

Introducing the written BNC2014

Lancaster University @ Lancaster Castle
19 November 2021

Intro

Vaclav Brezina

Lancaster University




Lancaster
University









12.30 – 12.45 **Vaclav Brezina** (Lancaster University): Welcome and Introduction to the event

12.45 – 12.50 **Elena Semino** (Lancaster University): Welcome from the Director of The ESRC Centre for Corpus Approaches to Social Science (CASS)

12.50 – 12.55 **Paul Connolly** (Lancaster University): Welcome from the Dean of Faculty of Arts and Social Sciences (FASS)


12.55-13.00 **Break**

13.00 – 13.15 **Tony McEnery** (Lancaster University): The idea of the written BNC2014

13.15 – 13.30 **Dawn Knight** (Cardiff University): Building a National Corpus: The story of the National Corpus of Contemporary Welsh

13.30 – 13.45 **Vaclav Brezina** (Lancaster University): Current British English

13.45 – 14.00 **Vaclav Brezina** and **William Platt** (Lancaster University): Exploring the BNC2014 using #LancsBox X



14.00 – 14.15 **Randi Reppen** (Northern Arizona University): Corpora in the classroom

14.15 – 14.30 **Alice Deignan** (University of Leeds): Corpora in education

14.30-14.45 **Dana Gablasova** (Lancaster University): Corpus for schools

14.45 – 15.00 **Bas Aarts** (University College London): Plonker of a politician NPs

15.00 – 15.15 **Break**

15.15-15.30 **Marc Alexander** (University of Glasgow): British English: A historical perspective

15.30 -15.45 **Michaela Mahlberg** (University of Birmingham): Corpora and fiction

15.45 – 16.00 **Martin Wynne** (University of Oxford): CLARIN – corpora, corpora tools and collaboration



Official welcome

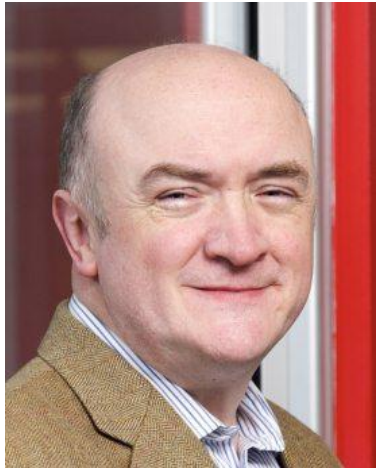


Professor Elena Semino
Lancaster University



Professor Paul Connolly
Lancaster University

Speakers



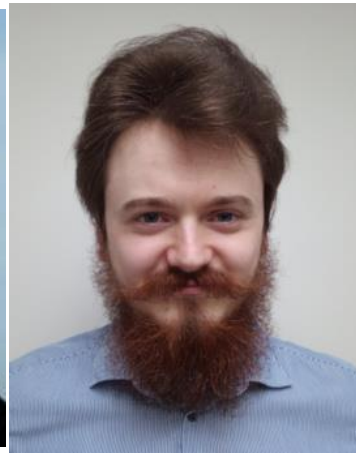
Lancaster
University



Cardiff
University



Lancaster
University



Lancaster
University



Northern
Arizona
University



University of
Leeds

Speakers



Lancaster
University



University
College London



University of
Glasgow



University of
Birmingham



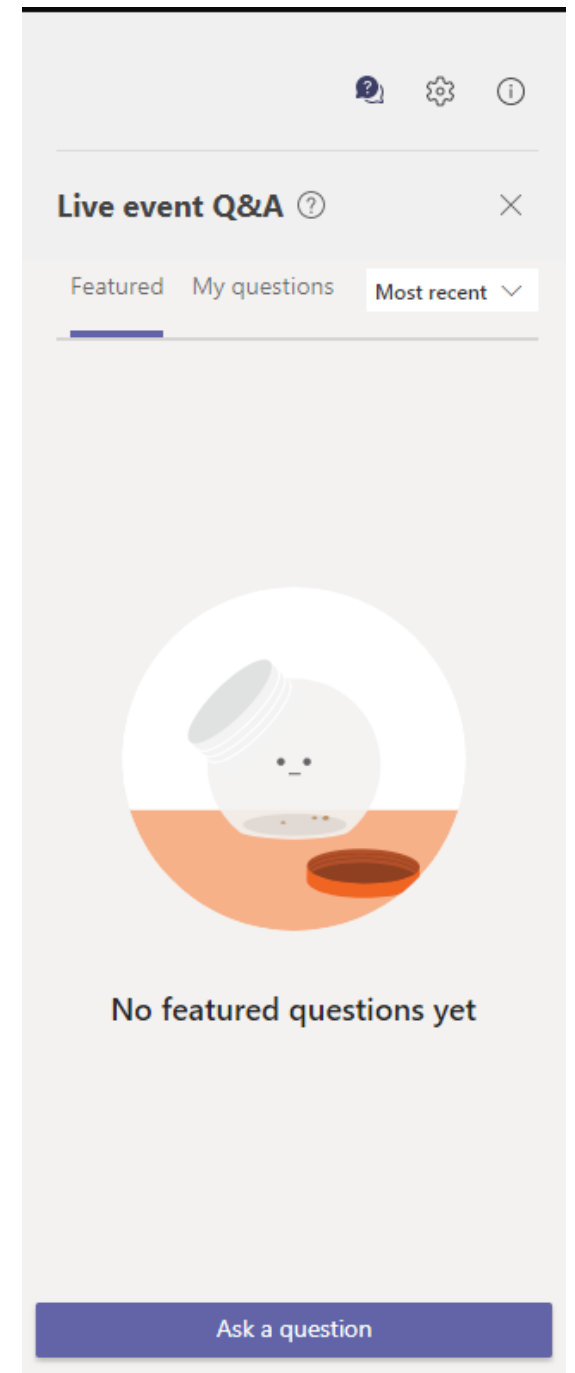
University of
Oxford

Over 1,000 participants online

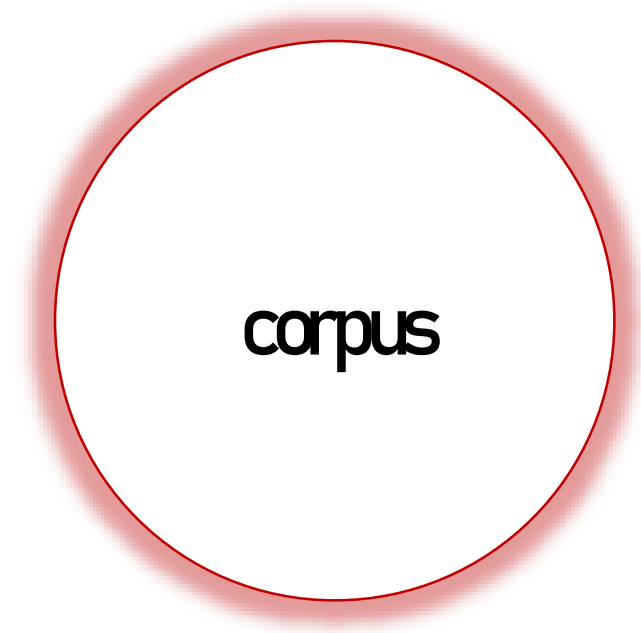
The word cloud features the central term 'corpus linguistics' in a large, bold, dark blue font. Surrounding it are numerous related terms in various sizes and shades of blue and teal. The terms include:

- Corpus in education
- Applied Linguistics
- Corpus translation
- second language
- contrastive linguistics
- historical linguistics
- linguistics and discourse
- corpus stylistics
- Cognitive Linguistics
- Forensic Linguistics
- English language
- Learner corpus
- corpus research
- Corpus annotation
- Corpus linguistics
- Corpus approaches
- Computational Linguistics
- language teaching
- discourse analysis

Q&A



Corpus as a sample



Corpus as a sample

Representative?
Balanced?
Unbiased?

BRITISH
NATIONAL
CORPUS
2014

BRITISH
NATIONAL
CORPUS
2014

B N C
BRITISH
NATIONAL
CORPUS
2014

BNC
BRITISH
NATIONAL
CORPUS
2014

BNC2014 in a nutshell

100M words (**90M writing**, 10M speech)

88,171 files

current

robust

8 genres, 52 subgenres



balanced

7 years in making

representative

British English



BRITISH
NATIONAL
CORPUS
2014



Lancaster
University

With the BNC2014 you can...



