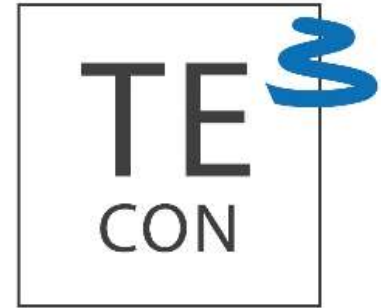


Instytut  
Lingwistyki  
Stosowanej

Erasmus+



# Corpus-based tasks

Agnieszka Leńko-Szymańska



UNIwersytet  
Warszawski

# Corpus



a collection of texts



in the digital format



compiled according to specific design criteria



for a specific purpose

# Purpose



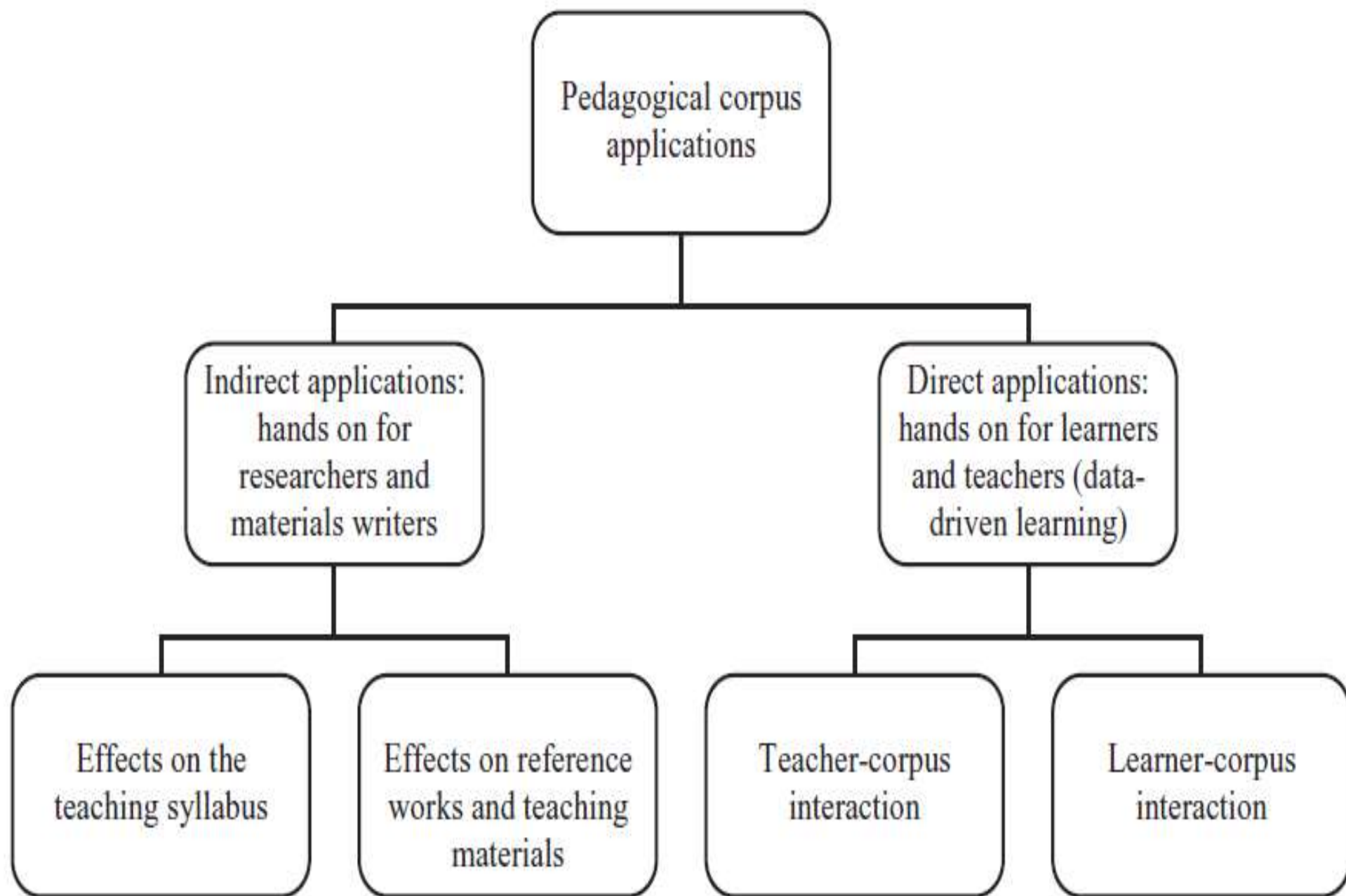
## research

linguistics

social sciences



## teaching (foreign languages)



# Researchers and materials writers



More accurate description of language

frequency  
meaning in context  
patterns  
etc.



