



Synchronic corpora

Kurkina Ekaterina

LMag-101



Synchronic Corpus

In linguistics, *synchrony* (as opposed to diachrony) refers to the analysis of language at one given point in time, usually the present, without taking its history or development into account. *Thus, a synchronic corpus provides data from one point in time.*

SPOKEN (300)	Dialogues (180)	Private (100)	Face-to-face conversations (90) Phonecalls (10)
		Public (80)	Classroom Lessons (20) Broadcast Discussions (20) Broadcast Interviews (10) Parliamentary Debates (10) Legal cross-examinations (10) Business Transactions (10)
	Monologues (120)	Unscripted (70)	Spontaneous commentaries (20) Unscripted Speeches (30) Demonstrations (10) Legal Presentations (10)
		Scripted (50)	Broadcast News (20) Broadcast Talks (20) Non-broadcast Talks (10)
WRITTEN (200)	Non-printed (50)	Student Writing (20)	Student Essays (10) Exam Scripts (10)
		Letters (30)	Social Letters (15) Business Letters (15)
	Printed (150)	Academic writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Popular writing (40)	Humanities (10) Social Sciences (10) Natural Sciences (10) Technology (10)
		Reportage (20)	Press news reports (20)
		Instructional writing (20)	Administrative Writing (10) Skills/hobbies (10)
		Persuasive writing (10)	Press editorials (10)
	Creative writing (20)	Novels & short stories (20)	

The International Corpus of English

- The ICE corpus consists of a collection of twenty corpora of one million words each, each composed of written and spoken English.
- All ICE corpora contain 500 texts of approximately 2,000 words each.
-
- Some corpora can be downloaded.

The following corpora are available free (under Licence) to **download** from this site:

CANADA (ICE-CAN - 1m words, lexical)

JAMAICA (ICE-JA - 1m words, lexical)

HONG KONG (ICE-HK - 1m words, lexical)

EAST AFRICA (ICE-EA - Kenya & Tanzania, v 2, 1m words, lexical, including a version for Wordsmith)

INDIA (ICE-IND - 1m-words, lexical)

SINGAPORE (ICE-SIN - 1m words, lexical)

PHILIPPINES (ICE-PHI - 1m words, lexical)

USA (ICE-USA, written component - c.400,000 words, lexical)

IRELAND (ICE-IRL - 1m words, lexical)

SPICE-IRELAND (SPICE-IRL - c.600,000 words with prosodic and pragmatic annotation)

The following corpora are also available, from the addresses shown:

GREAT BRITAIN (ICE-GB - 1m words, POS-tagged and parsed, distributed with ICECUP retrieval software) . Available from: Survey of English Usage, University College London, Gower St, London WC1E 6BT, UK. [Order Form](#)

NEW ZEALAND (ICE-NZ - 1m words, lexical)

Available from: School of Linguistics & Applied Language Studies, Victoria University of Wellington PO Box 600, Wellington, New Zealand. [Order form](#)

SRI LANKA (ICE-SL - written component; lexical and POS-tagged with CLAWS C7 tagset)

Available from the Department of English, the University of Giessen, Germany.

To obtain a copy of the corpus, please email ice-sl@anglistik.uni-giessen.de

NIGERIA (ICE-NG - written component). The corpus is [available here](#).

Please note that texts in all ICE corpora are protected by copyright law. They are available for **non-profit academic research** purposes only.

PEARSON

Longman

FIG 1.

the theatre, stayed to **haunt** her for the rest of her life.
his assassin he will not **haunt** him. Ah, the past is filled
woman began to **haunt** him, and not only in his dream
the eyebrow continued to **haunt** him, and the Coming, so
anxious fantasies will **haunt** him; but a witch he can push
eased for a single day to **haunt** him. What they whispered in
vulnerable expression **haunt** his eyes, and wished I hadn't

FIG 2.

haunt [T *not in progressive forms*] **1** [*often pass.*] (of spirit, esp. of a dead person) to visit (a place), appearing in a strange form: *The ghost of a headless man haunts the castle / a haunted house* **2** [*usu. pass*] (esp. of something strange or sad) to be always in the thoughts of (someone): *I was haunted by his last words to me. She had a haunted look, as if she were constantly anxious or afraid.* **3** [*infml*] to visit (a place) regularly; FREQUENT

The Longman/Lancaster Corpus

- It is 30 million words of written language taken from literature, magazines, papers and more ephemeral materials such as leaflets and packaging.
- Both imaginative and informative text categories are included.
- The corpus is not available for public access.

The Longman Written American Corpus

- 100 million words comprising running text from newspapers, journals, magazines, best-selling novels, technical and scientific writing, and coffee-table books.
- Words can be concordanced, wordlists created, and statistical features analysed, allowing lexicographers to compare and contrast usage in British and American English.
- This corpus does not appear to allow public access.

The CREA corpus of Spanish

- The corpus currently contains 133 million words sampled from a wide range of written (90%) and spoken (10%) text categories.
- The corpus is accessible on the Internet.



Real Academia Española - Corpus de Referencia del Español Actual (CREA)

Consulta:	<input type="text"/>		
Criterios de selección:			
Autor:	<input type="text"/>	Obra:	<input type="text"/>
Cronológico:	<input type="text"/> <input type="text"/>	Medio:	Geográfico:
		(Todos) ▲ Libros Periódicos Revistas Miscelánea Oral ▼	(Todos) ▲ Argentina Bolivia Chile Colombia Costa Rica ▼
Tema:	(Todos) ▲ 1.- Ciencias y Tecnología. 101.- Biología. 102.- Veterinaria. 103.- Ecología. 104.- Tecnología. ▼		

The LIVAC corpus of Chinese

- By 2020, 2.7 billion characters of news media texts have been filtered so far, of which 680 million characters have been processed and analyzed and have yielded an expanding Pan-Chinese dictionary of 2.3 million words from the Pan-Chinese printed media.

Input a word or a word fragment. Change parameters below as needed. Only traditional Chinese characters are accepted.

Enter query: Search

Match Mode: Partial Match Exact Match As Prefix Only As Suffix Only As Infix Only

Word Length: From To Characters

Word Class: No Tags Personal Names Place Names Organizational Names
 Reduplication Alias Haplogy Number
 Other Nouns Works Names

[Reset query parameters]
[Clear results and Start over]

6968 results have been retrieved with the above query parameters.

Please select words for geographical distributions by checking the boxes beside them. Specify the target date range, and click "Show Geographical Distribution" button.

Note: your current access level entitles you to access corpus data spanning from 1995-07-01 to 2011-06-28.

Shortcuts for selection:
[select all] [clear all] [toggle all]

Show Geographical Distribution From: To:

<input type="checkbox"/> (1) 車	Ordinary Word (一般詞類)	(Count article by year) (Count article by LIVAC yearcode)
<input type="checkbox"/> (2) 三車	Ordinary Word (一般詞類)	(Count article by year) (Count article by LIVAC yearcode)
<input type="checkbox"/> (3) 上車	Ordinary Word (一般詞類)	(Count article by year) (Count article by LIVAC yearcode)
<input type="checkbox"/> (4) 下車	Ordinary Word (一般詞類)	(Count article by year) (Count article by LIVAC yearcode)
<input type="checkbox"/> (5) 中車	Ordinary Word (一般詞類)	(Count article by year) (Count article by LIVAC yearcode)

- The corpus is accessible on the Internet.

Thank you for attention!