SMS

Reviews

Media

Newspapers



Television



Speech

EXPLORE

Magazines



Blog posts



Academic
Writing



Fiction

@Emails

What?

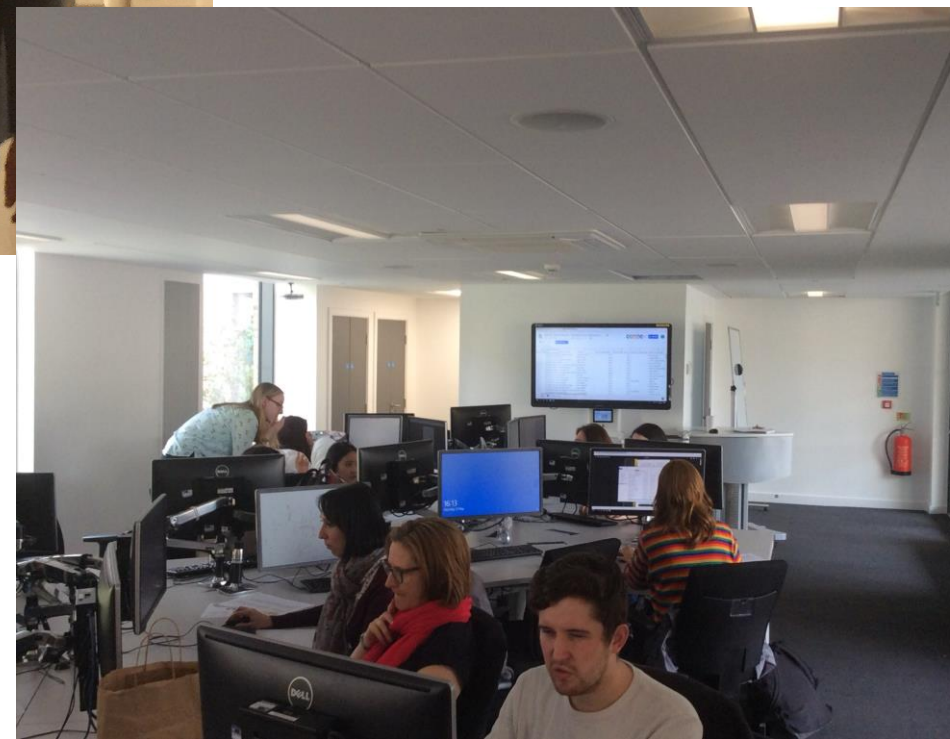
100 million words of
written and spoken English.

Why?

To analyse real language use
and how it is evolving.

Where?

#LancsBox X





BRITISH
NATIONAL
CORPUS
2014



Lancaster
University

With the BNC2014 you can...



SMS

Reviews

Media

Newspapers

Television



Speech

EXPLORE

Blog posts



Academic
Writing



Fiction

@Emails



What?

100 million words of
written and spoken English.

Why?

To analyse real language use
and how it is evolving.

Where?

#LancsBox X

BNC2014 Book Collection - Submission Form (Responses)

Timestamp	Title of the book:	Author(s):	Year of first publication:	Scanned book
5/12/2018 14:20:29	The Bedlam Stacks	Natasha Pulley		
5/12/2018 15:03:10	Hate	Alan Gibbons	2017	
5/12/2018 16:07:02	The Green Road	Anne Enright	2014	
5/12/2018 16:07:49	NW	Zadi Smith	2015	
5/12/2018 16:08:11	Jo Cox: More in Common	Brendan Cox	2017	
5/12/2018 16:08:34	The Wolf Border	Sarah Hall	2017	
5/12/2018 16:09:01	Loop of Jade	Sarah Howe	2015	
5/12/2018 16:09:12	The Casebook of Victor F	Peter Ackroyd	2015	
5/12/2018 16:09:37	Umbrella	Will Self	2008	
5/12/2018 16:10:33	The Lie Tree	Frances Hardinge	2012	
5/12/2018 16:11:39	The Betrayal	Helen Dunmore	2015	
5/12/2018 16:11:34	Conquest	Zoe Brigley	10	
5/12/2018 16:12:06	The Friday Gospels		12	
			13	





BRITISH
NATIONAL
CORPUS
2014



With the BNC2014 you can...



What? 100 million words of written and spoken English.

Why? To analyse real language use and how it is evolving.

Where? #LancsBox X

Acknowledgements

We would like to thank a number of researchers who helped collect data for the Written BNC2014, in particular Carmen Dayrell, who collected large amounts of data across different genres/register. In addition, we would like to thank Zoe Broisson, Mathew Gillings, Andressa Rodrigues Gomide, Vasiliki Simaki, Matt Timperley, Isolde van Dorst and Pierre Weill-Tessier. William Platt, Hanna Schmueck and Maggie Mi assisted with data processing.

Research funding: This work was supported by the Economic and Social Research Council (grant number EP/P001559/1, ES/K002155/1 and ES/R008906/1).

Publishers: Cambridge University Press, Dunedin Academic Press, Elsevier and John Benjamins.

More info...



Open Access Published by De Gruyter Mouton August 12, 2021

The Written British National Corpus 2014 – design and comparability

Vaclav Brezina, Abi Hawtin and Tony McEnery 

From the journal *Text & Talk*

<https://doi.org/10.1515/text-2020-0052>

Cite this

Abstract

The British National Corpus 2014 is a major project led by Lancaster University to create a 100-million-word corpus of present day British English. This corpus has been constructed as a comparable counterpart of the original British National Corpus (referred to as the BNC1994 in this article), which was compiled in the early 1990s. This article starts with the justification of the project answering the question of ‘Why do we need a new BNC?’. We then provide a general overview of the construction of the Written British National Corpus 2014 (Written BNC2014); we also briefly discuss some issues of data collection before looking in detail at the design of the corpus. Compiling a large general corpus such as the Written BNC2014 has been a major undertaking involving teamwork and collaboration. It also required generosity on the part of the many individuals and organisations who contributed to the data collection.


Keywords: [balance](#); [BNC1994](#); [comparability](#); [corpus description](#); [design principles](#); [representativeness](#); [sampling](#); [Written BNC2014](#)

1 Introduction

A synchronic corpus such as the British National Corpus 1994 (hereafter BNC1994) and the British National Corpus 2014 (hereafter BNC2014) provides a snapshot of language, and a window into social history, at the time of its compilation. Language and society, however, change over time. Consider the examples below. Example (1) is taken from the Written BNC1994, while example (2) represents a contribution to the Electronic language (E-language) subcorpus of the Written BNC2014.



The idea of the written BNC2014 (or how to build a corpus with nothing but hopes and wishes)

Tony McEnery
Lancaster University
 @TonyMcEnery

<http://cass.lancs.ac.uk>



WRITTEN

With a little help from my friends

CHALLENGES

- Capture regional and social variation.
- Capture register variation.
- Enhance balance.
- Represent 'British English'.
- Capture change in language and society.
- Deal with copyright.
- Stay true to the original while not repeating mistakes
- To learn from the scholarship on the BNC 1994

OPPORTUNITIES

- Update the BNC.
- Improve on the original.
- Understand **current** British English.
- Rethink sampling issues.
- Offer statistical sophistication (Xaira v. #LancsBox).
- Provide a free resource for researchers for non-profit making research.
- Give due weight to new forms of communication



Corpus sampling BNC2014

Genre/register	Proportion (BNC1994)	Proportion (BNC2014)
Academic prose	16.5%	20%
Fiction	16.6%	20%
Newspapers (Serious, mass market and regional)	9.6%	20%
Magazines	7.6%	15%
E-language	0.2%	5%
Other	39.5%	10%
Spoken	10%	10%

Myths and Legends – the Written BNC

- A good proxy for British English as written in the early 1990s “The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century”

- 1,718,449 words (1960-1974)

- 4,625,456 words (1975-1984)

- 79,728,156 words (1985-1993)

- “The BNC suggests that in the 1990’s”

2010	2%
2011	3%
2012	5%
2013	3%
2014	22%
2015	21%
2016	19%
2017	5%
2018	2%
2019	2%
2020	1%



BNC2014
data pool



Spoken
BNC2014



BNC2014
Baby +



Written
BNC2014

BNC 2014



Current British English

Vaclav Brezina (Lancaster University)

1994

2014



B BRITISH
N NATIONAL
C CORPUS

B **N** **C**
BRITISH
NATIONAL
CORPUS
2014

1994

2014



B RITISH
N ATIONAL
C ORPUS

1) Overall much more informal

2) Effect of language

3) The case of declining modals



BRITISH
NATIONAL
CORPUS

BNC214: AcaHumRa219

Faith As Extended Knowledge

You don't know that p unless it's on account of your cognitive abilities that you believe truly that p. Virtue epistemologists think there's some such ability constraint on knowledge. This looks to be in considerable tension, though, with putative faith-based knowledge. For at least on a popular Christian conception, when you believe something truly on the basis of faith this isn't because of anything you're naturally competent to do. Rather, faith-based beliefs are entirely a product of divine agency. Appearances to the contrary, I argue in this article that there's no deep tension between faith-based knowledge and virtue epistemology.