Identifying problematic areas (learner corpora)



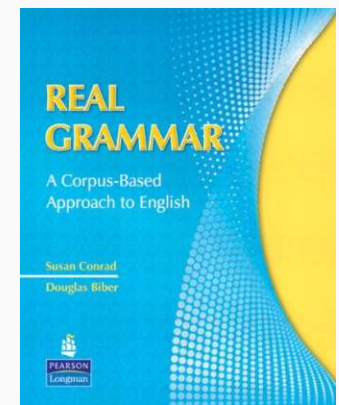
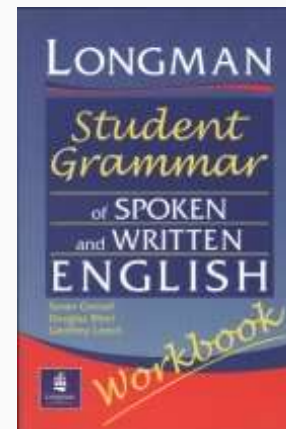
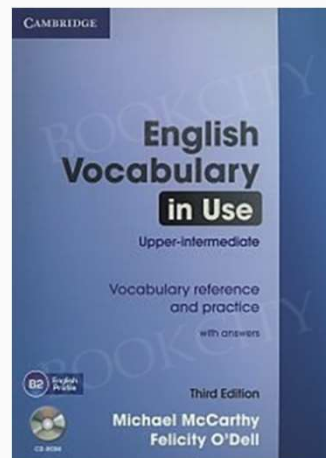
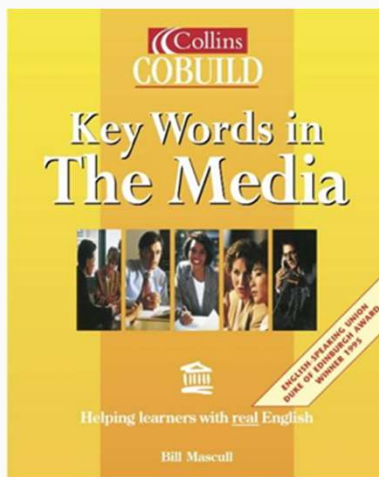
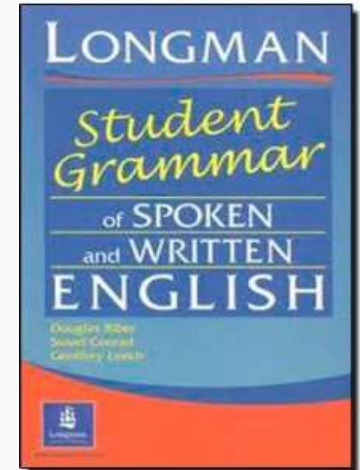
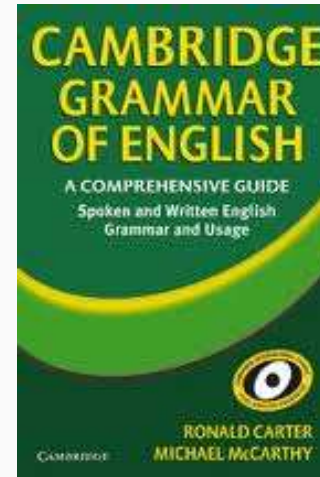
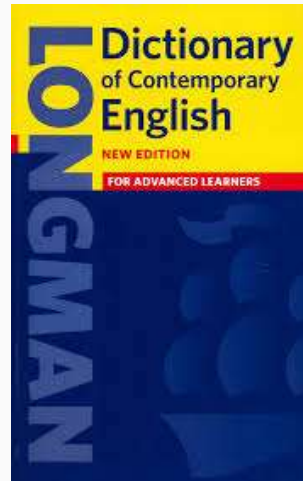
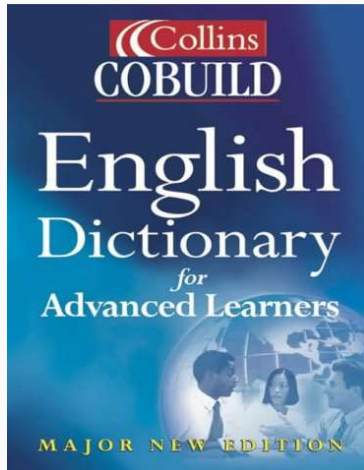
Syllabus design (especially ESP courses)



