable listing of the words and phrases typically produced by English language learners at each level of the CEFR. It provides information at word and sense level, as well as examples of the word in use by real learners, extremely useful for anyone in the ELT community. [Subscription](#) is free through this website.

Search

Free subscription  
English Vocabulary  
Profile



CAMBRIDGE  
UNIVERSITY PRESS