BNC214: AcaHumRa219

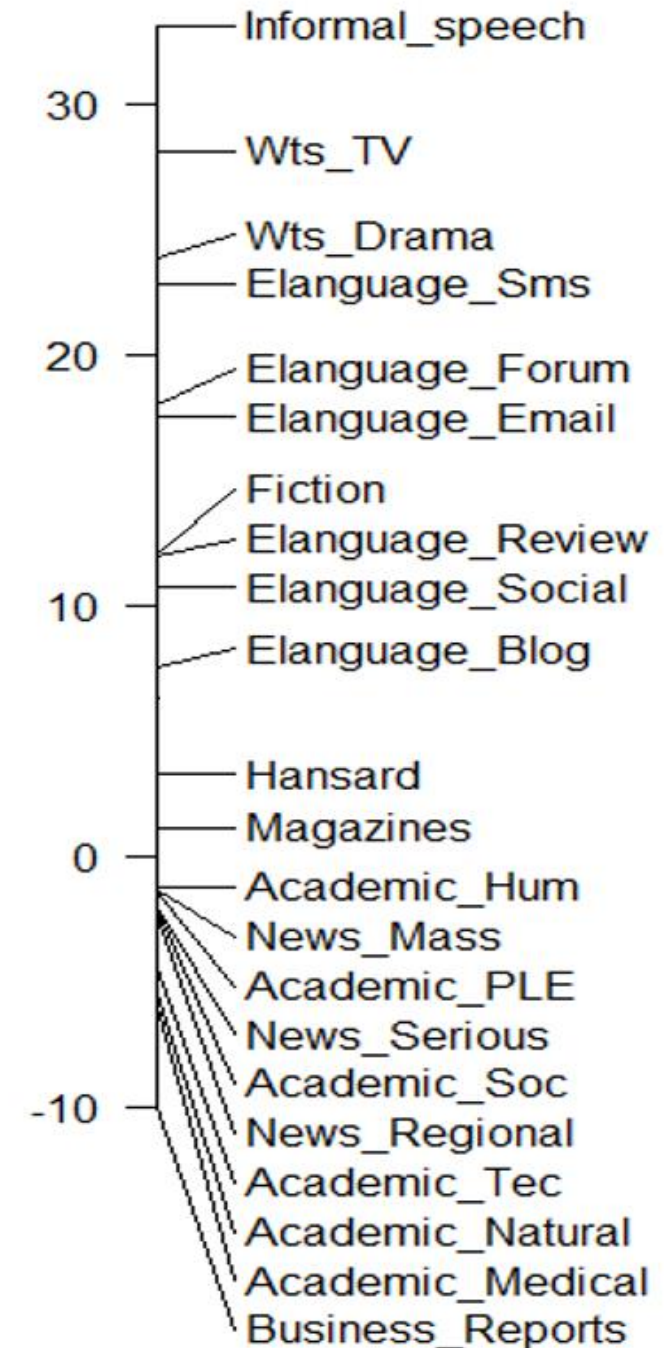
Faith As Extended Knowledge

You don't know that p unless it's on account of your cognitive abilities that you believe truly that p . Virtue epistemologists think there's some such ability constraint on knowledge. This looks to be in considerable tension, though, with putative faith-based knowledge. For at least on a popular Christian conception, when you believe something truly on the basis of faith this isn't because of anything you're naturally competent to do. Rather, faith-based beliefs are entirely a product of divine agency. Appearances to the contrary, I argue in this article that there's no deep tension between faith-based knowledge and virtue epistemology.

+ 28%

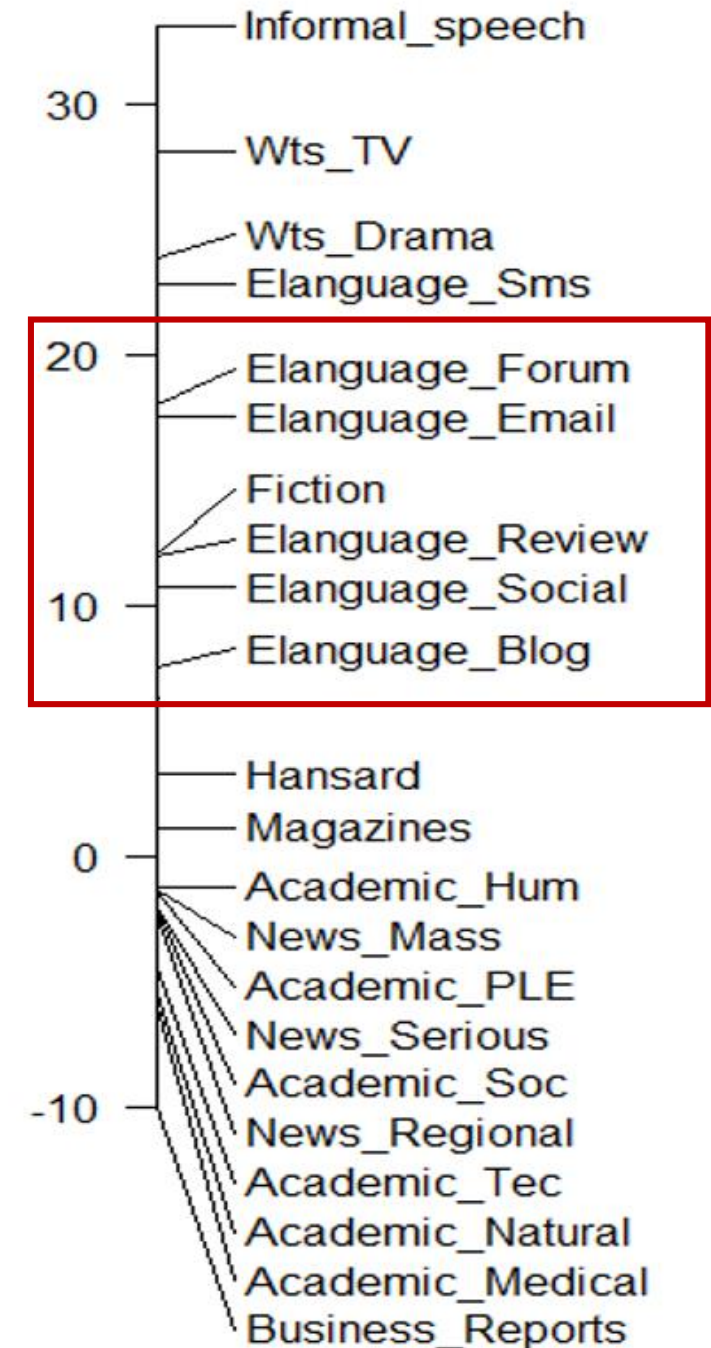
Involved
Contractions,
pronouns, private
verbs, emphatics...

Informational
Long words, nouns



Involved
Contractions,
pronouns, private
verbs, emphatics...