Production of teaching materials

authentic language examples  
information in learners' dictionaries

# Corpus-based materials for EFL learners



# For teachers



reference (information about language)



materials development (addressing students' specific needs)



**LSP**

# For learners



contact with authentic language



developing language awareness



developing learner autonomy



reference



# But corpora

are difficult to  
acquire or compile

require advanced  
technical skills

are cognitively  
challenging (require  
language analytical  
skills)

are demanding  
timewise

contain advanced  
and „messy“  
language

promote studying  
language over  
developing  
communication skills

# Corpus-based tasks in TE-Con3



4 MODULES



6 LSP CORPORA



11 DATA-DRIVEN  
ACTIVITIES

# Corpora

corpus	texts	size	compilation method
Architecture	708	1.4 M	journal articles
Bioethics	427	1.2 M	web crawler <- keywords from journals
Cars in cities	449	1.1 M	web crawler <- keywords from a text
Automated car manufacturing	828	2.5 M	web crawler <- keywords from a text
COCA, section MAG: Soc/Arts)		9.3 M	section of a general corpus
COCA, virtual corpus on Media		2.3 M	virtual corpus built from a general corpus <- keywords from tasks

# Activities

Module	Topic	Focus
Architecture	wood vs wooden	meanings
	design	collocations
	aim	colligations
	present perfect	grammar
	a magazine article	reading and writing
Biomedical sciences	case	terminology, collocations, grammar
	noun post-modification	grammar
Automotive engineering	noise	collocations
	verb + by vs with	colligations
Art and media	art vs arts	meanings and collocations
	however	writing

# Wood or wooden

## Task 1

Do you know the words *wood* and *wooden*? What do they mean? What is the difference between them?

## Task 2

Both words can be used before a noun. Study the sentences below. Can you see any differences in meaning and use between the two words?

- A brick wall features a bold mural; the **wood ceiling** and joists are painted bright yellow; furniture, floors, and walls are made of ash wood, treated with tongue oil or blackened.
- To achieve the transformation, ZGF played up the most striking aspect of the original building— its enormous clear-span floor space – leaving the vast **wooden ceiling** and its arching supports exposed, to stunning visual effect.
- **IMPROVED WOOD INTERIORS** Many commercial and residential interiors desire the use of wood for a more natural and comfortable indoor environment.
- The contemporary wing is as true to the spirit of the place as its 19th-century counterpart, with its envelope of corrugated steel, rough **wooden interior** finishes, and stone-hewn patio all exuding a rough-and-ready, pioneer spirit.
- To be durable, **wood structures** must be protected from water, insect, or fungal infiltration potentially through the use of sealants combined with the sound design and detailing of the building Mass Timber envelope.
- In his estimation, to adequately house Virginia's growing white constituency and government, construction practices needed to evolve beyond the production of the crude, ugly **wooden structures** and awkwardly proportioned brick buildings that were found in Williamsburg.

### **Task 3**

Study a random sample of examples (concordance lines) drawn from the Architecture corpus.

Which of the two words is more frequent?

Line	Left	KWIC	Right
1	House in Fontainhas, Porto, Portugal, 2019 Fluent white walls are stuck between striped	wooden floors	and light-blue ceilings, disturbed once in a while by blue doors, touches of marble or a
2	doors and windows to the outside, these Egyptian models are handily contained within the	wooden boxes	that serve as their outer walls. Not pictured here, the slaughterhouse also had a
3	. As the family became more established they upgraded the house by sheathing it in	wood clapboard	, a more weather-resistant and dignified finish. In time, successive occupants, tired
4	, with a glass facade overlooking the lake. Both garage and studio are covered with tarred	wood shingles	extending from their pitched roofs to the ground, "Tarred wood in Finland appears on
5	in with nature." The interiors of the studio (as well as the new black-waxed traditional	wood sauna	with a grass roof, next to the lake) are sheathed inside with pale spruce. "The interiors'
6	steep grade change to the classroom clusters. Students naturally have appropriated the	wood steps	of the tree-columned space—15 feet across at its widest point—for sunlight-drenched
7	BENSCHNEIDER; KEVIN SCOTT (OPPOSITE) THE LIBRARY (above and opposite) includes	wood blocks	that students can climb upon or rearrange to use as seating. 85 Copyright of
8	6 ELEVATOR 11 2 LOBBY 7 METAL SHOP 12 SEMINAR ROOM FLEX ROOM 3 FORUM STAIR 8	WOOD SHOP	13 MECHANICALS 4 ASSEMBLY ROOM 9 STUDIO ART 14 VIRTUAL REALITY AREA 5 PRINT
9	architects did reach into the adjacent studio-art center to build contiguous metal and	wood shops	. Then they created an open tri-level volume with a glazed assembly room and concrete
10	thermal performance properties of a metal roof. Shakes – Mimicking the look and style of	wooden shingles	, shake-inspired metal roofing panels offer a more classic look. Barbara
11	spaces. Here the architect divided each house into five individual but interconnected	wood structures	. Their exteriors are clad in Canadian cedar shingles, with roofs that are either pitched
12	piece de resistance is its 9-inch-deep dropped ceiling. Rendered in overlapping	wood slats	, it was inspired by the neckline of the kimono, which is created by folding layers of
13	customer's stocking feet to the cabinetry's smooth cedar cladding and the delicate	wooden fretwork	of the sliding doors for the six consultation rooms, where apparel options are presented
14	something else is going on; the piece is actually a sequence. The first stage is a simple	wooden hemisphere	; the second shows that same hemisphere cut into unequal segments but emanating from t
15	the Metropolitan Life Insurance Company tower in New York—a pencil drawing overlaid on a	wooden support	. Effectively, it's a model on a stick, made for Harvey Wiley Corbet in 1929 so the
16	PHOTOGRAPHY: © FRANCESCO RUSSO BY TIM ABRAHAMS 28 AT FIRST GLANCE, the	wooden assemblage	, about 2 feet long, could be a work of art—an early experiment by Noguchi, perhaps. But
17	something else is going on; the piece is actually a sequence. The first stage is a simple	wooden hemisphere	; the second shows that same hemisphere cut into unequal segments but emanating from t
18	the Metropolitan Life Insurance Company tower in New York—a pencil drawing overlaid on a	wooden support	. Effectively, it's a model on a stick, made for Harvey Wiley Corbet in 1929 so the
19	style; these materials also tend to be durable and strong. Classic combinations include	wood doors	and cabinetry with metal pulls and hardware; exposed wood beams and metal roofing; and
20	outdoors, and in both structural or non-structural applications. Of all the domestic	wood species	, coast redwood is among the most desirable. One of the lightest softwoods found in North
21	other biophilic materials and shapes, whether crisply painted cabinetry or a reclaimed	wood entry	door. Facilitating the Indoor-Outdoor Connection Custom home clients often request
22	. How to Calculate the Wood Carbon Footprint of a Building Expanding the possibilities of	wood building	design Sponsored by Think Wood   By Edie Sonne Hall, Ph.D. From an environmental
23	to develop a whole-building carbon footprint. 4. Define what is and is not included in a	wood EPD	and why. 5. Discuss the biogenic forest carbon cycle, and ways to track and assure forest
24	and assure wood comes from sustainable forests in North America and why demand for	wood products	supports investment in forest management. Source: © 2018 2030, Inc./Architecture 2030
25	LEED rating criteria include: • Renewable Resource: Redwood is a sustainable natural	wood product	. • Certified Wood: FSC certification indicating that redwood forests are
26	EDUCATION Photo courtesy of Bison Innovative Products Adjustable pedestals support	wood tiles	. of the product but does not "judge" whether the product meets any environmental quality
27	entrances, pathways, parking spaces, sitting spaces, and large open areas. Modular	wood elements	can be mixed with pavers or used by themselves to create unique aesthetics. Modular
28	installation through fastening kits that facilitate swift and secure installations of	wood surface	tiles and paver-tray-backed pavers without harm to the wood tiles or pavers. Deck System
29	installations of wood surface tiles and paver-tray-backed pavers without harm to the	wood tiles	or pavers. Deck System Components Deck systems are comprised of two major
30	/paver surfaces, including structural porcelain, stone, granite or concrete pavers,	wood tiles	, composite materials, fiberglass grating, or conventional joist and plank systems. As
31	multiple cultural influences. Traditional construction methods are centered around	wood framing	and product types that originated in the 19th and 20th centuries. Building on this
32	that sit on top of the roof membrane and allow the top surface to be leveled. Modular	wood tiles	or pavers then rest on top of the pedestals and allow rainwater to pass through the joints
33	access to other systems. If a lighter-weight surface material is preferred or needed,	wood tiles	are a good alternative, as they weigh only one-third as much as concrete tiles. Typically
34	, and Forest Stewardship Council (FSC)-certified hardwood options. If maintaining the	wood color	is desired, wood tiles can be periodically cleaned and sealed. Left to weather naturally
35	Council (FSC)-certified hardwood options. If maintaining the wood color is desired,	wood tiles	can be periodically cleaned and sealed. Left to weather naturally, they will develop a
36	just sized to accommodate the programmatic needs of the space. Site Furnishings Modular	wood cubes	are available with an array of design options to incorporate seating, storage, and
37	has been a surge of innovative decking products that are made from a combination of	wood fiber	and different types of plastic materials. These composite decking materials are
38	and separated from noncertified and noncontrolled material. This is important because	wood moves	through a supply chain, from the forest to the market, and can change hands and
39	of organizations. The certification process is an important step in verifying that	wood products	used in construction are in fact sustainable. Therefore, any design professional or
40	aesthetics to home designs. The use of hidden connectors and continuous visible	wood lengths	allow for clean, finished joints. There is also the benefit of a cleaner, lessobstructed