Informational
Long words, nouns



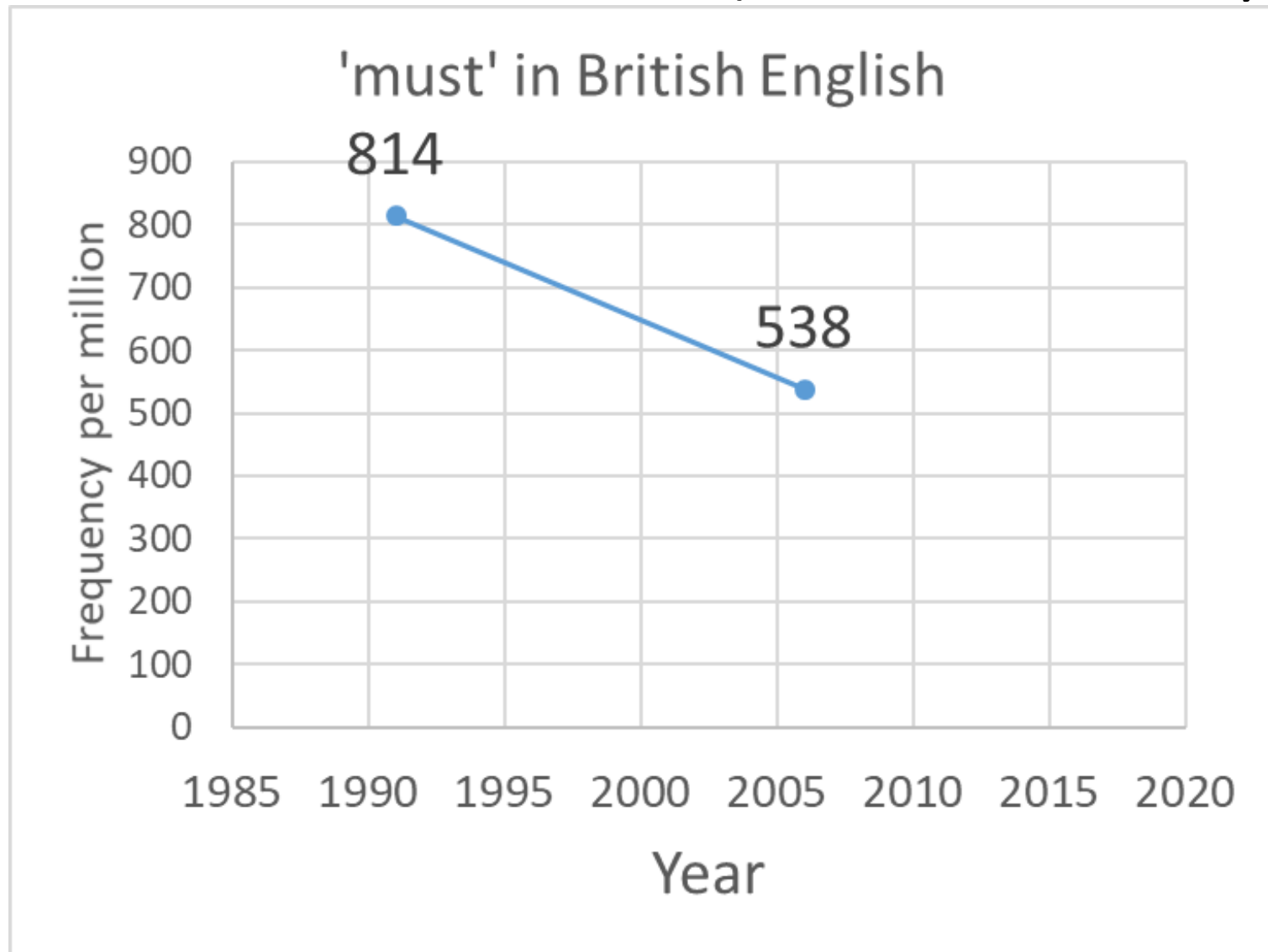
Elanguage

Popular words in elanguage in BNC2014: Alexa, App, Awesome, Blog, Congrats, Email, Fab, Facebook, Fitbit, Im, Instagram, ipad, iphone, Ive, Lol, Omg, Tbh, Tweet, Twitter and Website