- **Task 4**

1. In some cases, the words *wood* and *wooden* used before the same noun can have very different meanings. Compare the noun phrases below and explain what they mean.

wood shop

wooden shop

wood colour

wooden colour

wood profile

wooden profile

wood screw

wooden screw

2. Fill out the gaps in the sentences below with *wood* or *wooden*.

- 

- Although the architect's design is nothing particularly special, this \_\_\_\_\_ **shop** was masterfully built by an expert carpenter.
- By the 1st century BC, \_\_\_\_\_ **screws** were commonly used throughout the Mediterranean world in devices such as oil and wine presses. Metal screws used as fasteners did not appear in Europe until the 1400s.
- It was decorated suitably in dark \_\_\_\_\_ **colours** and had ornamental Grapevine signs hanging on the walls.
- The \_\_\_\_\_ **profile** is not shown on the perimeter of the bookcase in order to emphasise the nested quality of the wood.
- West coast cedar (*Thuja plicata*) can be found at many lumber stores and is much easier to work with, though it is prone to splitting when using \_\_\_\_\_ **screws** without pre-drilling.
- The clubhouse, which has been built in conjunction with an existing boathouse, is clad with narrow, black-painted \_\_\_\_\_ **profiles**, laid in varying directions, which give the facade a simple but also refined and captivating look.
- To accommodate an ambitious program that includes state-of-the-art production, printing, and multidisciplinary facilities, the architects did reach into the adjacent studio-art center to build contiguous metal and \_\_\_\_\_ **shops**.
- If maintaining the \_\_\_\_\_ **color** is desired, wood tiles can be periodically cleaned and sealed.



3. Fill out the sentences below with the words *wood* or *wooden*, where appropriate. Remember the difference between the two, which you studied in Task 2.

- \_\_\_\_\_ **supply** is the term used to describe the estimated volume of timber that can be harvested from an area while meeting environmental, economic, and social objectives.
- A prefabricated \_\_\_\_\_ **beam** with steel diagonals running east–west in the roof of each linear, gabled 21-foot-high structure takes seismic loads along that axis.
- According to the company, whole timbers are 50% stronger than the largest \_\_\_\_\_ **member** milled from the same tree—and they sequester twice as much carbon.
- Appalachian hardwood is available for the floor in your home at wholesale \_\_\_\_\_ **prices**.
- But when it reopens, chocolate lovers will once again be confronted with a fresh interior in a palette of white, brown, and black, with carefully detailed metal fixtures and \_\_\_\_\_ **displays** for the chocolate on offer.
- CCWD summarizes the most common requirements for \_\_\_\_\_ **construction** in commercial buildings according to the 2015 IBC.
- Clad in a simple palette of cedar and cementitious stucco, the building's façade grows increasingly transparent as it rises from a monolithic \_\_\_\_\_ **base** to a framed, open-air terrace on the third floor.
- Get the latest research, news, and updates on innovative \_\_\_\_\_ **use**.
- However, much survived, including the walls and the beautiful \_\_\_\_\_ **pillars** behind the altar, as well as those separating the nave from the entrance lobby.
- Many materials such as vinyl and composites try to replicate the look of real cedar, but it is difficult to achieve the \_\_\_\_\_ **grain**, color, and feel of Real Cedar.
- Mjösa Tower—a hotel, housing, and office development—opened in March 2019, and assumed the mantle of the "world's tallest \_\_\_\_\_ **building**."
- New industry-average EPDs for wood and timber products are expected in the first quarter of 2020, but declarations from specific \_\_\_\_\_ **manufacturers** would be even more useful.
- Of all the domestic \_\_\_\_\_ **species**, coast redwood is among the most desirable.
- To meet these requirements, the architect prioritized natural materials, such as \_\_\_\_\_ **floors**, inherently flame-resistant wool carpets, felt seating spots, and cork- and linseed-based tackboards.

# case

4. You will study the same pattern in a much larger corpus of General English (over 1 billion words). Go to the following address and register for a free account. Then follow the instructions below:

Corpus of Contemporary American English (COCA) <https://www.english-corpora.org/coca/>

- In the 'SEARCH' tab, type *in \* cases* in the query box. Next, press the button 'Find matching strings'.

The screenshot shows the COCA website interface. The top navigation bar includes 'SEARCH', 'FREQUENCY', 'CONTEXT', and 'ACCOUNT'. The 'SEARCH' tab is active. In the search box, the query 'in \* cases' is entered, with a '[POS]?' label to its right. Below the search box are buttons for 'Find matching strings' and 'Reset'. To the right of the search box, there is a 'NO LICENSE' button and a green button that says 'Download the corpus (and corpus-based frequency data) for offline use'. Below these buttons are links for 'PDF overview' and 'Five minute tour'. The main content area contains a paragraph describing the COCA corpus as the only large and representative corpus of American English, and its relation to other corpora.

# Corpus of Contemporary American English

SEARCH    **FREQUENCY**    CONTEXT    ACCOUNT

ON CLICK: [CONTEXT](#) [TRANSLATE \(PL\)](#) [ENTIRE PAGE](#) [GOOGLE](#) [IMAGE](#) [PRON/VIDEO](#) [BOOK](#) (HELP)

HELP		ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 42,647   UNIQUE 1,151 +
1		<a href="#">IN SOME CASES</a>	13856	<div style="width: 100%;"></div>
2		<a href="#">IN MANY CASES</a>	6986	<div style="width: 50%;"></div>
3		<a href="#">IN MOST CASES</a>	5375	<div style="width: 38%;"></div>
4		<a href="#">IN BOTH CASES</a>	2637	<div style="width: 19%;"></div>
5		<a href="#">IN THESE CASES</a>	1706	<div style="width: 12.5%;"></div>
6		<a href="#">IN ALL CASES</a>	1561	<div style="width: 11.5%;"></div>
7		<a href="#">IN OTHER CASES</a>	1385	<div style="width: 10.2%;"></div>
8		<a href="#">IN SUCH CASES</a>	1382	<div style="width: 10.1%;"></div>
9		<a href="#">IN THOSE CASES</a>	926	<div style="width: 6.8%;"></div>
10		<a href="#">IN THE CASES</a>	744	<div style="width: 5.5%;"></div>
11		<a href="#">IN CERTAIN CASES</a>	417	<div style="width: 3.1%;"></div>
12		<a href="#">IN EXTREME CASES</a>	364	<div style="width: 2.7%;"></div>
13		<a href="#">IN RARE CASES</a>	337	<div style="width: 2.5%;"></div>
14		<a href="#">IN CRIMINAL CASES</a>	252	<div style="width: 1.8%;"></div>
15		<a href="#">IN SEVERAL CASES</a>	235	<div style="width: 1.7%;"></div>
16		<a href="#">IN TWO CASES</a>	178	<div style="width: 1.3%;"></div>
17		<a href="#">IN SEVERE CASES</a>	147	<div style="width: 1.1%;"></div>
18		<a href="#">IN CIVIL CASES</a>	119	<div style="width: 0.8%;"></div>
19		<a href="#">IN INDIVIDUAL CASES</a>	98	<div style="width: 0.7%;"></div>
20		<a href="#">IN PARTICULAR CASES</a>	86	<div style="width: 0.6%;"></div>

# noise

## Task 3

Compare the collocational profile you created based on an automotive industry corpus to the collocation profile based on a corpus of general English. Follow the instructions below. Can you see any differences between the two profiles? How can you explain them?