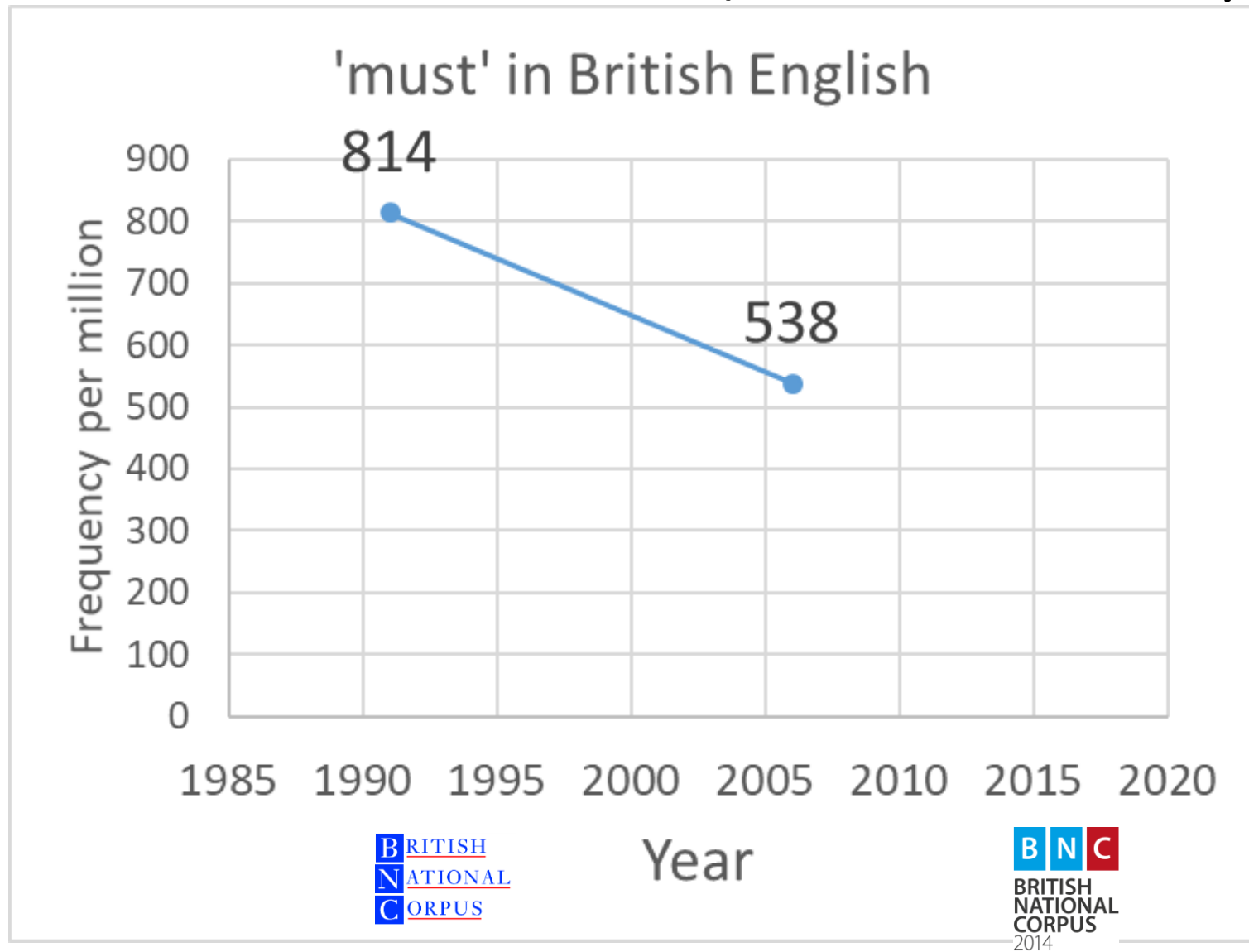


Software  App

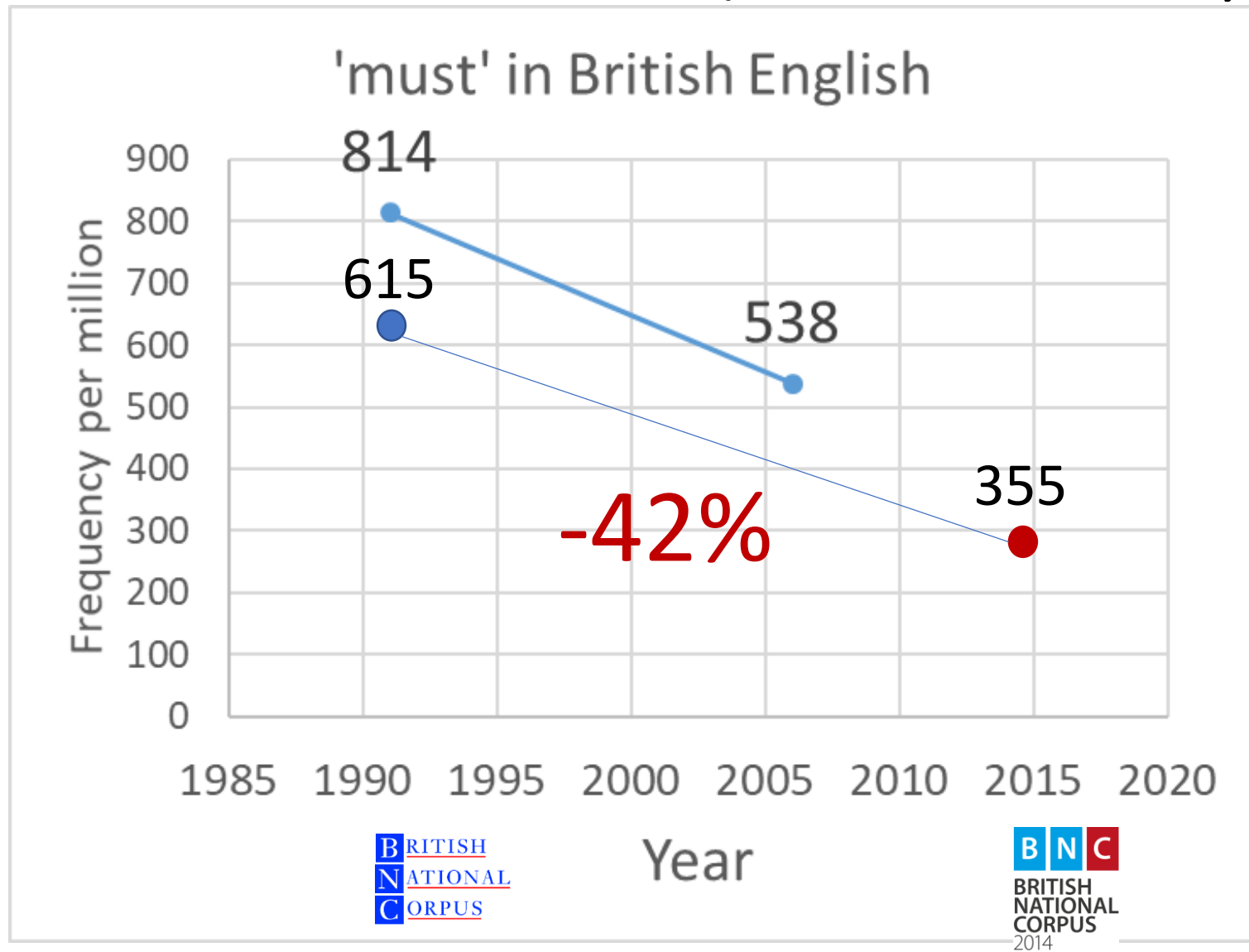
Modals on decline (Brown-family: 1M)



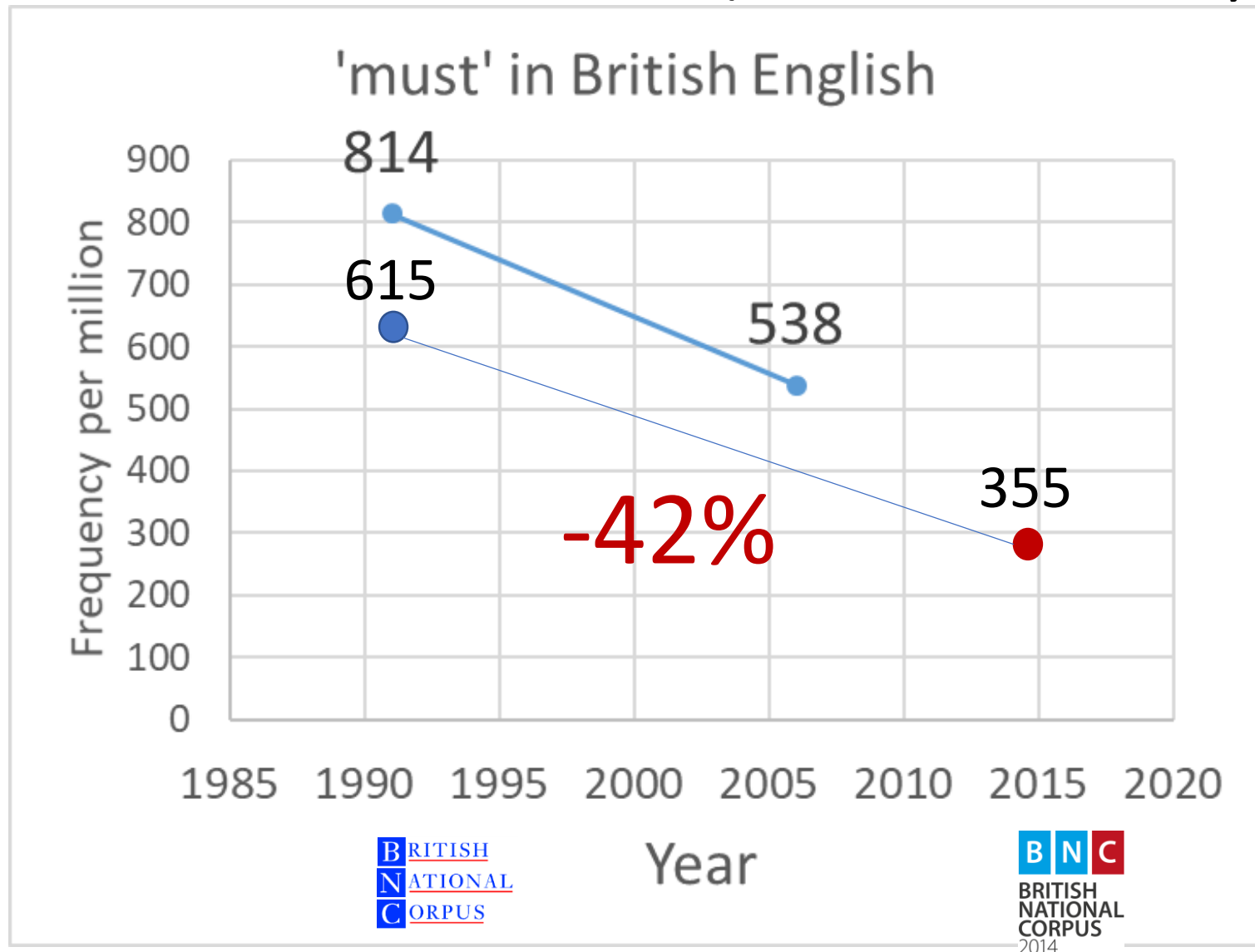
Modals on decline (Brown-family: 1M)



Modals on decline (Brown-family: 1M)



Modals on decline (Brown-family: 1M)