- Go to the following address: SKELL <https://skell.sketchengine.eu/>.
- Type 'noise' in the search box and press 'Enter'.
- Select the 'Word sketch' tab under the search box at the top of the screen.
- Study the collocates.
- Click on selected collocates and study their examples in sentences.



noise English

Examples Word sketch Similar words

noise noun Show context

verbs with noise as subject	verbs with noise as object	adjectives with noise
1. <a href="#">cancel</a> noise cancelling headphones	1. <a href="#">hear</a> heard a noise	1. <a href="#">audible</a> noise audible
2. <a href="#">deafen</a> The noise was deafening	2. <a href="#">buzz</a> buzzing noise	2. <a href="#">loud</a> noise was so loud
3. <a href="#">startle</a> noise startled	3. <a href="#">hiss</a> a hissing noise	3. <a href="#">bothersome</a>
4. <a href="#">disturb</a> disturbed by the noise	4. <a href="#">reduce</a> reduce noise	4. <a href="#">unbearable</a> noise was unbearable .
5. <a href="#">awaken</a> awakened by the noise	5. <a href="#">click</a> clicking noise	5. <a href="#">Gaussian</a>
6. <a href="#">shape</a> noise shaping	6. <a href="#">rattle</a> a rattling noise	6. <a href="#">ear-splitting</a>
7. <a href="#">emanate</a> noises emanating from	7. <a href="#">deafen</a> the deafening noise	7. <a href="#">inaudible</a>
8. <a href="#">frighten</a> frightened by the noise	8. <a href="#">emit</a> noise emitted	8. <a href="#">intrusive</a>
9. <a href="#">wake</a> noise woke	9. <a href="#">minimize</a> to minimize noise	9. <a href="#">non-stationary</a>
10. <a href="#">filter</a> noise filtering and	10. <a href="#">pop</a> popping noise	10. <a href="#">nuisance</a>
11. <a href="#">distract</a> distracted by the noise	11. <a href="#">grind</a> grinding noise	11. <a href="#">excessive</a>
12. <a href="#">drown</a> noise drowns out	12. <a href="#">muffle</a> muffle the noise	12. <a href="#">horrendous</a>
13. <a href="#">bother</a> bothered by the noise	13. <a href="#">rumble</a> a rumbling noise	13. <a href="#">unobtrusive</a>
14. <a href="#">measure</a> the noise measuring station	14. <a href="#">hum</a> a humming noise	14. <a href="#">uncorrelated</a>





noise

English

Examples

Word sketch

Similar words

## noise + audible 0.01 hits per million



1. The traffic **noise** is scantily **audible** on very quiet and still mornings , particularly on the ridges.
2. They do cause explosions, electrical issues, and other **noises audible** for blocks around the site.
3. It appears that while Mr. Cyrus Field was on watch in the tank, a little before the time of the accident a grating **noise** was **audible** as the cable flew over the coil astern.
4. They sailed overhead and swayed with the slightly erratic action of the cableway holding them, bumping into each other now and again and so producing the clinking **noise** just **audible** over the din.
5. The white noise that is introduced by the radio is then also de-emphasized at the higher frequencies (where it is most noticeable) along with the pre-emphasized program, making the **noise** less **audible** .
6. His **Noise** is **audible** once more, and he openly thinks about how beautiful Viola is and how he wants to hold her.
7. The exercise consists in calling attention, when perfect silence has been established, to the ticking of the clock, and to all the little **noises** not commonly **audible** to the ear.
8. By the late 1960s, disk reproducing equipment became so good that audiophiles soon became aware that some of the **noise audible** on recordings was not surface noise or deficiencies in their equipment, but reproduced tape hiss.
9. The show also includes an on-screen timer whose loud ticking **noise** is evidently **audible** on set, disturbing Walker as she tries to cook.
10. The Hum is a phenomenon, or collection of phenomena, involving widespread reports of a persistent and invasive low-frequency humming, rumbling, or droning **noise** not **audible** to all people.
11. The ship started to capsiz slowly at first but with increasing speed, accompanied by increasing **noise audible** on the