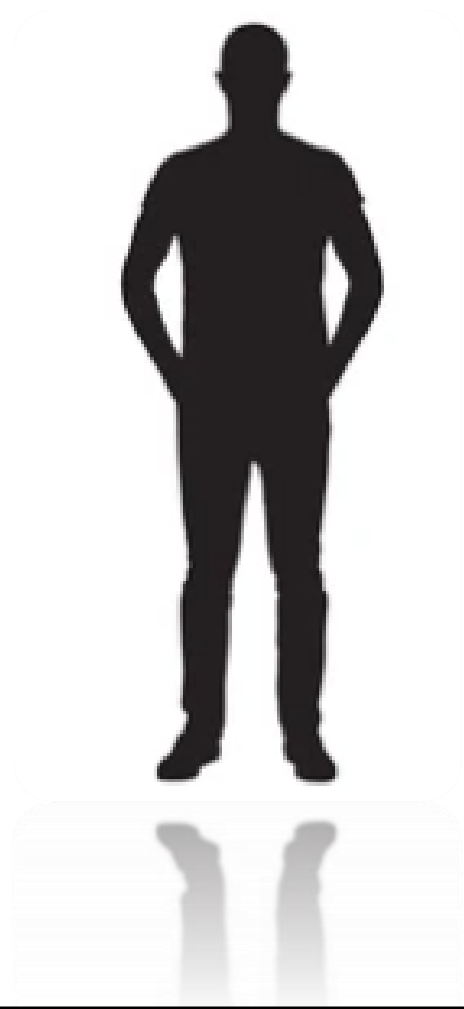
BRITISH
NATIONAL
CORPUS

BNC
BRITISH
NATIONAL
CORPUS
2014

May -40%

Shall -60%

Language development



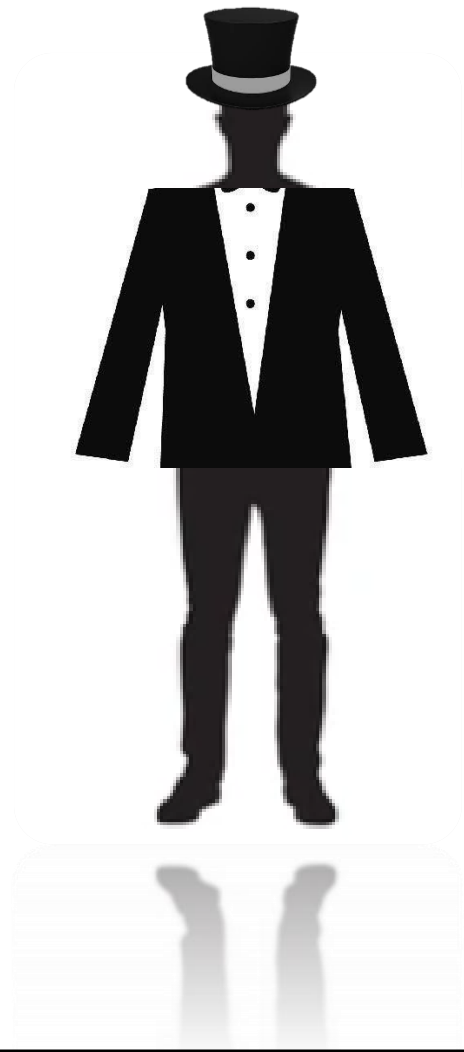
Language development



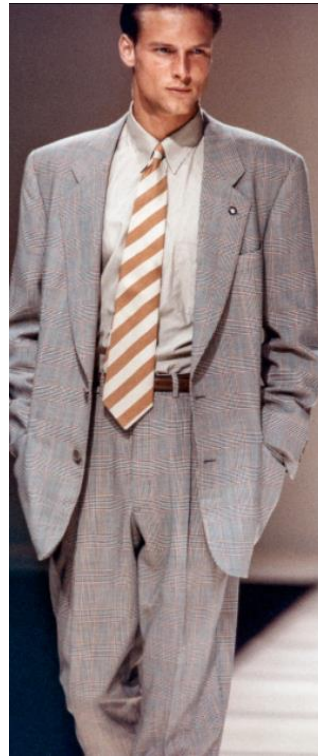
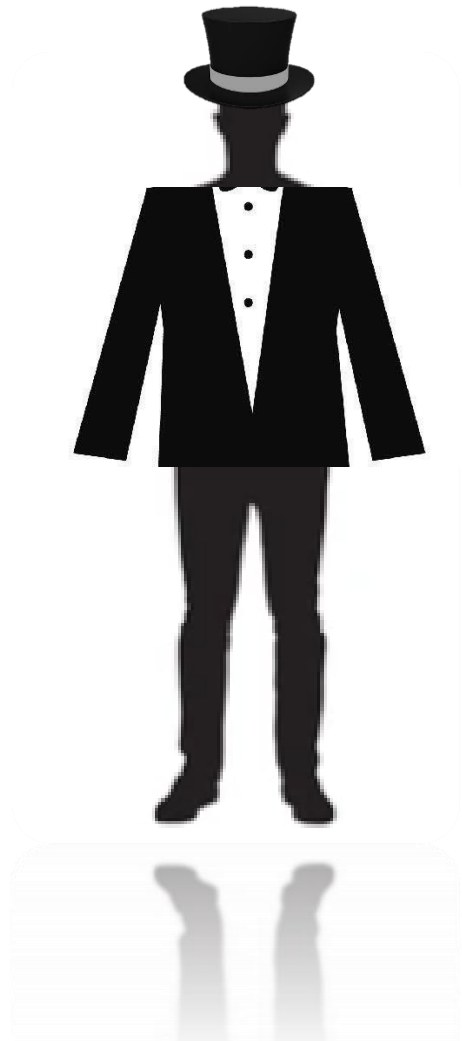
Language development



Language development



Language development



B BRITISH
N NATIONAL
C CORPUS



B **N** **C**
BRITISH
NATIONAL
CORPUS
2014

Exploring the BNC2014 using #LancsBox X

Vaclav Brezina & William Platt (Lancaster University)


```

</header>
<text id="ElanBlogAli1" mode="writing" genre="elanguage" subgenre="blog" section="personal" url="
http://www.alicesantics.co.uk/2014/02/the-lovehate-of-chick-flicks.html" sample="whole" source=
"AliceDodd" author="NA" pubDate="2014-01-22" words="721">
<title><s n="1"><w pos="AT" hw="the" class="ART" usas="Z5">The</w> <w pos="NN1" hw="love/hate" class=
"SUBST" usas="E2">Love/Hate</w> <w pos="IO" hw="of" class="PREP" usas="Z5">of</w><w pos="NN1" hw=
"'chick" class="SUBST" usas="Z99">'Chick</w> <w pos="NN2" hw="flicks" class="SUBST" usas="A1:1:1">
Flicks</w><w pos="GE" hw="'" class="UNC" usas="Z5">'</w></s></title>
<p n="1"><s n="2"><w pos="DA2" hw="many" class="ADJ" usas="N5">Many</w> <w pos="NN" hw="people" class=
"SUBST" usas="S2">people</w> <w pos="VV0" hw="know" class="VERB" usas="X2:2">know</w><c>,</c> <w pos=
"PPIS1" hw="i" class="PRON" usas="Z8">I</w> <w pos="VBM" hw="be" class="VERB" usas="A3">am</w> <w
pos="JJ" hw="partial" class="ADJ" usas="N5:1">partial</w> <w pos="II" hw="to" class="PREP" usas="Z5">
to</w> <w pos="AT" hw="the" class="ART" usas="Z5">the</w> <w pos="JJ" hw="odd" class="ADJ" usas=
"A6:2">odd</w> <c>'</c><w pos="NN1" hw="chick" class="SUBST" usas="L2">Chick</w> <w pos="NN1" hw=
"flick" class="SUBST" usas="A1:1:1">Flick</w><c>.</c><c>'</c></s> <s n="3"><w pos="PPIO1" hw="i"
class="PRON" usas="Z8">Me</w> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w pos="APPGE" hw=
"my" class="PRON" usas="Z8">my</w> <w pos="JJT" hw="good" class="ADJ" usas="A5:1">best</w> <w pos=
"NN1" hw="friend" class="SUBST" usas="S3:1">friend</w> <w pos="VH0" hw="have" class="VERB" usas="Z5">
have</w> <w pos="VBN" hw="be" class="VERB" usas="Z5">been</w> <w pos="VVN" hw="know" class="VERB"
usas="X2:2">known</w> <w pos="TO" hw="to" class="PREP" usas="Z5">to</w> <w pos="VVI" hw="sit" class=
"VERB" usas="M8">sit</w> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w pos="VVI" hw="watch"
class="VERB" usas="X2:2">watch</w><c>.</c></s> <s n="4"><w pos="JJ" hw="first" class="ADJ" usas="N4">first</w> <w pos="MC" hw="20" class="ADJ" usas="N1">20</w> <w pos="JJ"
hw="minute" class="ADJ" usas="N3:2">minute</w> <w pos="NN1" hw="dialogue" class="SUBST" usas="Q2:1">
dialogue</w> <w pos="IO" hw="of" class="PREP" usas="Z5">of</w> <c>'</c><w pos="ZZ1" hw="a" class=
"SUBST" usas="Z5">A</w> <w pos="NP1" hw="cinderella" class="SUBST" usas="Z1">Cinderella</w> <w pos=
"NN1" hw="story" class="SUBST" usas="Q2:1">Story</w><c>'</c> <w pos="IW" hw="with" class="PREP" usas=
"Z5">with</w> <w pos="DD1" hw="another" class="ADJ" usas="A6:1">another</w> <w pos="NN1" hw="friend"
class="SUBST" usas="S3:1">friend</w><c>,</c> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w
pos="EX" hw="there" class="PRON" usas="Z5">there</w> <w pos="VBZ" hw="be" class="VERB" usas="A3">is
</w> <w pos="PN1" hw="nothing" class="PRON" usas="Z6">nothing</w> <w pos="PPIS1" hw="i" class="PRON"
usas="Z8">I</w> <w pos="VV0" hw="like" class="VERB" usas="E2">like</w> <w pos="RRR" hw="more" class=
"ADV" usas="A13:3">more</w> <w pos="II" hw="on" class="PREP" usas="Z5">on</w> <w pos="AT1" hw="a"
class="ART" usas="Z5">a</w> <w pos="NP1" hw="couple" class="SUBST" usas="Z1">a couple</w> <w pos=
"s="Z1">last</w> <w pos=
"VVI" hw="sit" class="VERB" usas="M8">sit</w> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w pos="VVI" hw="watch"
class="VERB" usas="X2:2">watch</w><c>.</c></s> <s n="4"><w pos="JJ" hw="first" class="ADJ" usas="N4">first</w> <w pos="MC" hw="20" class="ADJ" usas="N1">20</w> <w pos="JJ"
hw="minute" class="ADJ" usas="N3:2">minute</w> <w pos="NN1" hw="dialogue" class="SUBST" usas="Q2:1">
dialogue</w> <w pos="IO" hw="of" class="PREP" usas="Z5">of</w> <c>'</c><w pos="ZZ1" hw="a" class=
"SUBST" usas="Z5">A</w> <w pos="NP1" hw="cinderella" class="SUBST" usas="Z1">Cinderella</w> <w pos=
"NN1" hw="story" class="SUBST" usas="Q2:1">Story</w><c>'</c> <w pos="IW" hw="with" class="PREP" usas=
"Z5">with</w> <w pos="DD1" hw="another" class="ADJ" usas="A6:1">another</w> <w pos="NN1" hw="friend"
class="SUBST" usas="S3:1">friend</w><c>,</c> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w
pos="EX" hw="there" class="PRON" usas="Z5">there</w> <w pos="VBZ" hw="be" class="VERB" usas="A3">is
</w> <w pos="PN1" hw="nothing" class="PRON" usas="Z6">nothing</w> <w pos="PPIS1" hw="i" class="PRON"
usas="Z8">I</w> <w pos="VV0" hw="like" class="VERB" usas="E2">like</w> <w pos="RRR" hw="more" class=
"ADV" usas="A13:3">more</w> <w pos="II" hw="on" class="PREP" usas="Z5">on</w> <w pos="AT1" hw="a"
class="ART" usas="Z5">a</w> <w pos="NP1" hw="couple" class="SUBST" usas="Z1">a couple</w> <w pos=
"s="Z1">last</w> <w pos=
"VVI" hw="sit" class="VERB" usas="M8">sit</w> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w pos="VVI" hw="watch"
class="VERB" usas="X2:2">watch</w><c>.</c></s> <s n="4"><w pos="JJ" hw="first" class="ADJ" usas="N4">first</w> <w pos="MC" hw="20" class="ADJ" usas="N1">20</w> <w pos="JJ"
hw="minute" class="ADJ" usas="N3:2">minute</w> <w pos="NN1" hw="dialogue" class="SUBST" usas="Q2:1">
dialogue</w> <w pos="IO" hw="of" class="PREP" usas="Z5">of</w> <c>'</c><w pos="ZZ1" hw="a" class=
"SUBST" usas="Z5">A</w> <w pos="NP1" hw="cinderella" class="SUBST" usas="Z1">Cinderella</w> <w pos=
"NN1" hw="story" class="SUBST" usas="Q2:1">Story</w><c>'</c> <w pos="IW" hw="with" class="PREP" usas=
"Z5">with</w> <w pos="DD1" hw="another" class="ADJ" usas="A6:1">another</w> <w pos="NN1" hw="friend"
class="SUBST" usas="S3:1">friend</w><c>,</c> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w
pos="EX" hw="there" class="PRON" usas="Z5">there</w> <w pos="VBZ" hw="be" class="VERB" usas="A3">is
</w> <w pos="PN1" hw="nothing" class="PRON" usas="Z6">nothing</w> <w pos="PPIS1" hw="i" class="PRON"
usas="Z8">I</w> <w pos="VV0" hw="like" class="VERB" usas="E2">like</w> <w pos="RRR" hw="more" class=
"ADV" usas="A13:3">more</w> <w pos="II" hw="on" class="PREP" usas="Z5">on</w> <w pos="AT1" hw="a"
class="ART" usas="Z5">a</w> <w pos="NP1" hw="couple" class="SUBST" usas="Z1">a couple</w> <w pos=
"s="Z1">last</w> <w pos=
"VVI" hw="sit" class="VERB" usas="M8">sit</w> <w pos="CC" hw="and" class="CONJ" usas="Z5">and</w> <w pos="VVI" hw="watch"
class="VERB" usas="X2:2">watch</w><c>.</c></s> <s n="4"><w pos="JJ" hw="first" class="ADJ" usas="N4">first</w> <w pos="MC" hw="20" class="ADJ" usas="N1">20</w> <w pos="JJ"
hw="minute" class="ADJ" usas="N3:2">minute</w> <w pos="NN1" hw="dialogue" class="SUBST" usas="Q2:1">
dialogue</w> <w pos="IO" hw="of" class="PREP" usas="Z5">of</w> <c>'</c><w pos="ZZ1" hw="a" class=
"SUBST" usas="Z5">A</w> <w pos="NP1" hw="cinderella" class="SUBST" usas="Z1">Cinderella</w> <w pos=
"NN1" hw="story" class
```



INTRO

DOWNLOAD

MATERIALS

USER GUIDE

NEWS

SEARCH



#LancsBox: Lancaster University corpus toolbox

Download version X: suitable for large corpora such as the British National Corpus 2014

Read [#LancsBox X manual \(English\)](#). [Japanese version](#) of the manual is also available.

#LancsBox X for Windows



Download



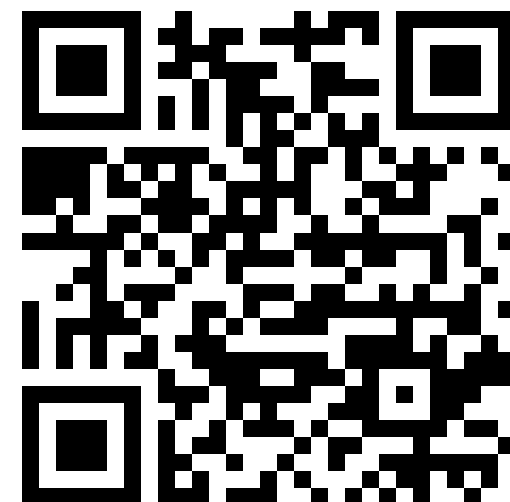
[Windows](#)



[Mac](#)



[Linux](#)



INTRODUCING #LANCSBOX X

#LancsBox X 1.0.0

Q ADJECTIVE [hw="food"]

BNC1994

whole corpus

97M

ADJECTIVE [hw="food"]

Hits: 4,467 (0.46)

Texts: 1,196/7,531

File	Left	Node	Right
Oth...	International plc operates a	fast food	franchise. Specialities refle...
Fict...	Cati's stomach like	undigested food;	Rosalba couldn't possibly
Fict...	...en included beans among	prohibited foods.	Although southern Italians...
Oin...	day. [Scene 4] Food,	good food,	comes to the ward
Fict...	the change from the	hot food	to the offerings of
Oth...	that you live on	dehydrated food,	and since at the
Mag...	not readily come to	hanging food	and birdtables. The Birdfo...
Oth...	are already too many	fast food	operators providing meat i...
Oth...	...ing enormous amounts of	expensive food	and drink if they
Fict...	I would be kept	alive. Food	and drink would be

BNC2014

whole corpus

100M

ADJECTIVE [hw="food"]

Hits: 5,764 (0.58)

Texts: 2,678/88,171

File	Left	Node	Right
Elan...	Having read about the	wonderful food	at Ducksoup on Rosie
Wts...	of students licking the	cheap food	out of their hideous
Wts...	eats a lot of	posh food.	He has no other
New...	peace and dine on	organic food	from the prison gardens.
New...	to afford to buy	good food (May 5). In my
New...	to spend time preparing	good food.	I raised my children
Fict...	saucers, some tinned and	dried food,	baking trays that have
Mag...	is perfect because the	Whole Foods	business model - organic and
Mag...	...iness model - organic and	expensive food -	was preventing it reaching
Mag...	do so, it launched	Whole Foods	365, effectively offering a

Search completed.

SMARTER.

SLEEKER.

FASTER.

> **SMARTER.**
SLEEKER.
FASTER.

X IS FOR XML

<u gender="male"/>

<pause/> within <u gender="male"/>

by within PREPOSITIONAL_PHRASE

<s/> within yes

(<s/> within yes) within <u gender="male"/>

> **SMARTER.**
SLEEKER.
FASTER.

SUBCORPORA

BNC2014 whole corpus 100M

Define new subcorpus Name: academic prose: humanities book

▼ Text

mode

- ☐ speech
- ☐ writing

genre

- ☐ academic prose
- ☐ elanguage
- ☐ fiction
- ☐ informal speech
- ☐ magazines
- ☐ newspapers
- ☐ official documents
- ☐ written-to-be-spoken

subgenre

- ☒ academic prose: humanities
- ☐ academic prose: medicine
- ☐ academic prose: natural sciences
- ☐ academic prose: politics/law/education
- ☐ academic prose: social sciences
- ☐ academic prose:

subsubgenre

- ☐ academic prose: humanities: archaeology
- ☐ academic prose: humanities:

sample

- ☐ beginning
- ☐ composite
- ☐ end

academic publication

- ☒ book
- ☐ journal

OK Cancel

> **SMARTER.**
SLEEKER.
FASTER.

SUBCORPORA

BNC2014 whole corpus 100M

Hits:

File	
------	--

- academic prose
- elanguage
- fiction
- informal speech
- magazines
- newspapers
- official documents
- written-to-be-spoken
- academic prose: humanit...


BNC2014 academic prose: humanities book 727K

ADVERB Hits: 39,490 (543.14) Texts: 97/97

727,072 words across 97 files

> **SMARTER.**
SLEEKER.
FASTER.

SUBCORPORA

Define new subcorpus Name: 

▶ Text

▼ Dialogue: utterance

gender	age	age range
<input type="checkbox"/> female	<input type="checkbox"/> 12	<input type="checkbox"/> 0-10
<input type="checkbox"/> male	<input type="checkbox"/> 13	<input type="checkbox"/> 11-18
<input type="checkbox"/> NA	<input type="checkbox"/> 14	<input type="checkbox"/> 19-29
	<input type="checkbox"/> 15	<input type="checkbox"/> 30-39
	<input type="checkbox"/> 16	<input type="checkbox"/> 40-49
	<input type="checkbox"/> 17	<input type="checkbox"/> 50-59
	<input type="checkbox"/> 18	<input type="checkbox"/> 60-69
	<input type="checkbox"/> 19	<input type="checkbox"/> 70-79

socio-economic category	education	dialect
<input type="checkbox"/> A	<input type="checkbox"/> 1-primary	<input type="checkbox"/> England
<input type="checkbox"/> B	<input type="checkbox"/> 2-secondary	<input type="checkbox"/> Northern Ireland

Define new subcorpus

▶ Text

▶ Dialogue: utterance

▶ Composite: element

▼ Paragraph

paragraph number

<input type="checkbox"/> 1
<input type="checkbox"/> 2
<input type="checkbox"/> 3
<input type="checkbox"/> 4
<input type="checkbox"/> 5
<input type="checkbox"/> 6
<input type="checkbox"/> 7
<input type="checkbox"/> 8

▶ Sentence

> **SMARTER.**
SLEEKER.
FASTER.

COOL COLUMNS

BNC2014		informal speech		10M
[hw="food"]		Hits: 2,863 (2.77)	Texts: 705/1,251	
Left	Node	Right	Text: genre	
them to pay for my	food	and what else and cos	informal speech	
I'm running out of	food	no I mean I have	informal speech	
I have run out of	food	I don't have anything	informal speech	
you've run out of	food	no that's fine that	informal speech	
there'll be like the	food	shop they don't even	informal speech	
we still need two single	food	place oh yeah I'm	informal speech	
is maybe the lack of	food	and maybe they were just	informal speech	
and that there or take	food	with you? yeah no eat	informal speech	
there selling all kinds of	food	hot food yeah so you	informal speech	
all kinds of food hot	food	yeah so you don't	informal speech	

> **SMARTER.**
SLEEKER.
FASTER.

COOL COLUMNS

BNC2014		informal speech		10M	
[hw="food"]		Hits: 2,863 (2.77)		Texts: 705/1,251	
Left	Node	Right	Text: genre	Dialogue: utterance: age	Dialogue: utterance: birthplace
...ay for my	food	and what el...	informal speech	20	Falmouth
...ing out of	food	no I mean I ...	informal speech	20	Falmouth
...run out of	food	I don't have ...	informal speech	20	Falmouth
...run out of	food	no that's fin...	informal speech	19	Liverpool
...e like the	food	shop they d...	informal speech	19	Liverpool
...wo single	food	place oh ye...	informal speech	57	New Zealand
...he lack of	food	and maybe t...	informal speech	31	Edinburgh
...re or take	food	with you? ye...	informal speech	27	Gillingham, Kent
...ll kinds of	food	hot food ye...	informal speech	69	London
...f food hot	food	yeah so you...	informal speech	69	London

SMARTER.

> SLEEKER.

FASTER.

SMARTER.
> SLEEKER.
FASTER.

SIDE-BY-SIDE TOOLS

#LancsBox X 1.0.0

Q [hw="food"]

BNC2014 informal speech 10M

[hw="food"] Hits: 2,863 (2.77) Texts: 705/1,251

File	Left	Node	Right
Sp0m2...	them to pay for my	food	and what else and cos
Sp0m2...	I'm running out of	food	no I mean I have
Sp0m2...	I have run out of	food	I don't have anything
Sp0m2...	you've run out of	food	no that's fine that
Sp0m2...	there'll be like the	food	shop they don't even
Sp2m2...	we still need two single	food	place oh yeah I'm
Sp2m0...	is maybe the lack of	food	and maybe they were just
Sp1m1...	and that there or take	food	with you? yeah no eat
Sp1m1...	there selling all kinds of	food	hot food yeah so you
Sp1m1...	all kinds of food hot	food	yeah so you don't
Sp1m1...	got enough f money for	food	and a couple of drinks
Sp1m1...	I don't each much	food	so that kind of just
Sp1m1...	kind of just covers my	food	no you're quite small
Sp1m1...	I wouldn't eat my	food	of that yeah again thank

Search completed.

SMARTER.
> SLEEKER.
FASTER.

REARRANGE AND RESIZE

#LancsBox X 1.0.0

apple

BNC2014 whole corpus

apple Hits: 4,192 (0.48) Texts: 2,236/88,171

File	Left	N...	Right
...	... Design Museum. Last year,	...	's Jony Ive and I
...	...d place behind Android and	...	s iOS . However, the Nokia
...	the likes of Samsung and	...	If you look at the
...	out? yeah after being the	...	of the eye and everything

BNC2014 informal speech

[hw="f... Hits: 2,863 (2.77) Texts: 705/1,251

File	Left	Node	Right
...	them to pay for my	food	and what else and cos
...	I'm running out of	foo...	no I mean I have
...	I have run out of	food	I don't have anything
...	you've run out of	food	no that's fine that
...	there'll be like the	food	shop they don't even
...	we still need two single	food	place oh yeah I'm
...	is maybe the lack of	food	and maybe they were just
...	and that there or take	food	with you? yeah no eat
...	there selling all kinds of	food	hot food yeah so you
...	all kinds of food hot	food	yeah so you don't
...	got enough f money for	food	and a couple of drinks
...	I don't each much	food	so that kind of just
...	kind of just covers my	food	no you're quite small

BNC2014 whole corpus

Hits: Texts:

File	Left	Node	Right
No content in table			

Search completed.

SMARTER.

SLEEKER.

> FASTER.

**SMARTER.
SLEEKER.
> FASTER.**

FASTER RESULTS

No corpus loading

Real-time updates

The screenshot shows the #LancsBox X 1.0.0 application window. At the top, there is a search bar with a magnifying glass icon and a vertical cursor. Below the search bar, a yellow header bar contains the text "BNC2014", a dropdown menu set to "whole corpus", and the text "100M". Below the header, there is a table with columns "File", "Left", "Node", and "Right". The table is currently empty, with the text "No content in table" centered below the header. At the bottom of the window, the text "Search completed." is displayed. A yellow border highlights the main content area of the application.

File	Left	Node	Right
No content in table			

Quick stats

**SMARTER.
SLEEKER.
> FASTER.**

BIGGER CORPORA

Search times on the BNC2014:

Word (freq. per 10K)	Time	
	Statistics	Full concordance
contextual (0.07)	< 1 sec	1 sec
food (2.13)	1 sec	12 sec
be (54.02)	1 sec	1 min 30 sec

Full times are most dependent on number of file matches, not whole corpus size, so even larger corpora are still fast for infrequent searches.

DOWNLOAD

<http://corpora.lancs.ac.uk/lancsbox/downloadx.php>



SMARTER. SLEEKER. FASTER.

“One more thing...”

